



# Children's Learning of the Compositional Number System

## Citation

Srinivasan, Akshita. 2024. Children's Learning of the Compositional Number System. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

## Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37378875>

## Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

## Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#)

# HARVARD

Kenneth C. Griffin



GRADUATE SCHOOL  
OF ARTS AND SCIENCES

## DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the  
Department of Psychology  
have examined a dissertation entitled  
Children's Learning of the Compositional Number System

presented by Akshita Srinivasan

candidate for the degree of Doctor of Philosophy and hereby  
certify that it is worthy of acceptance.

Signature

Typed name: Prof. Elizabeth Spelke

Signature

Typed name: Prof. Susan Carey

Signature

Typed name: Prof. Jesse Snedeker

Signature

Typed name: Prof. Jon R. Star

Signature

Typed name: Prof.

Date: April 23, 2024

# Children's Learning of the Compositional Number System

A dissertation presented

by

Akshita Srinivasan

To

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

April 23, 2024

© 2024 — Akshita Srinivasan

All rights reserved

## **Children’s Learning of the Compositional Number System**

### **Abstract**

The compositional system of natural numbers underlies our abilities to measure time, use money, and represent quantities. In the present dissertation, I examined how children master the compositional number system by (i) investigating the challenges that they face in learning it and (ii) evaluating game-based interventions to discern causal impacts of representations that facilitate their learning. In Chapter 2, I provide evidence using home-based games that training children’s innate numerical concepts does not support their learning of the symbolic number system taught in school. In Chapter 3, I provide evidence using cross-linguistic comparisons that the degree of transparency of number words in making the compositional rules of the number system explicit (e.g., “eleven” in English that translates to “ten one” in Mandarin) affects children’s counting skills. However, mastering the compositional number system is hard for children even when their language’s number words reveal its logic more transparently. In Chapters 4 and 5, I provide evidence using home-based games that training children’s understanding of the base-10 compositional number system, using multiple representations, successfully enhances their learning of it and the numerical concepts it supports. Crucially, some games in this dissertation were evaluated as a part of multifaceted school-based interventions that had distinct impacts on children’s math learning. This dissertation evaluated different components of the multifaceted interventions that could have accounted for their success. Together, this dissertation sheds light on children’s learning of the compositional number system and highlights the importance of small-scale home-based interventions conducted in synergy with large-scale school-based interventions to deepen our understanding of how children learn in their real-world environments.

## Table of Contents

|                                                                                                        |            |
|--------------------------------------------------------------------------------------------------------|------------|
| <b>Title page.....</b>                                                                                 | <b>i</b>   |
| <b>Copyright.....</b>                                                                                  | <b>ii</b>  |
| <b>Abstract.....</b>                                                                                   | <b>iii</b> |
| <b>Table of Contents.....</b>                                                                          | <b>iv</b>  |
| <b>Acknowledgments.....</b>                                                                            | <b>vii</b> |
| <b>1. Introduction.....</b>                                                                            | <b>1</b>   |
| 1.1 Children’s early emerging numerical knowledge.....                                                 | 2          |
| 1.2 Children’s learning of the base-10 compositional number system.....                                | 4          |
| 1.3 Challenges in learning the base-10 compositional number system.....                                | 6          |
| 1.3.1 Opaque number words and symbols.....                                                             | 6          |
| 1.3.2 Multiple representations of the same concept.....                                                | 8          |
| 1.4 Overview of the dissertation research.....                                                         | 10         |
| 1.4.1 Synergy between short-term home-based interventions and long-term field-based interventions..... | 11         |
| 1.4.2 Overview of Chapter 2.....                                                                       | 13         |
| 1.4.3 Overview of Chapter 3.....                                                                       | 15         |
| 1.4.4 Overview of Chapter 4.....                                                                       | 17         |
| 1.4.5 Overview of Chapter 5.....                                                                       | 18         |
| <b>2. Intuitive and Symbolic Numerical Concepts.....</b>                                               | <b>20</b>  |
| 2.1 Introduction.....                                                                                  | 20         |
| 2.2 Methods.....                                                                                       | 25         |
| 2.3 Results.....                                                                                       | 32         |

|                                                           |            |
|-----------------------------------------------------------|------------|
| 2.4 Discussion.....                                       | 38         |
| <b>3. Language and Numerical Concepts.....</b>            | <b>41</b>  |
| 3.1 Introduction.....                                     | 41         |
| 3.2 Study 1.....                                          | 52         |
| 3.2.1 Methods.....                                        | 52         |
| 3.2.2 Results.....                                        | 57         |
| 3.2.3 Discussion.....                                     | 64         |
| 3.3 Study 2.....                                          | 66         |
| 3.3.1 Methods.....                                        | 67         |
| 3.3.2 Results.....                                        | 69         |
| 3.3.3 Discussion.....                                     | 75         |
| 3.4 General Discussion.....                               | 77         |
| <b>4. Compositional Numerical Concepts.....</b>           | <b>81</b>  |
| 4.1 Research Highlights.....                              | 81         |
| 4.2 Abstract.....                                         | 81         |
| 4.3 Introduction.....                                     | 82         |
| 4.4 Methods.....                                          | 93         |
| 4.5 Results.....                                          | 101        |
| 4.6 Discussion.....                                       | 105        |
| <b>5. Compositional Numerical and Reading Skills.....</b> | <b>112</b> |
| 5.1 Introduction.....                                     | 112        |
| 5.2 Methods.....                                          | 118        |
| 5.3 Results.....                                          | 133        |

|                                                                                                       |            |
|-------------------------------------------------------------------------------------------------------|------------|
| 5.4 Discussion.....                                                                                   | 137        |
| <b>6. Conclusion.....</b>                                                                             | <b>142</b> |
| 6.1 Highlights of Chapters 2-5.....                                                                   | 142        |
| 6.2 What representations help children learn base-10 compositionality?.....                           | 144        |
| 6.3 Do children rely on compositional reasoning to acquire the meanings of single-digit numbers?..... | 145        |
| 6.4 Children's compositional reasoning beyond the number system.....                                  | 146        |
| 6.5 Final Conclusion.....                                                                             | 147        |
| <b>7. Appendix.....</b>                                                                               | <b>148</b> |
| 7.1 Supplementary Information (SI) for Chapter 3.....                                                 | 148        |
| 7.1.1 Tasks.....                                                                                      | 148        |
| 7.1.2 Math and language survey.....                                                                   | 150        |
| 7.1.3 Results from Study 1.....                                                                       | 150        |
| 7.1.4 Results from Study 2.....                                                                       | 155        |
| 7.2 Supplementary Information (SI) for Chapter 4.....                                                 | 161        |
| 7.2.1 Descriptions of the games.....                                                                  | 161        |
| 7.2.2 Descriptions of the assessments.....                                                            | 163        |
| 7.2.3 Analyses.....                                                                                   | 170        |
| <b>References.....</b>                                                                                | <b>178</b> |

## **Acknowledgments**

This work would not have been possible without the support of many people who supported my journey.

First, all the children and families who participated in my studies. I am so grateful that they volunteered their precious time to support Science.

My advisor, Elizabeth Spelke, who decided to take a chance and accept someone with very little background in developmental cognitive science into her lab. Her enthusiasm for Science inspires me every day. Conversations with her about research were some of the most generative and fun conversations I have ever had, and they will remain a highlight of my time at Harvard. I am grateful to Liz for encouraging me to think about the bigger impact of my work and for giving me the courage to pursue tough questions. Whenever I asked Liz for advice, she would say “well, what would YOU have fun doing?”, and I value the lesson that she was trying to impart with that question—that Science can be fun!

Members of my committee. Susan Carey from whom I learned so many lessons about mentorship, Science, and writing. Being a part of Susan’s writing workshop was one of the most enjoyable and transformative experiences of grad school for me. I am grateful for all the time Susan spent in helping me work my scientific arguments out and for pushing me to think about the learning story that underlies my work. Jesse Snedeker, who helped me think critically about my work. I am so grateful for Jesse’s engagement with me and my work as she made me a better thinker and communicator. Jon Star, who welcomed a very lost second year grad student to his lab meetings. I am grateful for Jon’s kindness in introducing the world of math education to me. I am always touched by Jon’s support and encouragement.

My mentors before and during grad school. Telli Davoodi, who gave me the first entry into the field while she was at Boston University. I am grateful to her for being the most fun and supportive mentor ever. Telli is the reason I pursued this field. Shari Liu, who nurtured my interest in the field and made me a rigorous scientist. Shari holds herself and the people around her to the highest standards and always gives me something to aspire to. Martin Zettersten, whom I got connected with via CDS's mentorship program in the final year of grad school. I am so grateful to Martin as he always made me feel like I've got this. Martin is one of the kindest and most generous people I had the privilege of meeting and he got me through the job market and an overwhelming final year of the PhD.

My collaborators and co-authors including Molly Dillon, Simge Topaloglu, Kexin Que, Stella Christie, Cigdem Ileri, Aylin Kuntay, Laura Mullertz and Chrissie Carvalho who worked hard with me on this research program.

My mentees including Gianna, Lucia, Sam, Cassie, Amber, and Gabe who were as invested in the success of my studies as me. Working with them was a highlight of my PhD and they gave me something to look forward to when progress felt slow.

Members of the Spelke Lab I had the privilege to overlap with including Brandon, Emily, Shari, Ashley, Heather, and Bill. I am grateful for their willingness to share their advice, wisdom, and expertise with me. They always embodied a team Science spirit and created a positive and supportive lab community. Special mention to our lab manager, Cristina, who does so much for our lab to enable all its research. Cristina rebuilt our lab community post-Pandemic and I am so grateful to her for helping us all feel welcome and supported in the lab.

Members of the Cambridge Writing Group including Ashley, Shari, Brandon, Leyla, Maya, Halie, Hayley, Junyi, and Qiong. Writing group transformed my relationship with writing and

made me feel like I belong in academia. I am grateful to Ashley for bringing writing group to Cambridge.

Faculty at Harvard whom I taught with: Mina Cikara, Fiery Cushman, and Leah Somerville. Teaching was a big part of my journey at Harvard and it informed what I wanted to do with my life going ahead. I am grateful that I got to learn about teaching from the best teachers Harvard Psychology has.

Patrick Mair, who taught me Statistics and R. I am so grateful for a teacher like Patrick who made Statistics fun and enjoyable. He is single-handedly responsible for every single analysis and plot that appears in this dissertation.

My academic community and support network including staff of the Psychology department (Allie, Krista, Joan, and Katie) and members of Harvard LDS for their feedback and continuous support of my research.

My friends within the department including Irene, Genia, Simge, Brandon, Hao, Yelina, Sarah, and Josh who made my time at Harvard so enjoyable and whose genuine care and affection deeply impacted me.

My mom who enabled me to be my own person and chart the course of my life the way I wanted to. I am so grateful and indebted to her for all the hard choices she made to give me this life. Her unconditional support and love are what got me so far.

My dad, who is always my number one cheerleader. I am grateful for his selflessness and belief in me.

My partner, Mohit, who in 2016 very casually said “well you can apply to Harvard LDS next year”. At the time, the idea sounded bizarre to me. I am grateful to him for believing in me

before I believed in myself. Sharing my life with him in Boston is an absolute joy and I am grateful I had him to go through the ups and downs of life with.

Finally, my dog and my best friend, Minnie. She came into my life in the middle of grad school and was single-handedly responsible for an increase in my happiness since then. Her enthusiastic welcome when I came back home from work got me through even the most tough days. She reminds me to laugh every day and be grateful for the little things in life.

## Chapter 1

### Introduction

Compositional reasoning is a key feature of rich and efficient human learning as it helps us make “infinite use of finite means” (Fodor, 1975; Fodor & Pylyshyn, 1988; Lake et al., 2017; Marcus, 2001; Piantadosi, 2011). For example, we can combine finite words to utter and understand infinite thoughts and we can combine a finite set of numbers to have unbounded abilities to calculate and quantify the world around us. My dissertation investigates how children learn the compositional rules that underlie the natural number system. More specifically, I (i) investigate why learning the compositional number system might be hard for children and (ii) evaluate game-based interventions to help children learn the compositional number system.

In the forthcoming sections of this introduction, I first briefly review children’s early numerical knowledge that develops prior to their learning of the compositional number system. Then, I outline the importance of learning the compositional number system and the challenges that children face as they seek to master it. This section also serves to highlight the foundational concepts that children must master in order to gain a complete understanding of the compositional number system. Finally, I provide an overview of the dissertation research by outlining the four empirical studies that comprise it and highlight the importance of small-scale home-based interventions, conducted in synergy with large-scale school-based interventions, to deepen our understanding of how children learn in their real-world environments. This research aims to shed light on the specific experiences that combine to give rise to children’s productive learning of school-relevant numerical concepts.

## 1.1 Children's early emerging numerical knowledge

Our ability to represent certain numerical concepts arises as early as infancy. Infants have numerical concepts that are supported by the core knowledge systems of number. These systems are innate and contain representations of (i) sets and their approximate numerical magnitudes (e.g., the distinction between 20 and 40), hereafter referred to as the Approximate Number System (ANS) and (ii) small sets and their exact magnitudes (i.e., the distinctions between 1, 2, and 3), hereafter referred to as the Parallel Individuation (PI) system (see Carey, 2009; Dehaene, 1997; Feigenson et al., 2004; Spelke, 2011; Spelke 2022 for reviews).

However, neither system of core knowledge (ANS or PI) on its own supports the exact representations of large quantities (e.g., the distinction between 20 and 21). Representations of large exact quantities arise only when children master natural number concepts. Natural number concepts are a crucial cultural achievement that children come to acquire: they underlie our abilities to measure time, use money, and represent quantities. Their developmental origins in humans have been studied for decades since Piaget (1952). Several proposals since then have outlined how children might acquire the earliest integers and the principles of natural numbers such as (i) “one” is the smallest unit that separates consecutive integers, (ii) integers are generated by the successive addition of one, and (iii) integers refer to exact cardinal values.

Gelman and Gallistel (1978) proposed that natural number concepts are *innate* and that they do not depend upon any specific experiences with sets of objects. Therefore, according to this proposal, natural number concepts are culturally universal. By contrast, Carey (2009) proposed that natural number concepts are *learned* based upon experience with a counting procedure. Therefore, according to this proposal, natural number concepts vary by culture as some languages (e.g., Tsimane) do not have a count list. Spelke (2017) proposed that natural number concepts arise

from the core knowledge systems of number, whose representations are productively composed via natural language. Therefore, according to this proposal, natural number concepts emerge universally upon learning one's language, and they do not depend upon mastering a count list.

All the proposals outlined above aim to explain how children come to learn the meanings of the smallest integers (e.g., 1-10) and how they learn that larger integers like “53” and “54” represent different quantities that vary exactly by one. These are important insights that children take several years to master (Fuson, 1988; Sarnecka & Carey, 2008; Wynn, 1990, 1992). Children must build on these insights and also master another crucial feature of the symbolic system that represents natural numbers: it is a compositional system where larger numbers are composed of just ten digits (0 to 9) multiplied by some power of the base number ten (e.g., “54” is “five tens and four ones”). How do children acquire this base-10 compositional number system?

Children start developing some insights about the symbolic system that represents the natural numbers much before they begin any formal instruction in school. For example, by age 5 or earlier they (i) have the intuition that numbers are hierarchically organized (Feigenson & Halberda, 2004; Rosenberg & Feigenson, 2013), (ii) can recite the count list and learn how counting represents number (Wynn, 1990, 1992), (iii) know that number words refer to exact cardinal values (Lipton & Spelke, 2006), (iv) know that two consecutive numbers in the count list vary exactly by one (Sarnecka & Carey, 2008), and (v) solve approximate arithmetic problems presented using number symbols and words (Gilmore, McCarthy, & Spelke, 2007).

They also have intuitions about large multidigit number words and symbols. For example, they can map number words (“two hundred thirty-five”) to their written symbols (“235” and not “523”) (Mix et al., 2014; Yuan et al., 2019) and judge which of the two symbols designates a larger number (Cheung & Ansari, 2021; Mix et al., 2014; Yuan et al., 2019). However, children can solve

such tasks using approximate heuristics. For example, when comparing magnitudes, they might use heuristics such as attending to the number of digits within numerals or to the leftmost digit when two numbers are composed of the same number of digits. When mapping between number words and symbols they might simply map the spoken order of number words to the left-to-right ordering of the number symbols (Byrge et al., 2014; Cheung & Ansari, 2021; Mix et al., 2014).

Children seem to be utilizing the structural relationships within the number symbols to develop intuitions about multidigit numbers without necessarily knowing why the heuristics are effective or what exact quantities the numbers designate. This kind of emerging knowledge about multidigit numbers is separate from a complete understanding of base-10 compositionality, which enables children to represent multidigit numbers as composed of sets of tens and ones (Mix et al., 2022). However, these intuitions about multidigit numbers that children possess by Kindergarten age are important as they predict base-10 understanding in first grade (Mix et al., 2022; Mix et al., 2023).

In summary, even though young children come to develop a host of numerical concepts informally (i.e., before any explicit instruction in school), they struggle to master the base-10 compositional number system even after formal instruction begins (Fuson, 1988; Mann et al., 2012; Ross & Sunflower, 1995). Therefore, my dissertation investigates why learning the base-10 compositional number system might be hard for children and evaluates game-based interventions to help children master base-10 compositionality and the mathematical concepts it supports.

## **1.2 Children's learning of the base-10 compositional number system**

Compositional systems have an extraordinary expressive power: a limited set of foundational concepts, together with the rules for combining them, lead to potentially unbounded concepts (Fodor, 1975, Fodor & Pylyshyn, 1988). In the number domain, just ten digits (0 to 9)

multiplied by some power of the base number ten can generate all the natural numbers. An understanding of the compositional rules that underlie the natural number system along with representations of multidigit numbers as sets of tens and ones is crucial for children's abilities to learn mathematics in school (Chan et al., 2014; Moeller et al., 2011; Ross, 1986) and productively use number in everyday life (Banerjee et al., in prep). For example, being able to compose and decompose sets of tens aids faster mental calculation (Laski et al., 2014; Laski et al., 2016).

Mastering the base-10 compositional number system provides children with both opportunities and challenges. It provides children with the ability to transcend the limits of the core knowledge systems of number and develop precise representations of large quantities. It also helps them convert large and otherwise intractable quantities into short strings of words and symbols, composed of a small number of elements that they can learn to read, write, and manipulate. For example, "234" can be expressed using just those symbols as long as children understand how the symbols combine to represent quantities.

How do children come to learn the base-10 compositional number system? Do they acquire it via exposure to multidigit number words or number symbols or both? Do they acquire it only after formal arithmetic instruction whose algorithms depend upon the base-10 number system? Or, do children benefit with symbol grounding where abstract symbols are instantiated through physical models such as abacus, numerical boards, or base-10 blocks (e.g., small cubes represent ones and bars that contain ten small cubes represent tens) that convey some aspects of base-10 compositionality?

Evaluations of interventions that systematically test the efficacy of each of these different representations to teach base-10 compositionality can help us establish their causal impacts and examine which representations are more effective than the others. Studying how children learn

base-10 compositionality can inform a challenge that any school curriculum aims to tackle: helping children become proficient in mathematics.

### **1.3 Challenges in learning the base-10 compositional number system**

Despite the usefulness and ubiquity of the base-10 compositional number system, it poses challenges for child learners for at least two reasons: (i) the rules for composing numerical concepts are not directly apparent in the language and symbols that express those concepts and (ii) it requires children to represent the same concept (e.g., “ten”) in multiple ways (as *one ten* and *ten ones*) and switch between the representations. I elaborate on both of these challenges below.

#### **1.3.1 Opaque number words and symbols**

First, some number words in English and in many other languages obscure base-10 compositionality. For example, in English, the words for the smallest multidigit numbers, “eleven” and “twelve,” do not explicitly convey the decomposition into their respective ten and ones. The other number words for the “teens” (e.g., “thirteen” and “fourteen”) and for the decades (e.g., “twenty” and “thirty”) also obscure base-10 compositionality as the morphemes “ty” and “teen” do not transparently map on to the “ten” that they represent.

In East-Asian languages (e.g., Mandarin), base-10 compositionality is more transparently reflected in the number words (e.g., “eleven” translates to “ten one” and “twenty” translates to “two tens”). Some studies have found that the transparency of number naming systems matters for children’s understanding of base-10 compositionality (Miura, 1987; Miura et al., 1988, Miura & Okamoto, 1989; Miura et al., 1993; Miura et al., 1994; Ho & Fuson 1998) and their counting skills (Miller et al., 1995; Miller & Stigler, 1987) such that children who speak transparent East-Asian languages outperform their age-matched peers who speak relatively non-transparent languages.

However, even in the transparent number naming systems of East-Asian languages, it is not clear which words relate to each other by addition and which ones by multiplication in order to reflect a single quantity. For example, “ten one” are related by addition to represent “eleven” and “two ten” are related by multiplication to represent “twenty”. There is some evidence demonstrating that children speaking East-Asian languages also learn slowly about the compositional rules that generate the decades. For example, while counting, children are more likely to stop at the number right before a decade (e.g., “forty-nine”) as compared to any of the other numbers between two decades (Miller et al., 1995; Srinivasan et al., in prep).

Second, the written symbols for numbers depend upon an understanding of the conventional system of place value: the positions that the digits occupy within a number matter for magnitude. Thus, the “3” in “32” and “23” represent two different quantities. The symbols that children learn to represent numbers do not explicitly convey the concept of place value and the important rule that numbers increase by powers of ten from right to left. Moreover, this ordering might be counterintuitive to children as it is opposite to the left-to-right ordering that is conventional for displaying numbers on rulers, number lines, and numerical board games for children (e.g., chutes and ladders).

Third, children also need to create mappings between spoken number words and written number symbols. However, creating these mappings can pose problems for children as in some languages, and for some number words, the ordering of digits within a number and the spoken numerical morphemes are reversed (Klein et al., 2013). For example, the English word for sixteen, places “six” before “teen”, where “teen” is the morpheme designating ten, whereas the written symbol for “sixteen” places the quantifier that designates the “tens” (i.e., “1”) before the “6” that designates the “ones”.

Children as young as 3 years of age use heuristics such as mapping the spoken order of number words (“four hundred and twenty-five”) on to the order of digits within a number (425 and not 542) (Mix et al., 2014; Yuan et al., 2019) without knowing what exact quantities those numbers represent or how they reflect base-10 compositionality. These initial heuristics might help them break into the base-10 compositional number system, but when the mappings between number words and symbols violate the left-to-right mapping, they might pose additional challenges for children.

In summary, the written and spoken symbol systems representing natural numbers pose challenges for children as they do not explicitly convey the compositional rules that underlie the base-10 number system.

### **1.3.2 Multiple representations of the same concept**

In order to master base-10 compositionality, children must learn to represent the same number in multiple ways and flexibly switch between the different representations. For example, “ten” must be represented both as *one ten* and *ten ones*. For example, when counting by tens, children must represent a set of ten as a single entity that they can count by. However, they must also represent the individual ones within a set of ten in order to be able to manipulate it when performing the algorithms of arithmetic such as carrying in addition and borrowing in subtraction.

Without being able to represent “ten” in two distinct ways, when adding numbers presented as symbols, children may apply the addition algorithms as a series of arbitrary steps for manipulating individual numerals rather than as a series of logical steps that treat quantities as compositions of tens and ones. Performing arithmetic by rote-learning arbitrary steps might lead to error prone calculations even if a single step is misapplied, and it does not lead to a productive

ability to calculate with numbers. Therefore, representations of ten as *one ten* and *ten ones* are crucial for children's mathematical learning.

Research in a different domain has found that young children find it challenging to understand multiple representations of the same concept. In classic theory of mind research, Flavell et al., (1983) showed 3-year-old children a sponge that was painted to look like a rock. Most 3-year-old children in the experiment were unlikely to say that the object is really a sponge but looks like a rock. Therefore, they had trouble maintaining representations of the object either as a sponge or a rock and failed to understand that there could be alternative representations to the same object.

Similarly, Gopnik and Astington (1988) showed 3-year-old children a closed box of candies and asked them what they thought was inside. Children answered the question correctly by saying "candies". Then the experimenters opened the box to reveal its unexpected contents of pencils. They then asked children what they thought was inside the box before it was opened. Most 3-year-olds failed to answer the questions correctly. That is, they said that they thought it contained pencils even before it was opened. Therefore, children failed to maintain distinct representations of their present belief of the contents of the box (pencils) and their past belief before the box was opened (candies).

In the same theory of mind tasks, children seem to understand and maintain multiple representations of the same object at around age 5 (Flavell, 1986; Flavell et al., 1983; Gopnik & Astington, 1988). However, in the numerical domain, it is unclear when children begin to represent ten both as *one ten* and *ten ones*. This understanding in the numerical domain might emerge at a later age as children become exposed to numerical concepts in school.

Some indirect evidence for children's lack of understanding of multiple representations of the same number comes from a task designed to measure base-10 understanding. Children were

asked to count objects (e.g., cookies) represented both as sets of tens (bags of ten cookies) and ones (individual cookies). The most common pattern of error was that children counted the total number of elements in the set rather than counting by tens and ones (Chan, Au, Tang, 2014; Srinivasan et al., in prep). That is, for example, when shown two opaque bags of ten cookies and five single cookies, children counted the total number of cookies as “seven” instead of “twenty-five”. Crucially, right before the trials for the task began, children correctly answered that each bag contains ten cookies inside (Srinivasan et al., in prep).

Since the ten cookies were not visible inside the bag, they had trouble holding its representation as a single set of ten ones and counting over different sets of quantities (i.e., tens and ones). This pattern of results does not help us distinguish between two reasons for children’s failures: an inability to count by tens or an inability to hold representations of ten as *one ten* and *ten ones*. However, it is possible that despite knowing how to count by tens (e.g., ten, twenty, thirty, and so on) children might struggle with representing the same concept of ten in multiple ways.

In summary, holding representations of ten as *one ten* and *ten ones* and switching between them might pose challenges for children when they aim to master the arithmetic algorithms that are supported by an understanding of the base-10 compositional number system.

#### **1.4 Overview of the dissertation research**

Children acquire crucial insights about natural numbers and the symbolic system that represents them via informal exposure either through their home environment or through play-based preschool activities. However, there is variation in the amount of this school-relevant knowledge that children possess even before beginning kindergarten and this variation predicts

their math achievement scores at least until middle school and possibly beyond (Duncan et al., 2007; Griffin, Case, & Siegler, 1994).

Given the importance of children's early developing mathematical knowledge for later school learning, my dissertation research asks 1) does early developing numerical knowledge causally support math learning in school, 2) why are some school-relevant numerical concepts hard for children to learn and 3) what kind of training best fosters the learning of such numerical concepts. I take two main approaches to answer the questions above: (1) I use home-based interventions both to evaluate their effectiveness as well as establish the causal impacts of certain numerical activities in supporting children's developing numerical concepts, and (2) I utilize the variation in languages' number naming systems to examine whether it impacts children's developing numerical understanding.

In Chapter 2, using home-based interventions, I studied whether games that recruit children's core knowledge systems of number (ANS and PI) together causally support their learning of school-relevant symbolic numerical concepts. In Chapter 3, I studied whether the degree of transparency of number words in five different languages in making the compositional rules of the base-10 number system explicit affects children's numerical understanding. In Chapters 4 and 5, I evaluated games using home-based interventions that aim to train children's understanding of base-10 compositionality.

#### **1.4.1 Synergy between short-term home-based interventions and long-term field-based interventions**

In this research, we aimed to evaluate our numerical games using home-based interventions for two reasons. First, children's home environments correlate with their school-relevant learning (Gunderson & Levine, 2011; Levine et al., 2010; Niklas & Schneider, 2014). Therefore, by

providing families with educational materials informed by research in developmental cognitive science, we can enhance children's learning as well as establish causal impacts of our home-based activities for school-relevant learning. Second, home-based interventions give us a snapshot into children's learning process as they unfold in their real-world environments that are much noisier and more complex as compared to the well-controlled lab environments in which many previous interventions have been evaluated. Such research can shed light on the specific home experiences that combine to enhance children's school-relevant learning.

Moreover, some of the games in our home-based interventions were also evaluated as a part of large-scale field studies in children's schools. Shorter interventions, administered in children's natural learning environments like their homes, can serve as a complement to large-scale field studies in two ways. First, they can serve as a prelude to large-scale school-based interventions: for example, if experiments were to establish that short-term play of four different games each led to enhanced learning of skills that schools aim to foster, then a school-based intervention using that suite of games, played over many months, could be worth implementing and evaluating. Second, they can serve as a follow-up to school-based interventions by probing the cognitive processes and states that are impacted by training and the effects of training that afforded those impacts, thereby shedding light on the aspects of the interventions that accounted for their success.

In order for large-scale, randomized interventions in children's schools to be effective, they must be long in duration, because children master the logic of primary school mathematics slowly. If they are long, however, they also must be multifaceted, both because multifaceted interventions provide richer opportunities for learning and because they are better able to maintain children's interest over the long term. When a multifaceted intervention succeeds, its success does not reveal

which components of the intervention were effective, or what specific effects they produced. As a result, it is not clear which aspects of such an intervention should be built on to improve mathematics education even further.

The present dissertation illustrates one manner in which large-scale, long, field experiments in education and small-scale, brief experiments in developmental cognitive science can synergize: home-based interventions can test different components of a large-scale interventions individually and determine their effectiveness over the short term. This approach demonstrates the potential for advances in the interdisciplinary field of developmental cognitive science when different methods, with complementary strengths and limitations, are brought to bear on the task of evaluating and improving pedagogical practices.

The home-based games that we developed and evaluated were not designed with the intention to replace the curriculum in school. Rather, they were designed to complement children's learning in school by engaging their school-relevant concepts in a fun manner at home. Below, I provide an overview of each empirical study in this dissertation.

#### **1.4.2 Overview of Chapter 2**

Children have intuitive numerical knowledge that is enabled by the core knowledge systems of number (ANS and PI) (see Carey, 2009; Dehaene, 1997; Feigenson et al., 2004; Spelke, 2011; Spelke 2022 for reviews). This intuitive knowledge does not depend upon knowing the symbols or language of mathematics. However, once children begin going to school, their thinking extends far beyond these intuitive abilities and depends upon the mastery of mathematical symbols and language. Does children's intuitive numerical knowledge causally support their learning of the symbolic numerical concepts taught in school?

Past work has shown that intuitive numerical knowledge is correlated with symbolic math skills (Halberda et al., 2008; Libertus et al., 2011). Moreover, a bi-directional relationship exists between intuitive and symbolic knowledge such that Amazonian adults who attended at least the first two grades of school and learned symbolic math had higher intuitive numerical skills than those who had not (Piazza et al., 2013). Beyond such correlational evidence, past work has shown that priming or briefly training intuitive numerical knowledge leads to improvements in symbolic skills in the short-term (Hyde et al., 2014; Khanum et al., 2016; Park et al., 2016; Park & Brannon, 2013, 2014). However, since most of this causal evidence comes from brief lab-based studies, it does not tell us whether the benefits of training intuitive abilities are sustained in the long-term, and if it can support the learning of school-based mathematics in children's noisy learning environments like their homes or schools.

In order to overcome the above limitations, Dillon et al., (2017) trained children's intuitive abilities over much longer periods (e.g., 17 weeks) in their schools and used a suite of games not only targeting children's numerical abilities but also their geometrical abilities. However, geometrical abilities are related to numerical abilities (Hawes & Ansari, 2020; Lauer & Lourenco, 2016; Mix et al., 2016; Mix & Cheng, 2012) and training children's geometrical abilities improves their numerical abilities (Cheng & Mix, 2014; Judd & Klingberg, 2021).

Therefore, based on a large-scale Randomized Controlled Trial (RCT) like Dillon et al., (2017) that train a host of mathematical skills (i.e., both numerical and geometry skills), it is not possible to establish the specific causal benefits of training intuitive numerical skills alone. Therefore, to address this limitation, we designed an experiment that trained first-graders' intuitive numerical abilities separately using games in their own homes for 2-3 weeks. We used a randomized controlled pretest-posttest design to compare the benefits of playing the number games

to playing geometry games or no games at home. We found that children who played the number games did not improve in their intuitive or symbolic numerical abilities in the short-term, both relative to children who played the geometry games or no games at game. However, children who played the geometry games improved in their geometry skills both relative to children who played the number games or no games at home.

Chapter 2 suggests that the sustained benefits seen in children’s intuitive mathematical abilities in Dillon et al., (2017) are likely driven by the intuitive geometry games and not the numerical games. Training children’s intuitive numerical skills might have limits towards sustainably improving their symbolic numerical skills. They might not work with children who are already in elementary school at the time of training (e.g., Bugden, Skudlarek, & Brannon, 2021) or when the duration of training is relatively short.

### **1.4.3 Overview of Chapter 3**

Given that we found no benefit of training children’s intuitive numerical skills towards improving their symbolic skills, in Chapter 3, we studied why symbolic numerical concepts are hard for children to learn. More specifically, we examined the difficulties (if any) that children face when trying to learn their language’s number words. Closely examining the sources of difficulty in children’s learning has the potential to inform future interventions aimed to improve their learning.

Number words in different languages vary in the degree of transparency by which they make the base-10 structure of the number system explicit. For example, in English, number words like “eleven” and “twenty” do not explicitly convey that they designate sets of “ten and one” and “two tens” respectively, whereas in Mandarin and other East-Asian languages (e.g., Japanese and Korean), they translate literally to “ten one” and “two tens”. Learning the meanings of multidigit

number words can pose challenges for children, especially when the rules for composing them are not explicit in their language's number words.

Most past studies that examined whether the degree of transparency of number words affects children's understanding of the base-10 compositional number system focused primarily on comparing transparent East-Asian languages with languages similar to English. They found that children who speak East-Asian languages outperform their age-matched English-speaking peers in tasks such as counting, understanding the logic of base-10 compositionality, and arithmetic (Bell, 1990; Fuson & Kwon, 1992; Ho & Fuson, 1998; Miller et al., 2005; Miller et al., 1995; Miller & Stigler, 1987; Miura et al., 1988; Miura et al., , 1993; Miura & Okamoto, 2003; Miura, et al., 1999; Song & Ginsburg, 1988; Towse & Saxton, 1998).

In our present study, we built on the prior literature in two important ways. First, we conducted a three-way comparison between children who spoke English, Mandarin, and Turkish. Including Turkish helped us examine whether the location of non-transparent number words in the count list affects the acquisition of base-10 compositionality. This is because Turkish contains transparent 'teens' (e.g., "eleven" translates to "ten one") and non-transparent decades as compared to English whose number words contain non-transparent teens (e.g., "eleven"). Second, we conducted a within country comparison between two languages, Tamil and Hindi spoken in India, that are extreme examples of transparency and non-transparency relatively. We studied children within India in order to examine the effects of number naming systems while trying to keep children's educational and SES backgrounds similar.

We found evidence that transparency of number words in different languages affects children's counting skills, but the acquisition base-10 compositionality is hard even when number

words are transparent in the child's language. This study suggests that all children would benefit from interventions that are aimed at improving their base-10 understanding.

#### **1.4.4 Overview of Chapter 4**

Given that base-10 compositionality takes years for children to learn even when their language's number words makes it more explicit, in Chapter 4, we evaluated an intervention that aimed to improve 5-6-year-old children's understanding of base-10 compositionality.

Our number game used three symbolic devices—grouped dots (by tens), a 10 x 5 spatial board, and transparent numerical language (e.g., “two tens and five ones”)—to convey the logic of the base-10 number system and enhance the learning of its words and symbols. We used a randomized controlled pretest-posttest design to compare the effect of this game, played at home, to the effect of home-based training of a game of similar difficulty that trained mental state reasoning. Children who played the number game for 2-3 weeks improved in their school-relevant numerical skills, relative to those who played the mental state reasoning game; improvements in mental state reasoning did not differ across the two conditions.

The number game in the present experiment was also evaluated in a large-scale field study along with three other mathematical games (Dean et al., in prep). The field study showed a positive impact on children's math knowledge, but it could not reveal whether all or only some of the four games contributed to this impact or what each game's specific impacts were. The present experiment therefore provides evidence that one of the four games used in the school-based math intervention was effective in enhancing children's understanding of the base-10 structure of the number system.

These findings suggest that a game training base-ten compositionality, played at home by children and their families or friends, can help foster children's numerical and arithmetic skills, at

least in the short term. This study, through its brief home-based intervention, was not designed to improve children's learning of math in the long term. However, it suggests that a game focused on base-ten compositionality would be worth including as a part of a large-scale intervention that uses a more comprehensive game-based curriculum to enhance learning in school.

#### **1.4.5 Overview of Chapter 5**

Building on the findings that our number game improved 5-6-year-old children's symbolic numerical concepts, we designed an improved version of the game to be administered to 6-7-year-old children who are in or are about to begin first grade in school. In the new version of the number game, we made several changes with the aim of increasing its effectiveness towards helping children learn the base-10 compositional number system.

First, we used all the two-digit numbers up to 99 instead of only using a subset of them as children might benefit getting exposed to a higher number of number words that reveal the logic of base-10 compositionality more transparently. Second, we changed the representation of the spatial board such that all the numbers within a decade (e.g., 20-29) appear in the same row, making it easier for children to analyze a double-digit number into its constituent tens and ones. Finally, we built in levels of the game that aim to teach children how to represent the same number in two different ways in order to aid faster mental calculation.

We used a randomized controlled pretest-posttest design, to compare the effect of the number game, played at home for 2-3 weeks, to the effect of home-based training of a game of similar difficulty that trained reading skills. The reading game focuses on the compositionality of the English alphabet and also uses a spatial board that systematically places rhyming words in the same column and words with the same beginning sound in the same row. In this intervention, both

games were designed to be more fun and engaging as they are two-player games with a winner that is determined by chance.

We found that children who played the number game for 2-3 weeks improved in their school-relevant numerical skills, relative to those who played the reading game; improvements in reading skills did not differ across the two conditions. This intervention took place entirely remotely and during disruptions to school-based learning due to the COVID-19 pandemic. Our experiment shows that making educational games available to families in their homes can help foster children's skills at the foundations of learning mathematics.

## Chapter 2

# Intuitive and Symbolic Numerical Concepts

Training children's intuitive mathematical knowledge improves their geometry but not numerical skills

## 2.1 Introduction

Much before children begin going to school, they have some knowledge about mathematics. They have numerical representations of large approximate quantities (e.g., the distinction between 20 and 30) and of small exact quantities (i.e., the distinctions between 1, 2, and 3) (see Carey, 2009; Dehaene, 1997; Feigenson et al., 2004; Spelke, 2011 for reviews). The system that represents approximate quantities, called the Approximate Number System (ANS), supports infants' ability to discriminate between sets (Brannon, 2002; Xu, 2003; Xu et al., 2005; Xu & Spelke, 2000) as well as young children's abilities to add, subtract, multiply and divide sets, as long as the ratios between the sets are sufficiently large (Barth et al., 2005; Barth et al., 2006; McCrink & Wynn, 2004; McCrink & Spelke, 2010). This means that while the ANS can support the discrimination between sets of 20 and 30, it cannot support the discrimination between sets of 20 and 21.

Children also have geometrical representations of large navigable layouts focused on the distances and directions of places in the layout, and of the shapes of small objects focused on their relative line lengths and angles (see Spelke, 2011 for a review). For example, they can distinguish small visual forms based upon geometrical properties like parallelism and right angles, even without knowing the names of shapes or their properties (Dehaene et al., 2006).

This kind of mathematical knowledge does not depend upon knowing the symbols or language of mathematics. However, once children start going to school, their thinking extends far beyond these intuitive abilities. For example, they learn to discriminate between large numbers exactly (i.e., they can tell 20 and 21 apart) by counting, and they perform exact arithmetic. They also learn the names of shapes, like “rectangle”, and learn about their properties, like sides and angles. All this knowledge depends upon children’s mastery of mathematical symbols and language. How does children’s intuitive mathematical knowledge relate to their learning of the symbolic mathematical concepts taught in school? In the current experiment we ask if training children’s early emerging, intuitive numerical and geometrical abilities improves, on the short term, their school-relevant performance in arithmetic and geometry. Training children’s intuitive mathematical abilities helps us understand whether and how they are causally related to symbolic mathematical concepts, and it has the potential to inform interventions aimed at improving their learning of symbolic mathematical concepts.

Children’s intuitive mathematical abilities are spontaneously developing and universal, that is, they do not depend upon any specific prior experience (Carey, 2009; Gordon, 2004; Pica et al., 2004; Spelke, 2011). One theory suggests that the ANS serves as a foundation for the development of symbolic concepts and supports symbolic mathematics throughout our lifespan (Dehaene, 1992). Three lines of evidence provide support for this theory and illustrate the relationship between ANS and symbolic mathematical concepts. First, the ANS has individual differences in its acuity (i.e., the precision with which one can discriminate between sets) and it correlates with individual differences in symbolic math achievement scores (Halberda et al., 2008; Libertus et al., 2011). Second, the ANS is trainable, that is, its acuity improve can improve with experience. For example, ANS acuity was found to be higher in Amazonian adults who attended

at least the first two grades of school than in those who had not (Piazza et al., 2013), illustrating the bi-directional relationship between ANS and symbolic mathematics. Third, going beyond correlational evidence, studies have found that briefly training the ANS has short-term benefits on symbolic mathematics (Hyde et al., 2014; Khanum et al., 2016; Park et al., 2016; Park & Brannon, 2013, 2014). It is unclear, however, whether these benefits are sustained and if they boost the learning of subsequent school-based mathematics. Interestingly, Hyde et al., (2014) found that briefly priming the ANS did not lead to improvements in ANS acuity itself despite leading to improvements in symbolic mathematics. Therefore, it is unclear if ANS is malleable by short-term priming or training. Crucially, since most of the training studies evidence is lab-based, it is unclear if the same mechanisms would hold in children's noisy environments where learning naturally unfolds, in their schools and homes.

In order to evaluate the impact of training children's intuitive abilities on their school-relevant learning in the long term, and in their noisy, real-world, learning environments like schools, Dillon et al., (2017) conducted an intervention among low-income preschoolers in Delhi, India. They randomly assigned preschool classes to receive four math games (two number and two geometry) that drew upon intuitive abilities, a matched set of four social-cognitive games, or no games. Children played the games in groups for a period of 17 weeks. Before the intervention and at three time points after it ended, children's mastery of both intuitive and symbolic math concepts was measured. At the first endpoint (0-3 months after the end of the preschool year), children who had played the intuitive math games showed greater mastery of both intuitive and symbolic math concepts, relative to both control conditions. At two other endpoints, (six months and one year later), the math-trained children's intuitive math concepts continued to exceed those of the children in the two other conditions, but they showed no greater mastery of the school math curriculum.

Thus, the math games had enduring effects on the preschool children's intuitive mathematical abilities, but not on their learning of symbolic mathematics.

Large-scale field experiments like these, conducted in preschool classes over a substantial part of the school year, can be of great value for evaluating programs to improve preschool children's learning. They also shed light on the activities by which children develop mathematical concepts and skills. Nevertheless, these studies have limitations that our present research aims to address. In order to make any games-based curriculum engaging and beneficial to children over many months of play, it is necessary to present children with multiple games that vary both in their content and in their difficulty. Because the impact of the intervention is evaluated as a whole, however, it is not possible to determine the specific effects of playing games in each of the two domains (number and geometry). For example, what was driving children's improved intuitive abilities, was it the number games or the geometry games or both? Indirect evidence bearing on these questions comes from children's performance on the outcome measures testing these abilities, and from process monitoring during the intervention, assessing children's attention to the games and success at playing them, but this evidence does not establish causal impacts of particular game experiences on children's math learning.

Moreover, geometrical and numerical abilities are related to each other. Geometrical abilities correlate with numerical abilities (Hawes & Ansari, 2020; Lauer & Lourenco, 2016; Mix et al., 2016; Mix & Cheng, 2012) and studies that train children's geometrical abilities have found improvements in their numerical abilities (Cheng & Mix, 2014; Judd & Klingberg, 2021). Therefore, in order to understand the differential effects of training children's geometrical and numerical skills, we must evaluate them separately.

In order to understand the specific effects of playing the number games and geometry games, we conducted a lab-based experiment evaluating a briefer, small-scale, home-based intervention using the versions of the games used in Dillon et al.'s (2017) field experiment. Although such experiments cannot reveal whether a given activity will be beneficial if played in school, they are a valuable complement to experiments in schools, in two ways. First, they serve as a prelude to school-based interventions: For example, if lab-based experiments were to establish that short-term play of four different games each led to enhanced learning of skills that schools aim to foster, then a school-based intervention using that suite of games could be worth implementing and evaluating. Second, they can serve as a follow-up to school-based interventions, shedding light on the aspects of the interventions that accounted for their success. The present research falls in the latter category.

The intuitive math games in the current study are similar to those used in Dillon et al., (2017) but scaled up in difficulty for 6-7-year-old children in the Greater Boston area, USA. One of the number games engaged children's ANS abilities and required them to add and compare large sets of dots approximately. The second numerical game engaged children's ability to represent small exact quantities (1-3). It required them to relate small numerical magnitudes to positions on a line, by establishing one-to-one correspondence relations between these small numerical magnitudes and sequences of movements on a linear board. These abilities also emerge in infancy (Feigenson & Carey, 2005) and training them improves children's symbolic numerical concepts in the short-term (Ramani & Siegler, 2008; Siegler & Ramani, 2009). One of the geometry games required children to find one shape that didn't belong with the rest based on a geometric property (e.g., shape, parallelism, connectedness). The second geometry game required children to place objects at locations in larger layouts based on small-scale geometric maps. Both

these geometrical abilities are early-emerging, universal, and might promote the learning of a variety of mathematical concepts (Dehaene et al., 2006; DeLoache, 1987; Huttenlocher et al., 1999; Spelke, 2011; Uttal et al., 2013; Winkler-Rhoades et al., 2013).

In the present study, we evaluated the benefits of playing the two intuitive number games and the two intuitive geometry games on children's intuitive and symbolic mathematical concepts. Children were quasi-randomly (equating for age, gender, and time during the school year) assigned to either play the number games, the geometry games, or no games at home. Children who received games played them at home for 2-3 weeks with their caregiver(s). All children completed a battery of assessments measuring their numerical, geometrical and executive functioning skills before and after the 2-3-week period. We hypothesized that children who played the number games will show a greater improvement in their numerical skills as compared to children who played the geometry games or no games at home. We also hypothesized that children who played the geometry games will show a greater improvement in their geometrical skills as compared to children who played the number games or no games at home.

## **2.2 Methods**

### **Participants**

Ninety-six children (48 females, mean age= 7.1 years, range= 6.2 to 7.9 years) were included in our final sample of participants. Participants were recruited from the greater Boston area by phone or email from a database of families at Harvard University who had expressed interest in participating in research studies. They received a small gift (e.g. a stuffed animal or a t-shirt) for participating. Our use of human subjects was approved by the Institutional Review Board at Harvard University. All participants were either enrolled in or had just finished first grade. An additional 7 children participated in the study but were excluded because they did not return for

the second of the study's two visits ( $n=6$ ) and were in Kindergarten ( $n=1$ ). Sample size and exclusion criteria were fixed before beginning data collection and were preregistered on the Open Science Framework. Children were assigned to one of three conditions: Number ( $n=32$ ); Geometry ( $n=32$ ); or No-treatment control ( $n=32$ ). This assignment was done in a manner that ensured that participant age was roughly equal across conditions and equal number of males and females were assigned to each of the conditions. Testing of children in the three conditions was interleaved so that children in all conditions were matched for the time of year of their participation, and therefore for the amount of schooling they had received at the time of testing.

## **Design**

This study used a pretest-posttest design with an intervention in between. It consisted of two 90-minute lab sessions separated by 2-3 weeks of home-based game play. Children were assigned to one of three conditions—Number, Geometry, or No-treatment Control—in a manner that ensured that participant age was roughly equal across conditions and that equal number of males and females were assigned to each condition. Assignments to condition were made by a Research Assistant who knew the ages and genders of the children previously tested in each condition but was unaware of the contents of the training/no training involved in specific conditions. Testing of children in the two conditions was interleaved so that children in both conditions were matched for the time of year of their participation.

In the lab sessions, all children completed two versions of the assessment battery (A and B) measuring their numerical, geometry, and executive function skills. The two versions of the assessments were matched in difficulty but consisted of different items. Half the participants in each treatment condition received version A at pretest and version B at posttest; the others received the two versions in the reverse sequence. Assessments were administered by an experimenter who

was naïve to the child’s assigned condition. In addition to completing the assessments, children also played one round of their assigned games in the lab, in the presence of an experimenter naïve to their performance on the assessment battery. Children were assigned to their game condition prior to the collection of any pretest data, by a Research Assistant who was not involved in the games training or in administering the assessments. All the methods and analyses were pre-registered unless otherwise specified. See <https://osf.io/knys4/> for the preregistration.

## Games

The Number condition consisted of training with two number games:

1. *Speed Sorting* (see Figure 2.1 (a)): This game aimed to exercise children’s representations of approximate quantities. It required children to add two dot arrays (in blue) and compare the sum of those arrays to a third array (in red) in an approximate manner. The dot arrays were displayed on cards and children sorted the cards into bins based upon the color of the right answer. It consisted of two levels with 4 decks in each level. Each deck had 24 cards.
2. *Race to the Top* (see Figure 2.1 (b)): This game aimed to exercise children’s representations of small exact quantities. It required children to establish one-to-one correspondence relations between one to four objects depicted on cards and relate it to one to four positions on a number line. It consisted of a spinner, which determined how many cards can be drawn from a deck in one turn. Children were required to move by the number of spaces as the objects on each of the cards that they drew. This game also consisted of two levels, with 24 cards in each level. The first level had a linear board that went from 1 to 20, and the second level had a linear board that went from 1 to 30.

The Geometry condition consisted of training with two geometry games:

1. *Odd One Out* (see Figure 2.1 (c)): This game aimed to exercise children’s sensitivity to shapes and their relative lengths and angles. It required children to determine which shape out of a group of six shapes did not belong with the rest based on a specific geometric property. All the shapes varied

in their sizes and orientations and they were depicted on cards. This game consisted of 4 decks with 24 cards in each deck.

2. *Reading Maps* (see Figure 2.1 (d)): This game aimed to exercise children's sensitivity to regularities in the navigable layout focusing on distances and directions of the places in the layout. It required children to place a token on a location on a large mat using information provided in small, purely geometric maps. All the maps were depicted on cards. This game consisted of two levels, where each level had its own large mat with different geometrical shapes on it. Each level consisted of 4 decks, with 24 cards in each deck.

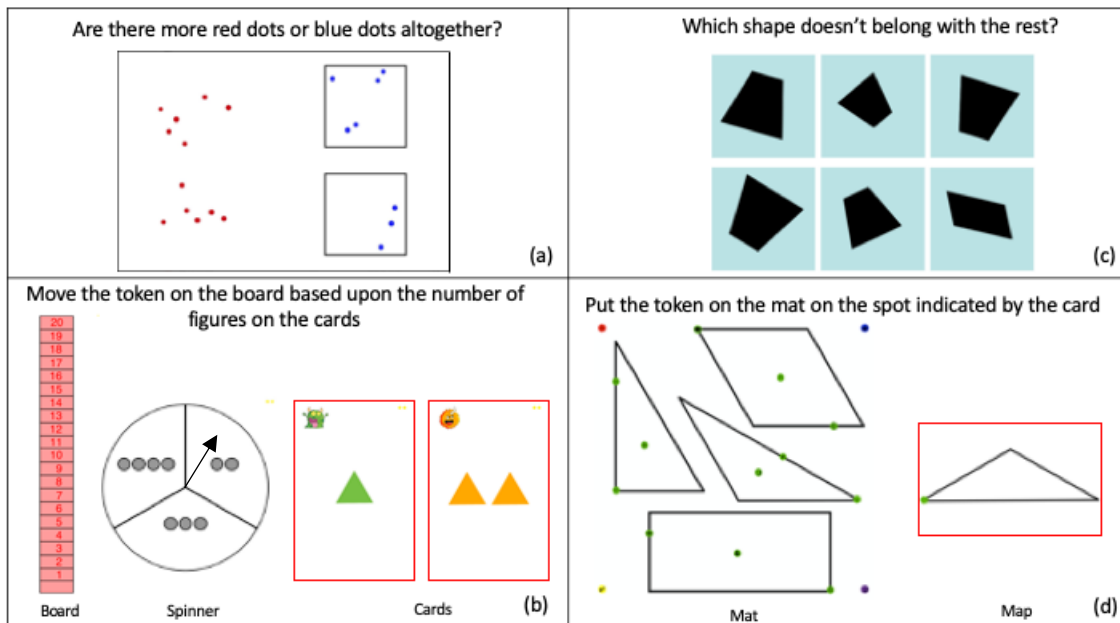


Figure 2.1. Number games: *Speed Sorting* (a) and *Race to the Top* (b). Geometry games: *Odd One Out* (c) and *Reading Maps* (d). In *Speed Sorting* (a) the card required children to compare the total number of red dots to the total number of blue dots and determined which is larger. Here, blue is the right answer, so the card would be sorted into a blue bin. In *Race to the Top* (b) children spun the spinner whose arrow indicated how many cards they could draw. They moved spaces on the board based upon the number of figures on the cards. If

they got green figure(s), they moved forward and if they got orange figure(s), they moved backwards. In *Odd One Out* (c) children were required to find the shape that didn't belong with the rest based upon a geometrical property. Here, the bottom right shape is the right answer. In *Reading Maps* (d) children were required to find the spot on the large map (left) that was indicated by the small map (right) on their card.

## **Assessments**

Two versions of assessment battery (A and B), each consisting of 7 tests were constructed. In each battery, 3 tests assessed children's numerical abilities, 3 tests assessed their geometric abilities and 1 test assessed their executive function skills. Half of the children in each condition completed the number assessments first, followed by the geometry assessments, and half of the children in each condition completed these sets of tests in the opposite order. The executive function test was always presented at the end.

The numerical tests were as follows:

1. Panamath (after Halberda et al., 2008), in which children were tasked with determining which of two dot arrays had more dots. This test measured the acuity of children's non-symbolic numerical discrimination.
2. Exact symbolic addition (after Hyde et al., 2014), in which children were presented symbolic addition problems with a pencil and paper. This test evaluated children's school-relevant performance of arithmetic.
3. Approximate symbolic addition (after Gilmore et al., 2007), in which children compared the sum of two rapidly presented Arabic numerals to a third numeral and judged which is larger. This test evaluated children's intuitive sense of the quantities represented by Arabic numerals as well as their ability to add and compare numbers approximately.

The geometry tests were as follows:

1. Geometric intruder (after Dehaene et al., 2006), in which children were tasked with determining which one of 6 shapes did not belong with the rest based on the one geometric property by which it differed (all shapes varied in size and orientation). This test measured children's intuitive judgments about the properties of 2D visual forms.
2. Shape knowledge (after Dillon et al., 2017), in which children were asked questions about the properties of shapes, e.g., "which shape has sides that are all the same length" and "which corner in a given shape is the biggest". This test measured children's school-relevant knowledge about shapes and their properties.
3. Geometric intuitions (Izard et al., 2011), in which children were asked questions about the properties of points and lines on the Euclidean plane. This test measured children's intuitions about abstract geometric objects.

The executive function test (after Dillon et al., 2017) required children to press keys either on the same side or the opposite side of characters presented on a screen. The test had four blocks, where in the first block, children had to press a key on the same side as a character. In the second block, they had to press a key on the opposite side as a character. In the third and fourth block, they had to switch between same side or opposite side keys based upon the character that appeared on the screen.

## **Procedure**

During the first lab session, an experimenter unaware of the child's condition administered one version (either A or B) of an assessment battery as a pretest. Following this, a second experimenter, unaware of the child's performance on the pretest, taught the child and their caregiver how to play their assigned game using practice decks of either the number games or the geometry games. Once they understood how to play their game, they were encouraged to play

independently with each other in the presence of the experimenter. At the end of the first session, only children and their caregivers in the Number and Geometry conditions were provided with their complete set of game materials to play at home during a 2-3-week period. They were asked to play the game on 8 separate days during the home-based training period either by themselves or by involving a sibling. In addition to the games, caregivers were provided with (i) an informational brochure containing the description of the games and instructions, (ii) experimenter contact information, and (iii) a website providing background information about the games, suggested feedback for gameplay, and a way to record performance.

During the home-based training period, caregivers in the training conditions were asked to fill out a brief online survey to indicate, for each game, which deck of the game they played, what time of the day they played, who played with the child, how many responses were correct, and how long it took. During the second lab session, an experimenter unaware of the child's condition administered a different version of the same assessment battery as a posttest. Finally, children and their caregivers from all conditions played one last level of their assigned games (same as pretest) with their caregiver in presence of the experimenter who had taught them their game. All the in-lab sessions were video recorded.

## **2.3 Results**

In order to determine the effectiveness of playing the number games and the geometry games, we pre-registered two primary outcome measures: a number composite score and a geometry composite score. The number composite score was calculated by averaging the Z-scores of the percent correct of the 3 numerical tests (Panamath, Exact symbolic addition, and Approximate symbolic addition). The geometry composite score was calculated by averaging the

Z-scores of the percent correct of the three geometrical tests (Geometric intruder, Shape knowledge, and Geometric intuitions). All Z-scores were calculated by standardizing to the pretest scores of the no-treatment control group. All the primary outcomes measures were analyzed using the following pre-registered regression framework:

$$y_i = \text{Number}_i + \text{Geometry}_i + \text{pretest}_i + \varepsilon_i \quad (1)$$

Where  $y_i$  represents the posttest score of a particular outcome for child  $i$ ,  $\text{Number}_i$  is an indicator variable for whether child  $i$  was treated with the number games,  $\text{Geometry}_i$  is an indicator variable for whether child  $i$  was treated with the geometry games, and  $\text{pretest}_i$  is the pretest score of an outcome  $y_i$  for child  $i$ .

We first tested whether children who received the number games had higher scores on the number tests at posttest than those in the no-treatment control condition. We regressed the number composite score using (1) and found that playing the number games led to no significant increase (-0.16 standard deviations) in the number scores relative to the no-treatment control condition [ $t(91) = -1.58$ ,  $p = 0.12$ ]. There was also no significant effect of playing the geometry games on children's number composite scores (see Figure 2.2 and Table 2.1). Therefore, playing the number games had no effect on children's numerical skills.

We then tested if children who received the geometry games had higher scores on the geometry tests at posttest than those who had not. We regressed the geometry composite score using (1) and found that playing the geometry games led to a significant increase in the geometry scores: 0.45 standard deviations relative to the no-treatment control condition [ $t(91) = 3.05$ ,  $p < 0.01$ ]. There was no significant effect of playing number games on children's geometry composite scores (see Figure 2.3 and Table 2.1). Therefore, only playing the geometry games had a positive effect on children's geometrical skills.

In order to determine whether the number and geometry games led to an improvement in children's symbolic skills, we also pre-registered two other primary outcome measures: a symbolic number composite score and a symbolic geometry composite score. The symbolic number composite was calculated by averaging the Z-scores of the percent correct of 2 symbolic number tests (Exact symbolic addition and Approximate symbolic addition). We regressed the symbolic number composite score using (1) and found that playing the number games led to no significant increase (-0.14 standard deviations) in the symbolic number composite scores relative to the no-treatment control condition [ $t(91) = -1.23$ ,  $p = 0.22$ ]. There was also no significant effect of playing the geometry games on children's symbolic number composite scores (see Table 2.1). Therefore, playing the number games had no effect on children's symbolic number skills.

The symbolic geometry composite score was calculated by averaging the Z-scores of the percent correct of the two symbolic geometry tests (Shape knowledge and Geometric intuitions). We regressed the symbolic geometry composite score using (1) and found that playing the geometry games led to no significant increase (0.23 standard deviations) in the symbolic geometry composite scores relative to the no-treatment control condition [ $t(91) = 1.18$ ,  $p = 0.24$ ]. There was also no significant effect of playing the number games on children's symbolic geometry composite scores (see Table 2.1). Therefore, playing the geometry games had no effect on children's symbolic geometrical skills when analyzed as a whole.

In order to determine whether the number and geometry games led to an improvement in children's non-symbolic/intuitive skills, we regressed our two non-symbolic measures (Z-scores of the percent correct of Panamath and Geometric intruder) separately using (1) and found that playing the number games led to no significant increase (-0.15 standard deviations) in the Panamath scores [ $t(90) = -0.83$ ,  $p = 0.41$ ] relative to the no-treatment control group. There was also

no significant effect of playing the geometry games on children's Panamath scores (see Table 2.1). Therefore, playing the number games had no effect on children's intuitive number skills. However, playing the geometry games led to a significant increase in the Geometric intruder scores: 0.86 standard deviations relative to the no-treatment control condition [ $t(91) = 4.82$ ,  $p < 0.001$ ]. There was no significant effect of playing the number games on the Geometric intruder scores (see Table 2.1). Therefore, playing the geometry games had a positive effect on children's intuitive geometrical skills.

We also regressed each of our symbolic measures (Z-scores of the percent correct of Exact symbolic addition, Approximate symbolic addition, Shape knowledge and Geometric intuitions) separately using (1). We found that playing the number games led to no significant increase in the Exact symbolic addition (-0.06 standard deviations) scores [ $t(92) = 0.34$ ,  $p = 0.73$ ] or the Approximate symbolic addition (-0.29 standard deviations) scores [ $t(91) = -1.72$ ,  $p = 0.09$ ] relative to the no-treatment control condition. There was also no significant effect of playing the geometry games on each of the symbolic number tests (see Table 2.1). Therefore, playing the number games had no effect on children's symbolic numerical skills even when analyzed separately.

We found that playing the geometry games led to a significant increase in the Shape knowledge scores: 0.39 standard deviations relative to the no-treatment control condition [ $t(92) = 2.01$ ,  $p < 0.05$ ]. However, playing the geometry games led to no significant increase (0.1 standard deviations) in the Geometric intuitions scores [ $t(91) = 0.06$ ,  $p = 0.95$ ] relative to the no-treatment control group. There was no significant effect of playing the number games on either of the symbolic geometry scores (see Table 2.1). Therefore, playing the geometry games had a positive effect on one of the symbolic geometry skills which measured children's explicit knowledge about

shapes and their properties. However, playing the geometry games had no effect on the second symbolic geometry test which measured children's explicit understanding of Euclidean geometry.

Finally, we regressed the executive function scores using (1). The executive function scores are the Z-scores of the average of 1) difference between the percentage of correct responding between blocks 1 and 2 and 2) difference in the percentage of correct responding between blocks 3 and 4 and 1 and 2. We found that neither the number [ $t(89) = 0.20$ ,  $p = 0.84$ ] nor the geometry [ $t(89) = -0.35$ ,  $p = 0.73$ ] games led to a significant increase on this measure as compared to the no-treatment control group (see Table 2.1). Therefore, games training had no effect on children's executive function skills.

Table 2.1. Coefficients from a linear regression model controlling for pretest scores for each of the primary outcomes. The number and geometry conditions are compared to the no-treatment control group respectively. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

|                               | Number | Geometry |
|-------------------------------|--------|----------|
| Number composite              | -0.16  | -0.03    |
| Geometry composite            | -0.003 | 0.45**   |
| Symbolic number composite     | -0.14  | 0.07     |
| Symbolic geometry composite   | 0.05   | 0.23     |
| Panamath                      | -0.15  | -0.18    |
| Geometric intruder            | -0.09  | 0.86***  |
| Exact symbolic addition       | 0.06   | 0.06     |
| Approximate symbolic addition | -0.29  | 0.04     |
| Shape knowledge               | 0.06   | 0.39*    |
| Geometric intuitions          | 0.1    | 0.02     |
| Executive function            | 0.04   | -0.06    |

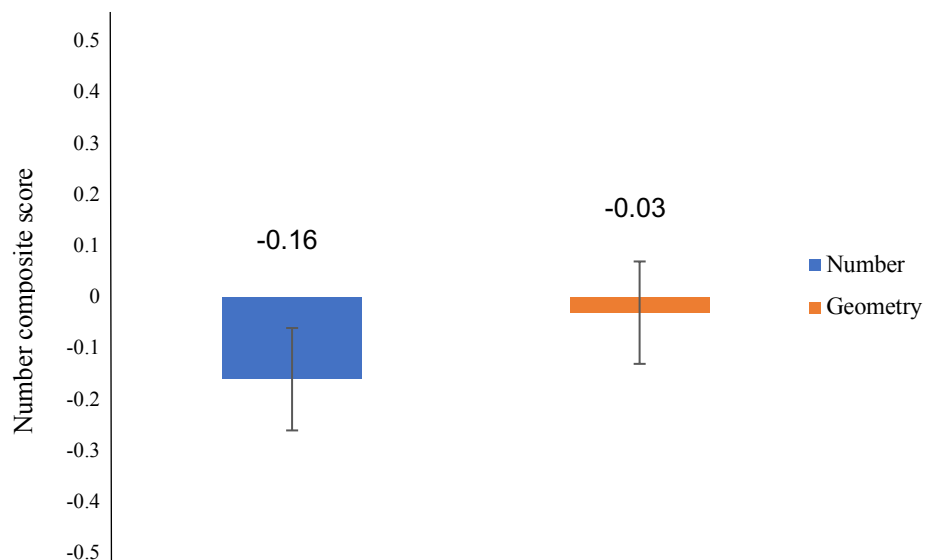


Figure 2.2. The colored bars show the impact of the number games and the geometry games on the number composite score (y-axis). Errors bars represent standard errors of the mean. Coefficient estimates indicate differences from the no-treatment control.

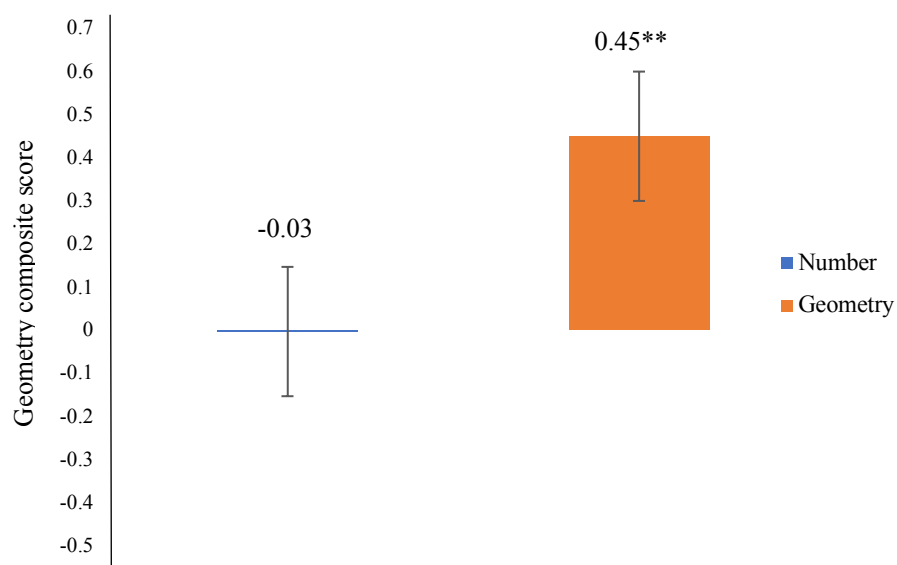


Figure 2.3. The colored bars show the impact of the number games and the geometry games on the geometry composite score (y-axis). Error bars represent standard errors of the mean. Coefficient estimates indicate differences from the no-treatment control. Asterisks beside the co-

efficients indicate a rejection of the null hypothesis of no difference compared to the no-treatment control group.  $**p<0.01$

In order to examine if there were any differences in the duration or frequency of playing the number and geometry games, we looked at the data from the caregiver report of gameplay at home. The average time spent playing Speed Sorting was the lowest, bringing down the average time of the number games. However, the number games were played more frequently than the geometry games (See Table 2.2).

Table 2.2. Caregiver report of children's gameplay at home in terms of duration and frequency of play

| Game            | Average time (in mins) per game | Total time (by condition) | Frequency of play per game | Total frequency (by condition) |
|-----------------|---------------------------------|---------------------------|----------------------------|--------------------------------|
| Odd One Out     | 6.07                            | 13.67                     | 253                        | 452                            |
| Reading Maps    | 7.6                             |                           | 199                        |                                |
| Speed Sorting   | 3.56                            | 9.76                      | 290                        | 503                            |
| Race To The Top | 6.2                             |                           | 215                        |                                |

## 2.4 Discussion

In this study, we examined if children's intuitive mathematical knowledge is casually related to their knowledge of symbolic mathematics taught in school. Towards this end, we evaluated the benefits of playing two intuitive number games and two intuitive geometry games at home for 2-3 weeks on children's mathematical skills. We found that children who played the intuitive number games did not improve in their intuitive or symbolic numerical skills both as compared to children who played the geometry games and to children who played no games at home. Whereas, children who played intuitive geometry games improved in their intuitive

geometrical skills but not in their symbolic geometrical skills taken together, both as compared to children who played the number games and to children who played no games at home.

Children who played the geometry games significantly improved in only one out of the two symbolic geometrical tests. This test measured their knowledge about shapes names and their properties. Children might have improved in this test because the geometry game called Odd One Out, that required children to find the shape that doesn't belong with the rest, might have created opportunities for them to talk about shapes explicitly. Though not required of them, it might have prompted children and their families to discuss why a particular shape was chosen as the answer, which in turn could have led to them explicitly talking about shapes using their names like "rectangle" and their properties like "it has four sides, two long and two short". Children who played the geometry games did not show any improvement on the test that measured their verbal knowledge about points and lines in a Euclidean plane as it was further away from the displays in any of the geometry games that they played.

None of the children showed any significant improvement in their executive function skills. Therefore, the improvements of the kids in the geometry condition on their geometrical skills aren't attributable to improvements in executive function skills. Moreover, children did not show any benefits in the domain that they weren't trained in. That is, children who played the number games did not show any improvement on the geometry measures and children who played the geometry games did not show any improvements in number measures. Therefore, the improvements of the children in the geometry condition are specific to the geometry games.

Despite playing the number games a little more frequently than the geometry games at home, why did children who played the number games show no improvement in either their intuitive or symbolic numerical skills? One possible reason for this might be due to the age of

children during training. Children in this study were already in first grade in contrast to Dillon et al.'s (2017) experiment with children who at the time of training were in preschool. More recently, Bugden, Skudlarek, & Brannon (2021) also found no benefit of training the ANS towards improving third and fourth grade children's symbolic abilities.

Another possible reason for the lack of benefits of training intuitive numerical skills might be due to the duration of the intervention. Children in our study had access to the games only for 2-3 weeks where as children in Dillon et al.'s, (2017) experiment played the game for 17 weeks. Training intuitive numerical concepts might be more beneficial when children are younger, and the training period is much longer. Our results stand in contrast to brief, lab-based, training studies that find benefits of ANS training on symbolic abilities (e.g., Hyde et al., 2014; Park et al., 2016; Park & Brannon, 2013, 2014). The mechanisms by which symbolic concepts improve after brief exposure to the ANS are yet to be fully understood. Our results are in line with failures to replicate the benefit of ANS training on symbolic math in school-aged children (Bugden et al., 2021) and adults (Szkudlarek et al., 2021).

Finally, this short home-based intervention serves as a complement to a large Randomized Controlled Trial (RCT) conducted by Dillon et al., (2017). It sheds light on the aspects of the intervention that might have accounted for their success. For example, the long-term improvement seen in the math trained children's intuitive abilities in Dillon et al., (2017), where children played both intuitive number and geometry games, might be due to the intuitive geometry games. Improving children's symbolic numerical skills might require something more. That is, the connections between the language and symbols of mathematics on the one hand, and intuitive mathematical knowledge on the other, might need to be made explicit to children.

## Chapter 3

# Language and Numerical Concepts

Does Language Shape Children's Understanding of the Base-10 Compositional Number System?

### 3.1 Introduction

In order to learn math efficiently in school, children need to master the base-10 compositional structure of the number system (Chan et al., 2014; Laski et al., 2014; Moeller et al., 2011; Ross, 1986), and the number words and symbols that represent this number system. That is, they need to represent multidigit numbers as quantities composed of sets generated by different powers of ten (e.g., “fifty-seven” is composed of “five tens” and “seven ones”). However, learning the meanings of multidigit number words can pose challenges for children, especially when the rules for composing them are not explicit in their language's number words.

Number words in different languages vary in the degree of transparency by which they make the base-10 structure of the number system explicit. For example, in English, number words like “eleven” and “twenty” do not explicitly convey that they designate sets of “ten and one” and “two tens” respectively, whereas in Mandarin and other East-Asian languages (e.g., Japanese and Korean), they translate literally to “ten one” and “two tens”. In the present study we examined whether the degree of transparency of number words across five different languages affects children's understanding of the base-10 compositional number system.

Most past studies primarily focused on comparing transparent East-Asian languages with languages similar to English that are relatively less transparent. In the present study, we built on the prior literature in two ways. First, we conducted a three-way comparison between children who

speak English, Mandarin, and Turkish. Including Turkish enabled us to examine whether the location of non-transparent number words in the count list affects the acquisition of base-10 compositionality. This is because Turkish contains transparent ‘teens’ (e.g., “eleven” translates to “ten one”) and non-transparent decades as compared to English, whose number words contain non-transparent teens (e.g., “eleven”).

Second, we conducted a within country comparison between two languages, Hindi and Tamil, spoken in India, that are extreme examples of non-transparency and transparency respectively. We studied children within India in order to examine the effects of number naming systems while trying to keep children’s educational and SES backgrounds similar. Therefore, in the present study, we reexamined the influence of different number naming systems on children’s base-10 understanding by expanding the languages studied as well as studying the effects of the position of the non-transparent number words in the count list (i.e., teens versus decades).

### **Effects of different number naming systems on children’s counting skills and base-10 understanding**

Past work has shown that English-speaking children fall behind their age-matched peers who speak languages like Mandarin in how high they can count (Miller et al., 1995; Miller & Stigler, 1987). For example, Miller et al., (1995) found that Mandarin-speaking 4-and 5-year-olds count much higher than their age-matched English-speaking peers. Critically, there were no differences in the percentages of Mandarin- and- English-speaking children who could count up to 10, as the number words in all languages from one to ten need to be rote-learned. The relatively large difference, favoring Mandarin-speaking children, emerged in the percentage of children who were able to count in the teens range (eleven to nineteen). This suggests that the irregularity of the English numbers in the teens might be playing a role. This difference in counting ability between

children who speak more transparent languages and those who speak relatively less transparent languages has been replicated in more recent studies and across a variety of languages (Cankaya et al., 2014; Guerrero et al., 2020; Lê & Noël, 2020; Schneider et al., 2020).

Do differences in counting ability lead to any differences in base-10 understanding? One of the most common tasks that has been used to elicit children's understanding of base-10 compositionality is called the blocks task (e.g., Miura, Kim, Chang, & Okamoto, 1988; Miura & Okamoto, 1989). In this task, children are shown small cubes that represent the numbers from one to nine, and bars that are composed of ten cubes to represent the number ten. Then they are asked to generate multidigit numbers (e.g., twenty-five) using such base-10 blocks. First-grade aged children who speak relatively transparent languages (e.g., Mandarin, Korean and Japanese) are more likely to use sets of tens and units of ones to generate multidigit numbers as compared to children who speak relatively less transparent languages (e.g., English, French, and Swedish) (Miura, 1987; Miura et al., 1988, Miura & Okamoto, 1989; Miura et al., 1993; Miura et al., 1994; Ho & Fuson 1998). For example, while Japanese-speaking children in these studies were more likely to use two bars of ten and five cubes of ones to generate twenty-five, English-speaking children were more likely to use twenty-five cubes of ones.

The difficulties that English-speaking children face when using base-10 blocks to represent multidigit numbers does not seem to come from a lack of understanding of the task. For example, Towse & Saxton (1997), in one condition, provided English-speaking children with a twenty bar by joining two bars of ten. When such a block was available, English-speaking used it to generate numbers in the twenties like twenty-five by composing the twenty bar and five cubes of ones. Therefore, they successfully decomposed twenty-five into “twenty and five” but failed to

decompose twenty into “two tens” and the teens into tens and ones, highlighting the influence of the non-transparent number naming system in English.

In addition to the blocks task, in another study, it was found that 5-year-old Mandarin-speaking children, who counted at least until 50, outperformed their age-matched English-speaking peers in representing teen quantities as tens and ones in an oral addition task (Ho & Fuson, 1998). For example, when told that there are ten objects hidden in a box to which ‘n’ (where n was a single-digit number) objects are being added, 5-year-old Mandarin-speaking children were more likely to correctly answer the question of how many objects are in the box altogether as compared to English-speaking children.

Together, these studies suggest that there is an effect of different number naming systems on children’s representations of multidigit numbers as composed of tens and ones. That is, children who speak transparent languages begin acquiring this base-10 understanding earlier than those who speak relatively non-transparent languages, thus lending support to the language hypothesis.

### **Challenges to the language hypothesis**

When comparing English-speaking children to those who speak East-Asian languages, cultural and educational practices covary with linguistic differences (Fuson & Li, 2009; Newman et al., 2007; Perry, 2000; Schleppenbach, et al., 2007; Stigler & Stevenson, 1992) and that might also lead to differences in understanding of base-10 compositionality. For example, elementary schools in China predominantly teach base-10 decomposition strategies during arithmetic instruction, whereas in American curricula, while base-10 decomposition is introduced, it is one of the many available strategies. Therefore English-speaking children in the US might have less practice with base-10 compositionality (Fuson & Li, 2009).

In an attempt to disentangle the effects of formal instruction and language, some past studies administered the blocks task with Kindergarten-aged children and found evidence against the language hypothesis (Vasilyeva et al., 2015, 2016; Wang et al., 2008). Kindergarteners, though fluent in their language, have fairly limited exposure to formal base-10 or arithmetic instruction as compared to first-graders. Therefore, Vasilyeva et al., (2015) administered the blocks task to Kindergarteners who spoke more transparent languages (Mandarin and Korean) and to those who spoke relatively less transparent languages (English and Russian).

Critically, they divided all children into two groups. In one group, children received practice trials with single digit-numbers. In this group, the experimenter showed children how to “make” the number seven, for example, with seven individual cubes. In the other group, children received mixed-digit practice trials. In this group, in addition to seven, the experimenter also showed children how to “make” the number fourteen with one bar of ten and four individual cubes. After these different types of practice trials between groups, all children completed trials of the blocks task where they had to generate double-digit numbers on their own. Among Kindergarteners, they found that all children, irrespective of their language, performed better if they received the mixed-digit practice trials (68% accuracy) as compared to the single-digit practice trials (39% accuracy). Therefore, they found an effect of the type of practice trial but not of language, replicating past work (Towse & Saxton, 1997). The relative degree of transparency of number words had no effect on children’s base-10 representations in this study.

If a single practice trial demonstrating how two-digit numbers can be generated using sets of tens and ones is sufficient in enabling all children, irrespective of their language, to generate a similar proportion of base-10 representations, then language might not be playing a substantial role in developing this concept as compared to other factors such as formal instruction. The

language advantage in the blocks task reported among first-graders (e.g., Miura et al., 1994) in past work might be due to more experience with formal instruction.

Similarly, in a study of Japanese and Singaporean Kindergartners, both groups who speak similarly transparent languages, it was found that Japanese children outperformed Singaporean children in oral and written addition problems with numbers less than 100. A higher proportion of Japanese children used base-10 strategies to perform arithmetic, and this likely reflects differences in children's experiences that are not related to their language's number naming systems (Marcruz et al., 2022).

Together, these set of studies that elicited base-10 understanding among Kindergarteners pose a challenge to the language hypothesis that focuses solely on the effects of different number naming systems on children's developing base-10 concepts. They suggest that factors that covary with language, such as differences in formal instruction, may also influence children's representations of multidigit numbers.

### **Re-examining the language hypothesis: the present study**

In order to reexamine the effects of different number naming systems on children's understanding of base-10 compositionality, we studied 4-8-year-old children who speak five languages namely English, Mandarin, Turkish, Hindi, and Tamil. These languages vary in how transparently their number words make the base-10 structure of the number system explicit, with Mandarin being the most transparent and Hindi being the least transparent. See Table 3.1 for an illustration using examples of a few number words. By recruiting children spanning a wide age range, we aimed to examine when the differences in the understanding of base-10 compositionality emerge (i.e., do they emerge only when formal base-10 and arithmetic instruction in schools begin).

Table 3.1. Examples of number words in five languages

|          | 2     | 5      | 10     | 12          | 20         | 25              | 52              |
|----------|-------|--------|--------|-------------|------------|-----------------|-----------------|
| English  | two   | five   | ten    | twelve      | twenty     | twenty-five     | fifty-two       |
| Mandarin | er    | wu     | shi    | shi er      | er shi     | er shi wu       | wu shi er       |
| Turkish  | iki   | beş    | on     | on iki      | yirmi      | yirmi bes       | elli iki        |
| Hindi    | do    | paanch | dus    | baarah      | bees       | pacchees        | baawan          |
| Tamil    | rendu | anju   | patthu | panni-rendu | iru-patthu | iru-patthu-anju | an-patthu-rendu |

### Effects of the positions of non-transparent number words in the count list

We selected languages that not only varied in terms of how many non-transparent number words they contain, but also in terms of where in the count list those number words occur. We included Turkish as it has transparent teens, just like Mandarin (i.e., eleven translates to “ten one”, twelve to “ten two” and so on) but it has non-transparent decades. That is, the number words for twenty, thirty, forty, and fifty have no similarity to the number words from one through five or to the number word for ten (see Table 3.1 for the examples of “twenty” and “fifty-two”). For the decades beyond fifty, while the number words have some phonological similarity to the single-digit number words from which they are composed, they still have no cues to their composition with the number word for ten. Therefore, for the decades, Turkish number words obscure the base-10 compositional number system. We compared Turkish-speaking children with English-speaking children, as English has irregular teens (i.e., eleven, twelve, and so on) and semi-regular decades, once children are able to figure out that “ty” in a decade like “twenty” refers to ten.

Comparisons of Turkish and English help us understand the effects (if any) of the position of the non-transparent number words on children’s understanding of base-10 compositionality. If earliest occurring multidigit number words, like eleven and twelve, are more crucial than the

decades for children to gain the insight of base-10 compositionality, then Turkish-speaking children should outperform English-speaking children in their counting skills and base-10 understanding. However, if the decades pose a bigger impediment to base-10 understanding, then English-speaking children should outperform Turkish-speaking children.

Very few past studies have probed Turkish-speaking children's numerical understanding. One study found that 3-and-4-year-old Turkish speaking children learn to count from 1-20 at a faster pace as compared to their age-matched English-speaking peers in the context of a game. However, this advantage did not translate into a better understanding of numerical concepts such as cardinality (Cankaya et al., 2014). Another study found that Turkish-speaking children acquire the knowledge of the cardinal principle (i.e., the concept that the last word used to count a set of objects refers to the cardinal value of the object) later than their English-speaking peers (Ceylan & Aslan, 2018).

Crucially, one study administered the blocks task in first-grade aged Turkish-speaking children and compared the results to Miura et al.'s (1994) previously reported findings with children speaking less transparent Western languages (English, French and Swedish) and more transparent East-Asian (Chinese, Japanese, and Korean) languages (Olkun et al., 2011). This study found that, across trials, Turkish-speaking children generated a lower percentage of base-10 representations in the blocks task as compared to children who spoke East-Asian languages. However, when Turkish-speaking children were compared with the children who spoke Western European languages, there was no consistent difference observed. The Turkish speaking children outperformed only the English-speaking children in Miura et al.'s, (1994) sample and performed similarly to the children who spoke French and Swedish. However, the Turkish-speaking children in this study were being compared to the English-speaking children recruited in Miura et al.'s,

(1994) study almost two decades ago. There may be generational shifts in English-speaking children's performance and one study found evidence supporting this hypothesis (Vasilyeva et al., 2016). Therefore, to our knowledge, our study is the first to directly compare Turkish-and-English-speaking children's base-10 understanding by sampling children across a wide age range simultaneously.

We also compared Mandarin-speaking children to English-speaking children in our study as it offers a perfectly transparent number word system to compare the performance of English-speaking children against. Since a majority of the past research focused on the contrast between East-Asian languages and English, it also helps us contextualize our findings with respect to previous findings.

### **Effects of a highly non-transparent language from a multilingual country**

In a second set of comparisons, we included Hindi and Tamil, both languages that are spoken in India. The reasoning behind studying children from multilingual countries is that those who speak relatively transparent or non-transparent languages are likely to be in schools that follow similar educational practices as compared to children who are from different countries. While Tamil is almost as transparent as Mandarin, Hindi is on the other end of the spectrum with a majority of the number words from 11-99 being non-transparent, such that they do not conform to a predictable pattern and would thus need to be rote-learned.

We aimed to compare Hindi-speaking and Tamil-speaking children who belong to similar SES backgrounds and attend schools that follow similar educational practices. Due to the prevalence of English-medium education in India, children who attend Tamil and Hindi medium schools typically come from low SES households and are likely to be first-generation school goers. By studying children from similar backgrounds and comparing their developing base-10

understanding, we attempted to account for the factors (e.g., SES and formal instruction) that tend to co-vary with language. If the transparency of number words has an effect on children's understanding of base-10 compositionality, even after accounting for SES and educational practices, then Tamil-speaking children should outperform their Hindi-speaking peers on tasks probing base-10 understanding.

While there are no studies directly comparing highly non-transparent languages like Hindi to transparent languages like Tamil, previous studies from multilingual regions have found mixed results. One previous study with older children (6-8-year-olds) from a multilingual region (Wales) found that children who spoke a transparent language (Welsh) outperformed those who spoke a relatively non-transparent language (English) in some numerical tasks relevant to base-10 knowledge (e.g., two-digit numerical comparison) but not others (e.g., arithmetic) (Dowker et al., 2008). Another study found effects of language, favoring children who spoke a transparent language (Chinese) as compared to those who spoke a relatively non-transparent language (Malay) in Malaysia, on a host of numerical tasks (e.g., counting and making numbers using smaller units of one, five, and ten) even after controlling for parental education and SES (Kheu and Nunes, 2017).

Lastly, one study compared Hindi-speaking children attending Hindi-medium schools to children attending English-medium schools, both within India. They found a difference in counting skills, favoring the children attending the English-medium schools (Schneider et al., 2020). However, children attending English-medium schools in India are more likely to be from middle and high SES backgrounds as compared to the children attending Hindi-medium schools who are typically from low SES backgrounds. Therefore, it is not possible to rule out the effects of other factors that covary with linguistic differences based upon this study. Moreover, the study above

did not measure children's base-10 understanding. Therefore, to our knowledge, our study is the first to compare base-10 understanding among children who speak a highly non-transparent language like Hindi to those who speak a transparent language like Tamil, within the same country.

### **Overview of the present study**

This study is a part of a larger project that aimed to examine the effects of number naming systems in different languages on a host of children's developing numerical capacities. In this paper, we will primarily focus on the findings from a task measuring base-10 understanding. We tested 4-7-year-old children who are monolingual speakers of English, Mandarin, or Turkish in Study 1 and tested 4-8-year-old children who are monolingual speakers of Hindi or Tamil in Study 2. In addition to the primary base-10 task, all children completed a verbal counting task, a task measuring their knowledge of the meanings of small number words (Give-N task, after Wynn, 1992), a task measuring their understanding that consecutive numbers in the count list vary by one (Unit task, after Sarnecka & Carey, 2008) and a task measuring their understanding that number words refer to exact cardinal values (Exact Equality or Logic task, after Lipton & Spelke, 2006). We will only report data from the primary base-10 task, verbal counting task and Logic task in this paper.

We designed a new base-10 task that was amenable to online testing based on Miura's (1987) base-10 blocks task. In our task, children reasoned about sets of tens and ones by comparing quantities represented both as sets of tens and ones, counting quantities represented both as sets of tens and ones, and decomposing multidigit numbers into their respective sets of tens and ones.

## **3.2 Study 1**

### **3.2.1 Methods**

#### **Participants**

A total of 287 children who were monolingual speakers of three languages (English= 96, Mandarin= 98, and Turkish= 93) were included in the final sample of participants. We aimed to recruit 124 3-7-year-olds in each of the three languages but excluded children who did not meet our preregistered inclusion criteria and 3-year-olds in this paper, as the primary base-10 task reported in this paper was optional for them (see [https://osf.io/zr5um/?view\\_only=1579dc9fb8874d8db46f52da49507abe](https://osf.io/zr5um/?view_only=1579dc9fb8874d8db46f52da49507abe) for the preregistration). We recruited English-speaking children from the US, Mandarin-speaking children from China, and Turkish-speaking children from Turkey. English-speaking children were recruited via a database of families who agreed to participate in our research, social media, or Children Helping Science. Mandarin-speaking children were recruited via a database of families who agreed to participate in our research and via social media. Turkish-speaking children were primarily recruited via social media and word-of-mouth.

Monolingual speakers were defined as children whose exposure to the target language made up at least 90% of their total language input, as reported by their caregiver. Each participant received a small gift (e.g., an Amazon gift card or a book) of a similar value across language sites. The sample size, exclusion criteria, and methods were preregistered prior to collecting any data. The analysis plan was preregistered prior to any examination of the data by the researchers (see [https://osf.io/zr5um/?view\\_only=1579dc9fb8874d8db46f52da49507abe](https://osf.io/zr5um/?view_only=1579dc9fb8874d8db46f52da49507abe) for the preregistered analyses). Our use of human subjects was approved by the Institutional Review Board at Harvard University.

Table 3.2. Participant mean (M) age in months and standard deviation (SD) by language

|          | 4 years                 | 5 years                 | 6 years                 |
|----------|-------------------------|-------------------------|-------------------------|
| English  | 35 (M= 53.43, SD= 3.34) | 38 (M=65.53, SD= 3.19)  | 23 (M= 75.74, SD= 3.43) |
| Mandarin | 38 (M= 53.03, SD=3.87)  | 39 (M= 65.23, SD= 3.13) | 21 (M= 76.29, SD= 3.10) |
| Turkish  | 29 (M= 53.97, SD= 3.61) | 41 (M= 65.95, SD= 3.33) | 23 (M= 76.65, SD= 3.24) |

## Tasks

This study was designed as a part of a larger project that aimed to investigate the relationship between a child’s language and their understanding of a variety of numerical concepts. Towards this end, we designed five tasks, but we will primarily focus on the findings from the first two tasks described in detail below. We will also report the findings from the Logic task in the Supplementary Information (SI). The tasks were exactly the same across all languages, except that they were translated from English into Mandarin and Turkish. Translations into Mandarin and Turkish were done by researchers who were native speakers of the languages and were checked by separate researchers who were also native speakers.

### Task 1: Verbal counting

In this task, children were asked to count as high as they could. If they hesitated to start, the experimenter prompted them by saying “one...”. At the point where the child first stopped or made a mistake, experimenters provided three sets of cues. First, if the child stopped, the experimenter prompted them by saying “can you keep going?”. If they did not continue or said the wrong number word, the experimenter then said, “what comes after n (n being the number the child stopped at/the last number word they said before making a mistake)?”. If they still did not

continue or said the wrong number word, the experimenter prompted them again by saying “n-2, n-1, n” in rising intonation. As a final cue, the experimenter corrected them and gave them the next number word, n+1, to see if they could continue counting. The task was terminated when (i) the child could not count beyond the number where they made a mistake, even after the final cue of providing the correct next number or (ii) when the child had counted up to 121. For all the primary analyses, per our preregistration, we considered children’s count range as the number they counted up to before being given the last cue of the final number word.

### Task 2: The Base-10 Task

In this task, we measured children’s ability to manipulate sets of tens and units of ones. Children were introduced to a fictional cartoon character who runs a cookie shop (see Figure 3.1). Children were told that the cookie shop sells single cookies in small bags and a set of ten cookies in opaque large bags. They then saw a demonstration where ten cookies were being packed into the large bag one by one. This was done in order to make sure they understood that there are ten cookies inside the large bags. Before beginning the test trials, they were asked how many cookies are inside the large and small bags, and they proceeded to the test trials only when they answered this comprehension check correctly with or without feedback.

The task had three types of trials. For the first type of trials, called the *Compare* Condition (see Figure 3.1a), each trial presented images of cookies that two characters owned in the form of large bags and small bags. The child was asked to say which character had more cookies. For the second type of trials, called the *Count* Condition (see Figure 3.1b), each trial presented a certain number of large bags and small bags on the screen. The child was asked to say how many cookies there were altogether. For the third type of trials, called the *Create* Condition (see Figure 3.1c),

each trial presented a character who asked for a certain number of cookies. Children were asked to say how many large bags and how many small bags the shopkeeper should give the character.

These conditions were blocked and were presented in the following order: *Compare-Count-Create*. The *Count* and *Create* conditions had five trials each and the *Compare* condition had seven trials. The *Count* and *Create* conditions were preceded by one practice trial each (see SI for the full trial structure). All the numbers used in this task were below 100.

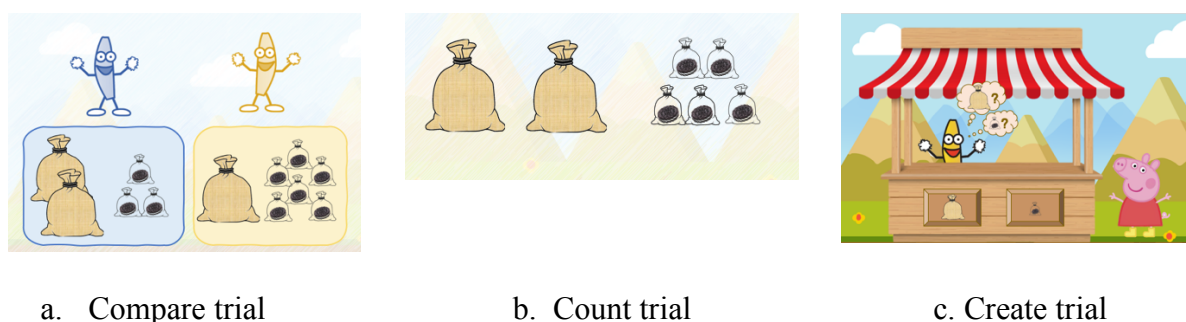


Figure 3.1. Example trials from the base-10 task. Children were told that the large opaque bag has 10 cookies and that the small transparent bag has one cookie. In (a) children had to decide whether the blue character or yellow character had more cookies, the right answer here is blue. In (b) children had to count the total number of cookies, the right answer here is 25. In (c) children had to decide how many large bags and how many small bags to give to a customer when they ask for a specific number of cookies. Here, the customer asked for “16” cookies so the right answer is one large bag and six small bags.

### Task 3: The Logic Task

In this task, we measured children’s understanding of the logic of exact equality, that is, the concept that certain transformations (e.g., stirring objects or removing one object and then adding the object back to the set) preserve set size and certain transformations (e.g., removing an

object or adding an object) do not. This task was adapted from Lipton & Spelke (2006) and it involved showing children animations where a certain number of balls were dropped into a transparent jar. The number of the balls was declared by the experimenter. Then certain transformations were performed to the set. They either changed the quantity of the marbles in the jar (e.g., removing one marble from the jar) or did not change the quantity (e.g., stirring the marbles in the jar with a spoon). Children were asked a two-choice question of whether the jar still has X number of marbles or not. There were 14 trials in this task. 7 of these trials involved quantity-preserving operations, while the remaining 7 involved quantity-changing operations.

In addition to the three tasks above, children also completed two other tasks in this study which are described in the SI but will not be reported in this paper. We administered the logic task in addition to the base-10 task in order to examine which numerical concepts were or were not affected by the degree of transparency of number word's in one's language. We did not have any language specific hypotheses for the logic task.

## **Procedure**

All children participated in the experiment virtually using video conferencing platforms (i.e., Zoom). The tasks were administered by a trained experimenter proficient in the child's language. In China and Turkey, children were tested by researchers who native speakers of Mandarin and Turkish respectively. In the US, children were tested by researchers who were native speakers of English and also by non-native speakers who were proficient in English. All children saw the stimuli displayed on a laptop screen shared via screen sharing and answered questions verbally.

### 3.2.2 Results

#### Effects of age and language on count range

In order to determine the effects of age and language on count range, we built linear regressions hierarchically using the `lm` function in R. We first analyzed our dependent variable, count range, as a function of age, and then sequentially added language and the interaction between age and language as predictors, in that order. We then performed model comparisons using Akaike Information Criterion (AIC) and the ANOVA function to determine the model with best fit. The ANOVA function compares each subsequent model with the previous model. See Table 3.3 for model specifications along with the results of the model comparisons. Before fitting any models, we transformed all numerical variables (i.e., age in months and count range) into Z-scores. In all the models, English is the reference group and against which both Mandarin and Turkish are being compared.

From table 3.3, we found Model 2 to be the best fitting model as it had the lowest AIC and it explained the data better than Model 1 ( $p < 0.01$ ) and Model 3 ( $p = 0.57$ ). In Model 2, we found significant effects of age ( $b = 0.49$ ,  $SE = 0.05$ ,  $t = 9.93$ ,  $p < 0.01$ ) and language (Mandarin ( $b = 0.37$ ,  $SE = 0.12$ ,  $t = 3.12$ ,  $p < 0.01$ ) and Turkish ( $b = -0.46$ ,  $SE = 0.12$ ,  $t = -3.82$ ,  $p < 0.01$ )). This indicates that children's count range improved with age and Mandarin-speaking children counted higher than English-speaking children and English-speaking children counted higher than Turkish-speaking children. See figure 3.2 for the predicted values of count range by language along with the raw data.

Table 3.3. Model comparisons for the linear regressions predicting count range

| Model Specifications                          | AIC    | <i>F</i> | df | p      |
|-----------------------------------------------|--------|----------|----|--------|
| M1: CountRange ~ Age + (1 Subject)            | 752.09 | -        | -  | -      |
| M2: CountRange ~ Age + Language + (1 Subject) | 711.24 | 23.86    | 2  | <0.01* |
| M3: CountRange ~ Age * Language + (1 Subject) | 714.08 | 0.57     | 2  | 0.57   |

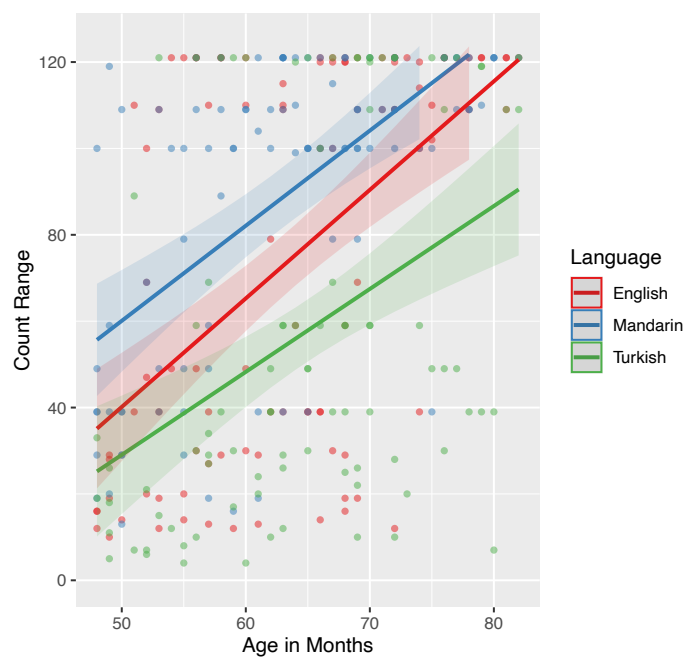


Figure 3.2. Predicted values of Count Range by Language. Individual dots represent the raw data.

### Effects of age, language and count range on base-10 task performance

Before conducting any statistical modeling, we first examined the effects of age and count range separately on children's base-10 performance across the three language groups. Towards this end, we plotted average base-10 scores (i.e., the percent correct) by age in months (see Figure 3.3) and then by count range (see Figure 3.4) for each language group.

Trends from figure 3.3 indicate that English-speaking and Mandarin-speaking children improved in their base-10 scores at similar rates with age, but Turkish-speaking children improved at a slower rate.

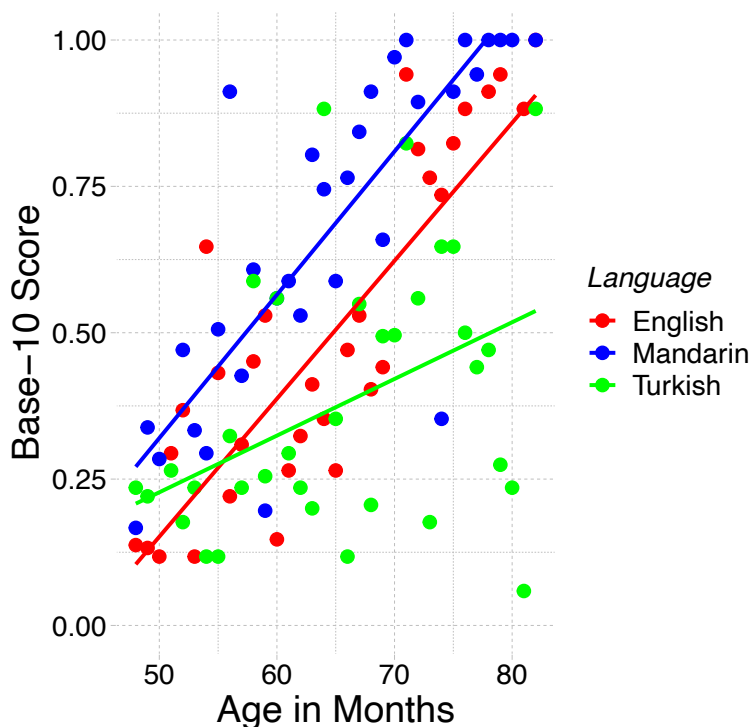


Figure 3.3 Average Base-10 Scores by Age by Language

Trends from figure 3.4 indicate that when matched for count range, English-speaking and Turkish-speaking children had similar average base-10 scores, but Mandarin-speaking children had slightly higher base-10 scores than English-speaking children at the higher count ranges.

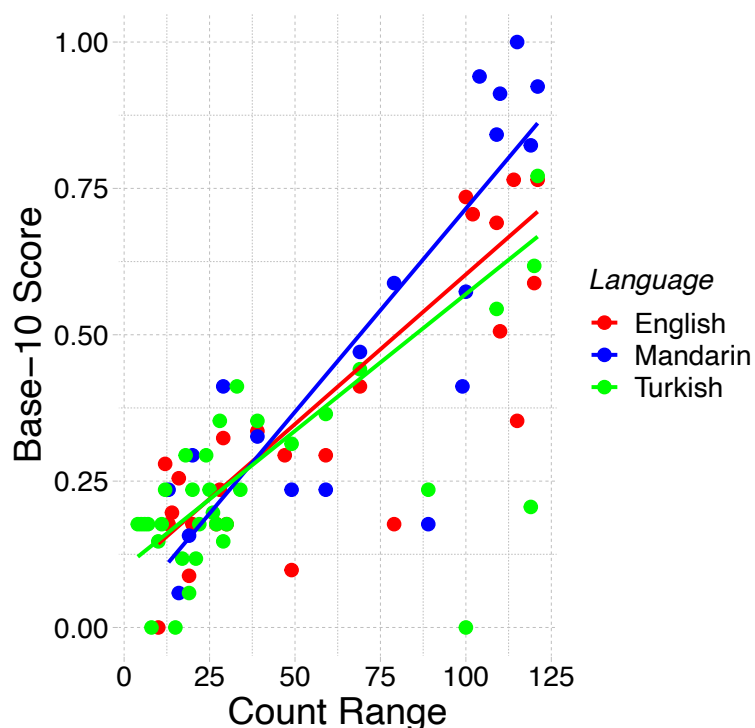


Figure 3.4. Average Base-10 Scores by Count Range by Language

In order to further examine these trends and determine the effects of age and language on base-10 performance after controlling for count range, we turned to statistical modeling. In order to determine the effects of age, language, and count range on children's base-10 task performance, we built logistic mixed-effects models in R using the glmer function from the lme4 package (Bates et al., 2015). We included subject and trial as random intercepts in all models.

We first analyzed our dependent variable, base-10 task performance (coded as 0 or 1), as a function of age, and then sequentially added the predictors of language, and the interaction between age and language, in that order. As a final step, we added count range as a predictor to the same model. We then performed model comparisons using AIC and the ANOVA function to determine the model with best fit. See Table 3.4 for model specifications along with the results of the model comparisons.

From table 3.4, we found that model Model 4 was the best fitting model as it explained the data better than Model 3 ( $p < 0.01$ ). In Model 4, we found main effects of age ( $b = 1.02$ ,  $SE = 0.2$ ,  $z = 5.19$ ,  $p < 0.01$ ), and language for Mandarin ( $b = 1.22$ ,  $SE = 0.27$ ,  $z = 4.58$ ,  $p < 0.01$ ) but not Turkish ( $b = -0.25$ ,  $SE = 0.26$ ,  $z = -0.95$ ,  $p = 0.34$ ). We also found main effects of count range ( $b = 1.29$ ,  $SE = 0.13$ ,  $z = 10.21$ ,  $p < 0.01$ ). This indicates that children's base-10 task performance improved with age and count range, and that Mandarin-speaking children performed better than English-speaking children but there was no difference between English-speaking children and Turkish-speaking children. We also found a significant interaction between age and language for Turkish ( $b = -0.94$ ,  $SE = 0.26$ ,  $z = -3.6$ ,  $p < 0.01$ ) but not for Mandarin ( $b = 0.48$ ,  $SE = 0.27$ ,  $z = 1.77$ ,  $p = 0.08$ ) in this model. This indicates that the differences in the rates of improvement in base-10 task performance by age differed significantly between English and Turkish but not between English and Mandarin, when controlling for count range. See figure 3.5 for the predicted probabilities of base-10 scores by age by language.

Table 3.4. Model comparisons for the mixed-effects models predicting base-10 task performance

| Model Specifications                                                    | AIC     | $\chi^2$ | df | p       |
|-------------------------------------------------------------------------|---------|----------|----|---------|
| M1: Performance ~ Age + (1 Subject) + (1 Trial)                         | 4190.47 | -        | -  | -       |
| M2: Performance ~ Age + Language + (1 Subject) + (1 Trial)              | 4130.86 | 63.61    | 2  | < 0.01* |
| M3: Performance ~ Age * Language + (1 Subject) + (1 Trial)              | 4107.44 | 27.42    | 2  | < 0.01* |
| M4: Performance ~ Age * Language + CountRange + (1 Subject) + (1 Trial) | 4016.75 | 92.68    | 1  | < 0.01* |

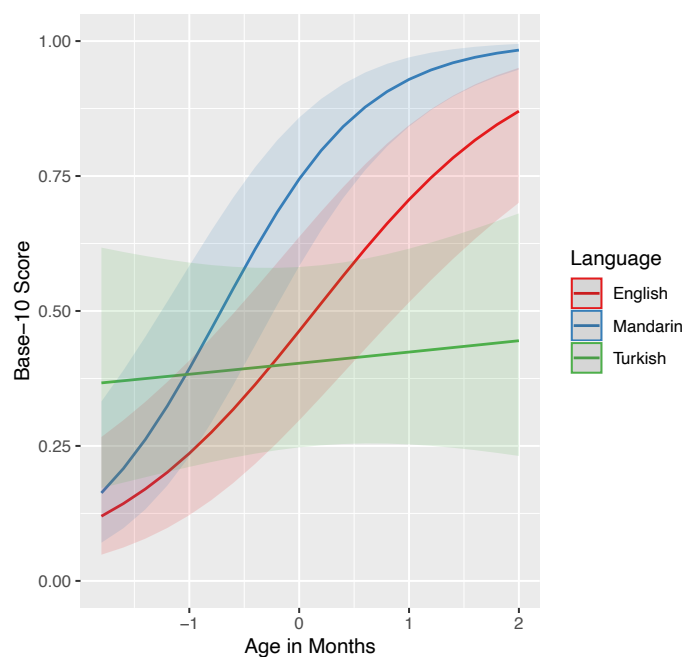


Figure 3.5. Predicted probabilities of Base-10 Scores from Model 4 by Age by Language

### **Exploratory analyses: Children's failure points in their count range**

In order to examine where children's stopping/failure points are in their count range (before being provided with the next number), we coded the failure points as follows:

1-9: Below 10

11-18: Teen

x9: The number before the decade (e.g., nineteen, twenty-nine, thirty-nine, and so on)

Not-x9: Anywhere between the decades (e.g., twenty-one to twenty-eight or fifty-one to fifty-eight, for example)

100: If they reach 100

Max: If they reach 121

We then plotted the percentage of children reaching these various stopping points by language. See Figure 3.6 for the plot. From the plot, we do not see any big differences in the percentage of children whose stopping points are either in the teens or right before the decade between English-and-Turkish-speaking children.

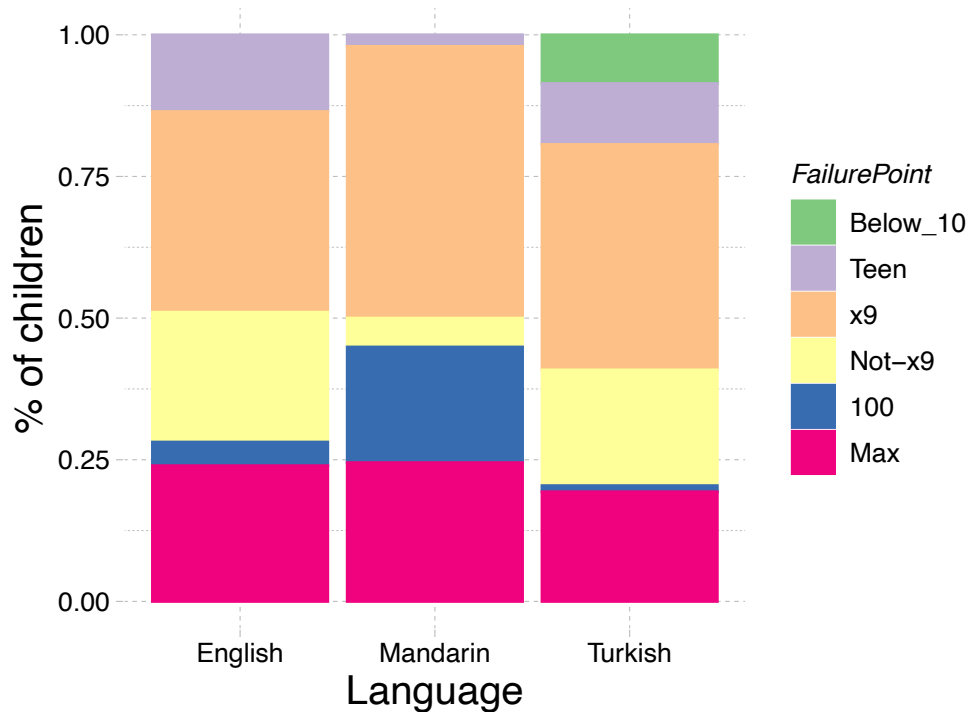


Figure 3.6. Percentage of children at various stopping points in count range by language

### 3.2.3 Discussion

In study 1, we conducted a three-way comparison between children who speak English, Mandarin, and Turkish in their count skills and in their understanding of the base-10 compositional number system. We compared English-speaking children with Turkish-speaking children in order to examine the effects (if any) of more transparent teens versus decades in the count list. We compared English-speaking with Mandarin-speaking children in order to compare English with a perfectly transparent language like Mandarin and also because most past research compared these two languages.

When we examined how high children could count, we found that Mandarin-speaking children counted higher than English-speaking children, replicating past work (Miller et al., 1995). We also found that English-speaking children counted higher than Turkish-speaking children. This

trend might not be due to the differences in the relative degree of transparency of number words in English and Turkish as children in both groups were similarly likely to stop at the teens or right before a decade (e.g., fifty-nine) when counting. If the non-transparent decades in Turkish were an impediment to developing productive counting skills, we should have seen a larger percentage of Turkish-children stopping right before a decade as compared to English-speaking children. In fact, a large percentage of Mandarin-speaking children also stopped right before a decade when counting, suggesting that learning the productive rules of counting in a base-10 number system might be hard for children even when their language's number naming systems are relatively more transparent. We also found no differences between languages in the rates at which children's counting skills improved with age.

When we examined children's base-10 task performance, we found differences in the rates at which English-speaking and Turkish-speaking children improved with age, even after controlling for count range. Turkish-speaking children improved at a much slower rate as compared to English-speaking children. This might be due to the non-transparent decades in Turkish, which could be a bigger impediment to the acquisition of base-10 compositionality as compared to the non-transparent teens in English.

However, if the differences in base-10 performance were only due to the non-transparent decades in Turkish, Turkish-speaking children should perform similarly to English-speaking children in a task that does not depend upon a precise understanding of base-10 compositionality. Children in our study completed the logic task, which measured their understanding of the principle that number words refer to exact cardinal values. Children can succeed in this task if they know that "fifty" refers to an exact quantity, but they need not know that it refers to "five tens", for example. We found that Turkish-speaking children had lower scores than English-speaking

children in the logic task where we did not expect any specific language related trends (see SI for the results from statistical modelling). This suggests that the difference between English-speaking and Turkish-speaking children's base-10 understanding might not be solely due to the differences in their number naming systems.

Additionally, there were no differences in the rates of improvement on the base-10 task performance with age between English-speaking and Mandarin-speaking children, after controlling for count range. However, Mandarin-speaking children had higher scores as compared to English-speaking children. Since there were no differences between the two groups in the logic task (see SI for the results from statistical modelling), it is possible that the difference in base-10 understanding might either be due to the differences in the number naming systems between English and Mandarin or due to other factors such as differences in formal instruction.

Overall, in study 1 we found that there were differences in how high children could count based on language. While it led to no difference in the rates of improvement in base-10 understanding by age between English-and-Mandarin-speaking children, it led to differences between English-and-Turkish speaking children such that Turkish speaking children improved at a slower rate. Since Turkish-speaking children performed worse than English-speaking children in numerical tasks that we expected to not be influenced by differences in the number naming systems, we think that factors that covary with language (e.g., formal instruction) might be influencing children's developing numerical understanding.

### **3.3 Study 2**

In study 2, we compared a highly non-transparent language, Hindi, to a transparent language, Tamil, both spoken within India. By studying children from similar educational and SES backgrounds within the same country and comparing their developing counting skills and base-10

understanding, we attempted to account for the effects of language that are independent of factors that tend to co-vary with language.

### **3.3.1 Methods**

#### **Participants**

A total of 230 children who were monolingual speakers of Hindi ( $n = 111$ ) and Tamil ( $n = 119$ ) within India were included in the final sample of participants. We aimed to recruit 124 4-8-year-olds in each language but excluded those who did not meet our preregistered inclusion criteria (see [https://osf.io/zr5um/?view\\_only=1579dc9fb8874d8db46f52da49507abe](https://osf.io/zr5um/?view_only=1579dc9fb8874d8db46f52da49507abe) for the preregistration). We recruited Tamil-speaking children from four different schools in various parts of Tamil Nadu and Hindi-speaking children from four different schools in Gujarat and Hyderabad. The schools that the children were recruited from serve low-income families and their primary language of instruction were Hindi or Tamil.

In the schools that we worked with, we found that the age at which children started school was variable. Therefore, in order to equate for amount of experience with mathematics in school, we roughly aimed to test equal number of children in grades 1, 2, and 3. Children in both languages were tested between the months of June 2023 and September 2023. See Table 3.5 for sample size broken down by grade in each language.

Each participant received a small gift (e.g., a book) of the same value. The sample size, exclusion criteria, and methods were preregistered prior to collecting any data. The analysis plan was preregistered prior to any examination of the data by the researchers (see [https://osf.io/zr5um/?view\\_only=1579dc9fb8874d8db46f52da49507abe](https://osf.io/zr5um/?view_only=1579dc9fb8874d8db46f52da49507abe) for the preregistered analyses). Our use of human subjects was approved by the Institutional Review Board at Harvard University.

Table 3.5. Participant grade, mean (M) age in months and standard deviation (SD) by language

|       | Grade 1                  | Grade 2                 | Grade 3                 |
|-------|--------------------------|-------------------------|-------------------------|
| Hindi | 50 (M= 70.30, SD= 12.46) | 37 (M=78.95, SD= 7.73)  | 24 (M= 80.17, SD= 7.15) |
| Tamil | 48 (M= 72.08, SD=7.08)   | 39 (M= 76.77, SD= 4.74) | 32 (M= 84.19, SD= 7.12) |

## Tasks

All the tasks were exactly the same as in Study 1, except that they were translated into Hindi and Tamil. Translation into Hindi was done by an elementary school Hindi teacher who was a native speaker of Hindi and was checked for accuracy by a researcher who was a native speaker of Hindi. Translation into Tamil was done by a researcher who was a native speaker of Tamil and was checked by separate researcher who was also a native speaker of Tamil.

## Procedure

Participants were tested in their schools by trained experimenters who were native speakers of Hindi or Tamil. Each participant was tested separately in a quiet and empty classroom. Participants were introduced to the experimenter by a teacher in the school and were told that the teacher would be waiting outside the classroom. All children saw the stimuli displayed on a laptop screen and answered questions verbally. For the Give-N task, the experimenters used familiar objects to the children (e.g., identical erasers) and a flat plate.

### 3.3.2 Results

#### Effects of age and language on count range

In order to determine the effects of age and language on count range, we built linear regressions hierarchically using the `lm` function in R. We first analyzed our dependent variable, count range, as a function of age, and then sequentially added language and the interaction between age and language as predictors, in that order. We then performed model comparisons using Akaike Information Criterion (AIC) and the ANOVA function to determine the model with best fit. The ANOVA function compares each subsequent model with the previous model. See Table 3.6 for model specifications along with the results of the model comparisons. Before fitting any models, we transformed all numerical variables (i.e., age in months and count range) into Z-scores. In all the models, Hindi is treated as the reference group against which Tamil is being compared.

From table 3.6, we found Model 3 to be the best fitting model as it had the lowest AIC and it explained the data better than Model 2 ( $p < 0.01$ ). In Model 3, we found no significant effect of age ( $b = 0.02$ ,  $SE = 0.08$ ,  $t = 0.27$ ,  $p = 0.78$ ) but we found a significant effect of language ( $b = 0.58$ ,  $SE = 0.12$ ,  $t = 69$ ,  $p < 0.01$ ) and a significant interaction between age and language ( $b = 0.36$ ,  $SE = 0.13$ ,  $t = 2.84$ ,  $p < 0.01$ ). This indicates that the rates of improvement in count range by age differed significantly by language, with Tamil-speaking children improving at a faster rate as compared to Hindi-speaking children. See figure 3.7 for the predicted values of count range by language along with the raw data.

Table 3.6. Model comparisons for the linear regressions predicting count range

| Model Specifications                          | AIC    | <i>F</i> | df | p      |
|-----------------------------------------------|--------|----------|----|--------|
| M1: CountRange ~ Age + (1 Subject)            | 650.34 | -        | -  | -      |
| M2: CountRange ~ Age + Language + (1 Subject) | 631.15 | 22.59    | 1  | <0.01* |
| M3: CountRange ~ Age * Language + (1 Subject) | 625.09 | 8.06     | 1  | <0.01* |

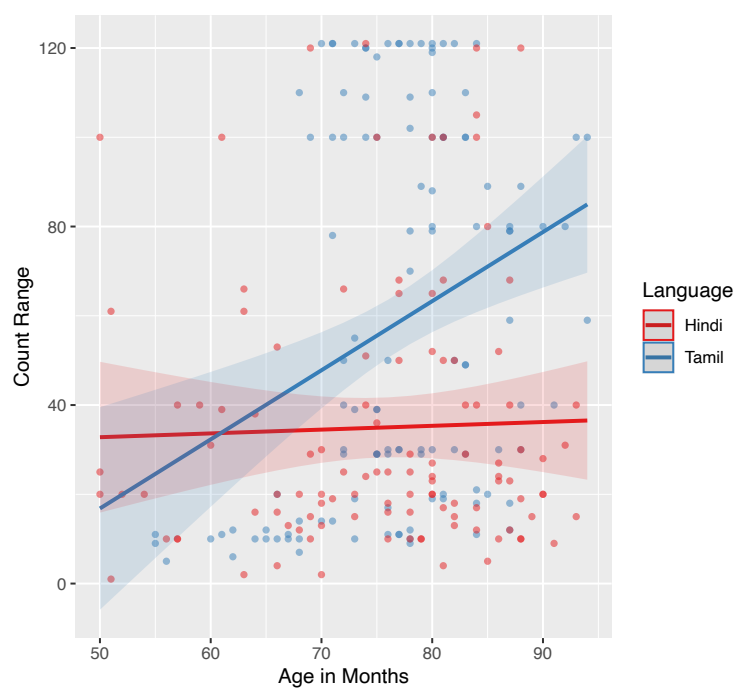


Figure 3.7. Predicted values of Count Range by Age by Language. Individual dots represent the raw data

### Effects of age, language and count range on base-10 task performance

Similar to study 1, before conducting any statistical modeling, we first plotted average base-10 scores (i.e., the percent correct) by age in months (see Figure 3.8) and then by count range (see Figure 3.9) for each language group.

Trends from figure 3.8 indicate that Tamil-speaking children improved in their base-10 scores at a faster rate with age as compared to Hindi-speaking children.

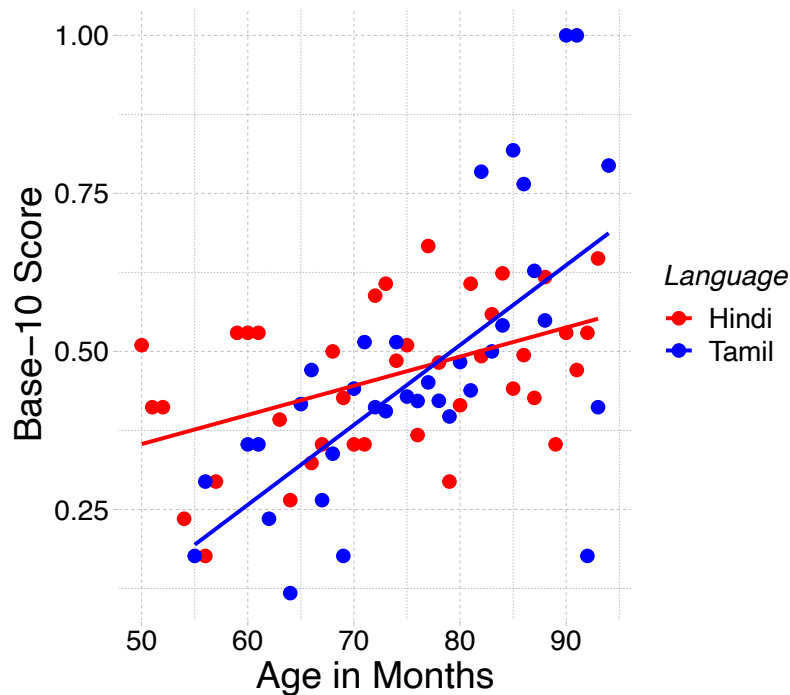


Figure 3.8. Average Base-10 Scores by Age by Language

Trends from figure 3.9 indicate that when matched for count range, Hindi-speaking children have higher average base-10 scores as compared to Tamil-speaking children.

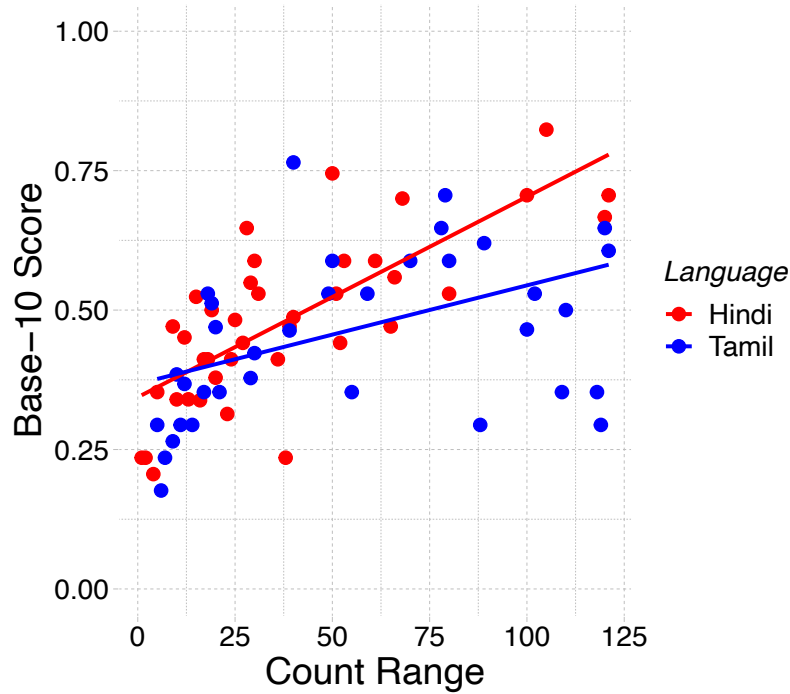


Figure 3.9. Average Base-10 Scores by Count Range by Language

In order to further examine these trends and determine the effects of age and language on base-10 performance after controlling for count range, we turned to statistical modeling. In order to determine the effects of age, language, and count range on children’s base-10 task performance, we built logistic mixed-effects models in R using the `glmer` function from the `lme4` package (Bates et al., 2015). We included subject and trial as random intercepts in all models. We first analyzed our dependent variable, base-10 task performance (coded as 0 or 1), as a function of age, and then sequentially added the predictors of language and the interaction between age and language, in that order. As a final step, we added count range as a predictor to the same model. We then performed model comparisons using AIC and the ANOVA function to determine the model with best fit. See Table 3.7 for model specifications along with the results of the model comparisons.

From table 3.7, we found that model Model 4 was the best fitting model as it explained the data better than Model 3 ( $p < 0.01$ ). We found that there were significant effects of age ( $b = 0.29$ ,  $SE = 0.09$ ,  $z = 3.02$ ,  $p < 0.01$ ), language ( $b = -0.38$ ,  $SE = 0.16$ ,  $z = -2.34$ ,  $p < 0.05$ ), and count range ( $b = 0.5$ ,  $SE = 0.08$ ,  $z = 6.03$ ,  $p < 0.01$ ) on base-10 task performance. We did not find a significant interaction between age and language ( $b = 0.26$ ,  $SE = 0.15$ ,  $z = 1.7$ ,  $p = 0.09$ ) in this model. This indicates that the differences in the rates of improvement in base-10 task performance by age did not differ significantly by language when controlling for count range. See figure 3.10 for the predicted probabilities of base-10 scores by age by language.

Table 3.7. Model comparisons for the mixed-effects models predicting base-10 task performance

| Model Specifications                                                    | AIC     | $\chi^2$ | df | p       |
|-------------------------------------------------------------------------|---------|----------|----|---------|
| M1: Performance ~ Age + (1 Subject) + (1 Trial)                         | 3825.88 | -        | -  | -       |
| M2: Performance ~ Age + Language + (1 Subject) + (1 Trial)              | 3827.73 | 0.14     | 1  | 0.71    |
| M3: Performance ~ Age * Language + (1 Subject) + (1 Trial)              | 3823.72 | 6.01     | 1  | < 0.05* |
| M4: Performance ~ Age * Language + CountRange + (1 Subject) + (1 Trial) | 3792.18 | 33.54    | 1  | <0.01*  |

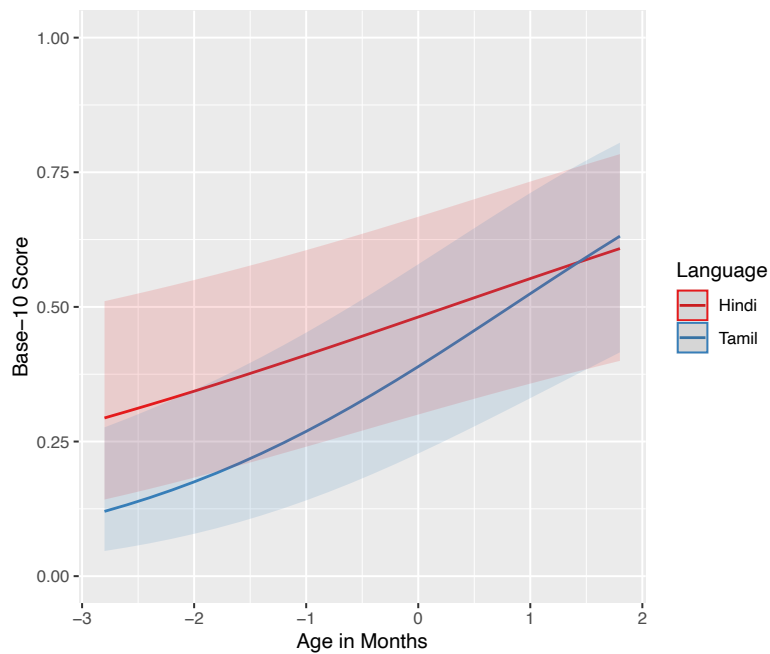


Figure 3.10. Predicted probabilities of Base-10 Score from Model 4 by Age by Language. Age in Months on the x-axis reflect scaled values

### Exploratory analyses: children's failure points in their count range

In order to examine where children's stopping/failure points are in their count range (before being provided with the next number), we coded the failure points similarly to study 1. We then plotted the percentage of children reaching these various stopping points by language. See Figure 3.11 for the plot. From the plot, we found that a larger proportion of Hindi-speaking children were stopping in between the decades when counting as compared to Tamil-speaking children.

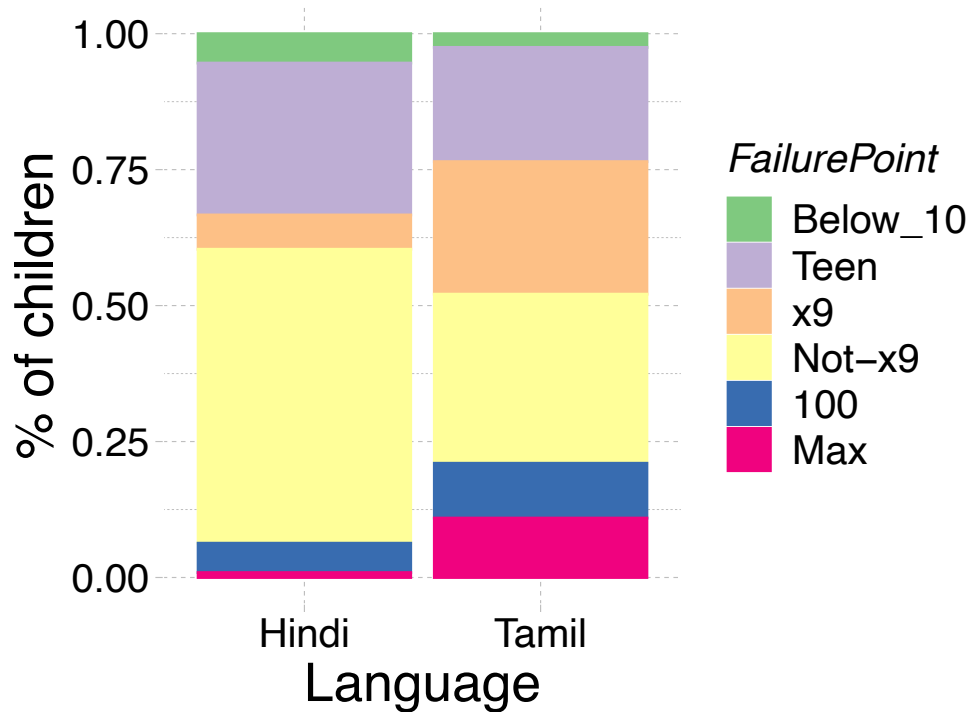


Figure 3.11. Percentage of children at various stopping points in count range by language

### 3.3.3 Discussion

In study 2, we aimed to compare Hindi-speaking and Tamil-speaking children's counting skills and their understanding of the base-10 compositional number system in order to examine the effects a highly non-transparent language like Hindi while accounting for children's SES and educational backgrounds.

When we examined how high children could count, we found that Tamil-speaking children counted higher and also improved in their count range with age at a faster rate as compared to Hindi-speaking children. This pattern might be due to the relative degree of transparency of the number words between the two languages as we found that a larger proportion of Hindi-speaking children stopped in between the decades when counting as compared to Tamil-speaking children.

Since Hindi number words are not fully transparent even between the decades, children might be rote learning the number names without being able to extract any systematic rules. Moreover, we also found that Tamil-speaking children produced an average of eight number words after being provided with the next number in the count list, whereas Hindi-speaking children produced an average of only two number words (see SI for the details of these analyses). This suggests that due to the more transparent number naming system in Tamil, children have extracted some systemic rules that allow them to count more productively as compared to Hindi-speaking children.

When we examined children's base-10 task performance, we found that Hindi-speaking children performed better than Tamil-speaking children when controlling for count range. This pattern challenges the hypothesis that favors the effects of different number naming systems on children's base-10 understanding. While Hindi-speaking children learn to count at a much slower rate as compared to Tamil-speaking children, once they have the same count range as Tamil-speaking children, they perform better in the base-10 task. Some of the Hindi-speaking children in our sample entered school at an earlier age as compared to the Tamil-speaking children. Therefore, formal instruction in school could be driving children's performance in the base-10 task as opposed to different number naming systems.

We also found the same pattern for the logic task, where we did not expect any language specific trends. We found that Hindi-speaking children performed better than the Tamil-speaking children after controlling for count range.

Over all, in study 2, we found differences in how high children could count based on language, where Tamil-speaking children counted higher than Hindi-speaking children. However, once children's count range is accounted for, it does not lead to better understanding of base-10

compositionality based on language. We think that other factors such as formal instruction in school might be influencing children's developing base-10 understanding more strongly.

### **3.4 General Discussion**

In this study, we examined whether the degree of transparency of number words across five different languages affects children's understanding of the base-10 compositional number system.

In study 1, we compared English (with non-transparent teens and semi-transparent decades) to Turkish (with transparent teens and non-transparent decades) and English to Mandarin (with fully transparent number words). We found differences in how high children could count based on language, such that Mandarin-speaking children counted higher than English-speaking (replicating past work e.g., Miller et al., 1995) children who in turn counted higher than Turkish-speaking children.

We found that though English-speaking and Mandarin-speaking children improved in their base-10 understanding at similar rates with age, Mandarin-speaking children had higher base-10 scores than English-speaking children at any given age. These differences in base-10 understanding appeared as early as 4-years in our sample, that is, prior to any formal instruction in base-10 concepts or arithmetic. This finding contradicts some past work that found no language effects among Kindergarteners (e.g., Vasilyeva et al., 2015).

Critically, since we found no major differences between English-speaking and Mandarin-speaking children in a numerical task that does not depend upon a precise base-10 understanding, any differences in base-10 understanding could either be attributable to the number naming systems or exposure to numerical concepts at home or in preschool activities.

We also found a much slower improvement in base-10 understanding among Turkish-speaking children as compared to English-speaking children, after controlling for count range. While this could be due to the non-transparent decades in Turkish, we think that factors other than language might be explaining this difference as Turkish-speaking children performed worse on a numerical task that we expected to not be influenced by language's number naming systems.

Turkish-speaking children's performance in our sample might be attributable other factors such as formal instruction or a lack of exposure to online testing with an unfamiliar experimenter. The English-speaking and Mandarin-speaking children in our sample were mostly recruited through a website or database of families who agreed to participate in research. Therefore, they are likely to have been exposed to online testing with different experimenters for other studies in the past. The Turkish-speaking children in our sample were mostly recruited through social media and word of mouth, therefore, they may have less exposure to online testing, and this may have affected their performance in all our numerical tasks.

In study 2, we compared Hindi (a highly irregular number naming system) with Tamil (a regular number naming system). Similar to study 1, we found differences in how high children could count, with Tamil-speaking children counting higher than Hindi-speaking children. However, this did not lead to better understanding of base-10 compositionality among Tamil-speaking children as compared to Hindi-speaking children. Like in study 1, we think that factors other than language might be explaining this difference. It seems that formal instruction in school and exposure to math helps Hindi-speaking children learn base-10 concepts even though their number words obscure them.

## **Limitations and future directions**

While we tried to account for children's SES and educational backgrounds as much as possible in study 2, there still might be critical differences among the Hindi-and-Tamil-speaking children in our sample. For example, while the Tamil-speaking children were recruited from Tamil Nadu, a state in India that predominantly speaks Tamil, the Hindi-speaking children were recruited from Hyderabad and Gujarat, states where Hindi is one of the languages that is spoken there. Therefore, the Hindi-speaking children in our sample might be exposed to multiple languages and this might also influence their performance. Future work must aim to recruit children from more similar language environments within multilingual countries.

Future work must reexamine the effects of different number naming systems on children's understanding of base-10 compositionality by administering control tasks in a different domain (e.g., reading or vocabulary) that children also come to engage with in school. If children's performance in the control tasks are similar across languages, then it would enable us to more directly examine any differences in numerical understanding that could be attributable to the languages' number naming systems.

In our study, we did not measure children's understanding of the number symbols that also serve to represent base-10 compositionality. Future work must also aim to test children's grasp of multidigit number symbols across different languages. This would enable us to examine whether children can grasp number symbols and the base-10 compositional rules that underlie them independent of their language's number naming systems.

Ultimately, we believe that this research has implications for educational interventions, as it suggests that base-10 compositionality is hard and takes several years for children to master, even if their language's number naming system is as transparent as Mandarin or Tamil. It suggests

that representations in addition to number words (e.g., number symbols), that children get exposed to informally or via school, might be crucial in helping them learn base-10 compositionality.

Together, across both our studies we found that number naming systems in different languages affect children's counting skills, but once they are accounted for, any differences in numerical understanding, particularly of the base-10 number system might not be due to language.

## Chapter 4

### Compositional Numerical Concepts

A home-based game training base-10 compositionality improved 5-6-year-old children's numerical concepts and skills

#### 4.1 Research Highlights

- A randomized experiment training compositional math or mental state concepts enhanced 5- to 6-year-old children's mathematical skills but not their mental state reasoning.
- A short-term home-based intervention shed light on one effective feature of a successful long-term school-based mathematics intervention exercising children's grasp of the base-10 number system.
- Home-based experiments complement field experiments in education shedding light on the effectiveness of one component of the multifaceted educational intervention.
- Children learn best in school when supported by family activities at home. Home-based educational interventions with children and their caregivers may foster this supportive environment.

#### 4.2 Abstract

The base-10 number system is foundational for mastery of mathematics in primary school, but it is challenging for young children for at least two reasons: number words and symbols do not explicitly convey the compositional rules that underlie them, and arithmetic algorithms learned in school require switches between representations such as *one ten* and *ten ones*. Building upon past research, we evaluated a game that aimed to improve 5- to 6-year-old children's understanding of

the base-10 number system, together with the addition algorithm, using arrays of dots grouped by tens and a 10 x 5 spatial board, coupled with numerical language. Using a pretest-posttest design, we compared the effects of this game, played at home, to the effects of playing a game that aimed to improve children's compositional reasoning about mental states of belief and desire. After 2-3 weeks of home-based play, children who played the numerical game improved their school-relevant numerical skills, relative to those who played the mental state game; improvements in mental state reasoning did not differ across the two conditions. These findings provide evidence that home-based game training of compositional base-10 number concepts improved children's numerical skills and compositional reasoning in the short term, but a similar game training compositional mental state concepts did not. Each game tested in this experiment formed a part of a different successful, large-scale, school-based educational intervention using multiple games; our findings therefore shed light on the components of those large-scale interventions that contributed to their success.

Key words

compositionality, numerical understanding, education

### **4.3 Introduction**

Throughout school, children are challenged to master compositional systems. To learn mathematics, they must grasp how numbers combine to form larger numbers, and how these combinations are used for performing arithmetic. To learn to read, they must grasp how speech sounds combine to form words, how words combine to form phrases and sentences, and how these combinations are used to convey meaning. To develop the skills of pedagogical learning, or “learning to learn,” children must grasp how their own and others' mental states of perception, intention, belief and desire combine to give rise to knowledge.

In the present study, we focused on the first challenge: Using a home-based training paradigm, we evaluated a game that aimed to improve children's understanding of numerical concepts generated by the base-10 system. This numerical game, called *Find and Move*, previously featured in a successful large-scale Randomized Controlled Trial (RCT) in Delhi, India along with three other math games (Dean et al., in prep). In order to begin evaluating the reasons for the success of the RCT, we evaluated the specific, short-term impact of *Find and Move*, when played alone, by comparing its effects to those of a similar game that focused on mental state concepts.

The social-cognitive game, *Mind Hunter*, previously featured, together with a second social-cognitive game, in a successful RCT in Salvador, Brazil (Andrade et al., in prep). The current experiment therefore assessed the distinct contributions of two individual games focused on compositional reasoning in two different domains, in a single home-based study. Here we focus primarily on the numerical game.

An important feature of compositional systems is their extraordinary expressive power: a limited set of foundational concepts, together with the rules for combining them, lead to potentially unbounded abilities to calculate with numbers, read texts, and infer the mental states of the self and others (Fodor, 1975, Fodor & Pylyshyn, 1988). In the number domain on which we focus, ten digits (0 to 9) each designate an integer from 0 to 9 when presented singly, a multiple of 10 when presented to the left of a second digit, and a multiple of 100 when presented to the left of two digits. Thus, a number such as '35' can be expressed using only these symbols if one realizes that it designates three tens plus five ones. This system allows for the conversion of large quantities into short strings of words and symbols, composed of a small number of elements that children can learn to read, write, and manipulate.

## **The Logic of Base-10 Compositionality is Hard for Children to Learn**

Though learning the compositional number system is essential for the meaningful learning of mathematics, it is hard for children to master because the rules for composing number concepts from powers of ten are rarely fully apparent in the language and symbols that express those concepts. For example, children as young as 3 years of age have partial knowledge of large multidigit numbers. For example, children aged 3-6-years can map number words (two hundred thirty-five) to their written symbols (235 and not 523) (Mix et al., 2014; Yuan et al., 2019) and judge which of the two symbols designates a larger number above chance (Cheung & Ansari, 2021; Mix et al., 2014; Yuan et al., 2019). By age 5, children also can solve approximate addition and subtraction problems involving two-digit numbers: for example, they can judge above chance that if one child has 21 candies and she gets 30 more candies she will have more candies than a child with 34 candies (Gilmore et al., 2007).

Closer examination reveals, however, that children fail to grasp the logic of base-10 compositionality: their success in these tasks reflects their use of heuristics, such as attending to the number of digits, or to the leftmost digit when two numbers are composed of the same number of digits, when determining magnitude (Byrge et al., 2014; Cheung & Ansari, 2021; Mix et al., 2014). Children may apply these heuristics without realizing why they are effective or what exact quantities the symbols designate.

Children need to go beyond such intuitions to learn exact arithmetic in a meaningful manner: they need to represent multidigit numbers as composed of sets generated by different powers of ten (initially, only numbers composed of 1 to 9 tens; later, also numbers whose composition includes 1 to 9 hundreds). This ability is crucial for understanding calculation procedures such as carrying in addition and borrowing in subtraction.

Learning the logic of base-10 compositionality is hard for children, however, for several reasons. First, the number words that children learn in many languages, including English, obscure base-10 compositionality. For example, number words like “eleven” and “twenty” do not explicitly convey that they designate sets that are composed from “ten and one” and from “two tens,” respectively, or that the compositional rule for “eleven” is to add the numbers, whereas the compositional rule for “twenty” is to multiply them. In East-Asian languages, where base-10 compositionality is more transparently reflected in number words (e.g., the number words for *eleven* and *twenty* in Mandarin translate literally to *ten one* and *two ten*), children are better at representing multidigit numbers as composed of sets of tens and ones (Ho & Fuson 1998; Miura et al., 1988, Miura et al., 1993, Miura & Okamoto, 2003), although they too learn slowly that the number words expressing decades are related by multiplication (Miller et al., 1995).

Second, written symbols for numbers do not convey the rule that numbers increase by powers of ten from right to left: the opposite of the left-to-right ordering that is conventional for displaying the numbers from 1 to 9 on a ruler or number line. Thus, the “3” in “32” represents  $3 \times 10^1$ , whereas in “23” it represents  $3 \times 10^0$ . Third, problems arise as children seek to master the relationships between spoken number words and written numbers. In some languages, and for some number words, the orderings of numerals and spoken numerical morphemes are reversed. The English word for sixteen, for example, places “six” before the morpheme designating ten, whereas the written symbol for sixteen places the quantifier that designates the “tens” before the “6” that designates the “ones” (Klein et al., 2013).

Finally, to engage in meaningful learning of arithmetic algorithms, children need to represent two-digit numbers in two different ways: *ten* must be represented both as “one ten” and

as “ten ones,” as the algorithms of arithmetic such as carrying and borrowing depend on conversions between these two representations.

It is not surprising, therefore, that base-10 compositionality poses problems for child learners. When challenged to add two-digit numbers, presented with symbols and words, they may apply the addition algorithms not as a series of logical steps that treat quantities (e.g., *twenty-seven*) as compositions of tens and ones, but rather as a series of arbitrary procedures for manipulating individual numerals (here, 2 and 7). Children who fail to master the logic of base-10 compositionality have difficulties in arithmetic (Fuson, 1990) and are at a greater risk for low math achievement in elementary school (Chan, 2014; Moeller et al., 2011). The importance of understanding base-10 compositionality, together with the sources of difficulty children face in learning it, underscore the need for developing and evaluating effective interventions to elicit and strengthen this understanding.

### **Previous Interventions to Improve Children’s Understanding of Base-10 Compositionality**

Previous interventions have aimed to aid children’s discovery of base-10 compositionality using blocks that contain sets generated by different powers of ten (Fuson & Briars, 1990; Laski et al., 2016; Mix et al., 2017; Yuan et al., 2021). Such blocks are meant to help children map number words and symbols to the physical quantities they represent. Base-10 blocks, commonly used in Montessori schools, consist of small cubes to represent numbers from one to nine, sticks composed of ten cubes to represent tens, and flat squares composed of ten sticks to represent hundreds. The efficacy of base-10 blocks has been studied both by using short interventions (Fuson & Briars, 1990; Mix et al., 2017; Yuan et al., 2021) and also by comparing children with years of Montessori education, who consistently played with base-10 blocks in elementary school, to those

studying in non-Montessori schools where such materials are not commonly used (Mix et al., 2017; Laski et al., 2016). The evidence bearing on the efficacy of base-10 blocks is mixed.

Laski, Vasilyeva, and Schiffman (2016) found that kindergarten children in Montessori schools had a better understanding of base-10 compositionality than their non-Montessori peers, as measured by a task that required them to use base-10 blocks to represent multidigit numbers, but first grade children in the two groups showed no differences. The differences found in the kindergarten children also disappeared in second and third grade when children's arithmetic accuracy was measured.

Mix et al. (2017) found the opposite pattern in a cross-sectional comparison between children in Montessori and non-Montessori schools: the Montessori advantage appeared in second grade but not kindergarten, on tasks of place-value understanding and arithmetic. Such discrepancies might be due to differences in tasks used in the two studies or due to the math curricula within each type of school (Basargekar & Lillard, 2021).

Shorter interventions with first- and second-grade children in non-Montessori schools showed that exercises using base-10 blocks improved children's arithmetic skills and place-value understanding (Fuson & Briars; 1990; Mix et al., 2017). However, kindergarten children, on the threshold of arithmetic instruction, did not benefit from base-10 block exercises when learning multidigit numbers (Yuan et al., 2021). In present study, we aimed to train Kindergarten children's understanding of base-10 compositionality.

### **The Synergy Between Long-term, Randomized Controlled Field Interventions and Short-term, Home-based Interventions to Improve Children's Learning of School Mathematics**

The research reviewed above has several limitations. Studies that presented children with well-controlled and focused interventions were necessarily of short duration, since children would

not likely maintain interest in a single activity if it were presented over the course of a full school year. However, children master the logic of the base-10 number system slowly. As a result, short-term interventions may fail to capture the gradual experience of learning this system. Longitudinal studies have compared children who received different math curricula for one or more years. Although this is an appropriate time frame for assessing the effects of different curricula on children's emerging base-10 knowledge, schools with different math curricula are likely to differ in multiple ways. Therefore, they do not clearly reveal what aspects of the curriculum fostered base-10 learning.

It is possible, however, that small- and large-scale studies will synergize if they focus on the same interventions, because they have complementary strengths and limitations. Large-scale interventions, over long time scales, can shed light on children's gradual learning in their noisy and rich school environments, but such interventions necessarily are multi-faceted. Therefore, evaluations of their impacts don't reveal which facets of a successful intervention account for its success.

Small-scale interventions, over short time scales, have the opposite strengths and weaknesses: They can shed light on the specific effects of individual aspects of a large-scale intervention, by testing them one by one, but they provide only a snapshot of the learning process over a small interval of time. Here we ask whether two individual, brief, small-scale randomized interventions in children's homes can shed light on one of the potential reasons for the success of two extended, large-scale interventions in children's schools: one focused on mathematics and one focused on mental state reasoning.

The math game used in the present experiment, called *Find and Move*, featured, along with three other math games, in a large-scale field experiment conducted in Kindergarten and First

Grade classrooms in Delhi, India (Dean et al., in prep). The intervention was administered by the children's regular teachers. *Find and Move* trained children's understanding of base-10 compositionality using arrays of dots grouped by tens and a spatial board with numbers, coupled with numerical language. The other three games exercised representations of approximate number (but with arrays of dots grouped by ten), geometric properties of angle, relative length and symmetry, and geometric properties of distance and direction.

The math games curriculum, consisting of all four games, was presented for three hours a week throughout the 9-month school year. To assess the efficacy of this curriculum, kindergarten and first grade children were randomized at the class level to receive the math games curriculum or no treatment (a condition that provided them with 3 more hours of the standard math curriculum, relative to the treatment condition). The analyses revealed a positive effect of the math games curriculum with an effect size above the median effect size of published evaluations of educational interventions in mathematics (Evans & Yuan, 2022).

Since children in the treatment condition received four math games, however, the experiment does not reveal whether all or only some of the four games contributed to this impact, or what each game's specific impacts are. For example, it is possible that only the number game training approximate numerical magnitudes had an impact on children's numerical abilities, because the base-10 game, *Find and Move*, was too demanding. Since the assessment tests focused on a broad range of math skills, the experiment does not reveal whether the intervention enhanced children's base-10 knowledge specifically. For example, it is possible that the math games increased children's motivation for math learning but had no specific impact on what they learned.

## Overview of the Present Intervention

In order to complement the strengths of long-term classroom-based interventions, we presented an English translation of the base-10 game, *Find and Move*, to 5- to 6-year-old U.S. children in their homes. We evaluated its effects by assessing children's numerical knowledge in the lab, before and after this intervention. To distinguish effects of the game's numerical content from the general effects of home-based interactions on children's motivation and learning, we compared the numerical abilities of children who played this game to those of children who played *Mind Hunter*: a game that was developed to train compositional abilities in the domain of mental state reasoning, and that has been evaluated through a randomized school-based experiment on 5- to 6-year-old children in Salvador, Brazil (Andrade et al., in prep). Below we describe each game in detail.

### ***Find and Move*: A Number Game to Foster Base-10 Understanding**

*Find and Move* trains children's understanding of base-10 compositionality, using the numbers from 1 to 50. It uses multiple cues to make base-10 compositionality more transparent to children. First, it presents displays that group dots into sets of ten within a circle, such that the number of tens in a given number could be determined by counting the circles (see Figure 4.1 b-d).

Second, it spatially arranges the Arabic number symbols on a 10x5 board that visually depicts the numbers of tens and ones into which each number can be decomposed by their vertical and horizontal positions (see Figure 4.1a), similar to previous interventions (Laski & Siegler, 2014).

Numerical board games have been used to enhance children's mastery of the principle that numbers increase linearly in magnitude by successive addition of one: i.e., the difference between

1 and 2 is equal to the difference between 87 and 88 (Laski & Siegler, 2014; Ramani & Siegler, 2008, 2011; Siegler & Ramani, 2009; Whyte & Bull, 2008). For example, 5-6 year old children in one study played a board game where they moved 1-6 spaces (as determined by a spinner) on a 10x10 board from 1 to 100 (Laski & Siegler, 2014). This intervention improved children's numerical skills of counting, numerical comparison, numerical identification, and arithmetic.

These findings raise the possibility that number boards can be used to train children's understanding of base-10 compositionality, because each row of a 10x10 number board contains ten numbers, each number in a row contains ten fewer numbers than the number above it, and each column on the board contains the same number of ones that remain after grouping all the other elements into sets of ten.

Two further features of *Find and Move* aimed to enhance children's understanding of base-10 compositionality. First, it is played with cards that present Arabic number symbols together with the arrays of dots grouped by sets of ten, with the aim of fostering children's mastery of 2-digit numbers from 1 to 50. Second, it encouraged children and their caregivers to decompose number words into tens and ones. For example, "twenty-five" was modeled during training as "two tens and five ones". We hypothesized that the explicit mapping between physical quantities (grouped dots), the arrangement of number symbols on the 5 x 10 board and on the cards, and the use of transparent numerical language would help children in Kindergarten improve their grasp of base-10 compositionality.

### ***Mind Hunter*: A Social Cognitive Game to Foster Mental State Understanding**

To assess the specific benefits of playing *Find and Move*, we randomized children to receive either this game or a game with similar procedures but no math content. We used a social-cognitive game, *Mind Hunter*, that that was evaluated in a randomized experiment with 5- to 6-

year-old children in Salvador, Brazil (Andrade et al., in prep). *Mind Hunter* aimed to train children's abilities to reason about other people's mental states by combining evidence concerning their intentions, desires, beliefs, or perceptions of the surrounding environment. Like number concepts, mental state concepts are compositional, they are referred to frequently in the language that children hear, and they often are accompanied by spatial symbols, including the gestures that accompany speech (e.g., thumbs up) and the emojis used in texting.

Mental state concepts also are critical for children's learning in school: to learn from a teacher, children must discern what lessons she intends to impart; to learn effectively, moreover, children must evaluate their own mental states of ignorance and knowledge. Finally, children progressively master the use of mental state concepts in their language during the early school years (Bartsch & Wellman, 1995). Thus, a game aimed at improving children's mental state concepts could have a positive impact on school-relevant skills, including children's sensitivity to what they do and do not know and their capacity for pedagogical communication and learning (Gweon, 2021).

Consistent with these considerations, the Brazilian children who were randomized to play *Mind Hunter*, together with a second game exercising social-cognitive abilities, showed gains on social cognitive but not math measures. As in the case of the field experiments in India, however, it is not clear whether, or to what extent, the gains shown by children were attributable to *Mind Hunter* or to the other social cognitive game, and whether the games enhanced children's understanding of, or interest in, mental state reasoning. Thus, our experiment serves as an initial test of the effectiveness of *Mind Hunter* at exercising and improving children's social cognitive skills in a short-term, home-based intervention in the U.S.

## Overview of the Present Experiment

Children were quasi-randomly assigned (equating for age, gender, and time during the school year) to play either *Find and Move* or *Mind Hunter* at home with their caregiver (see below for details). Before and after playing their assigned game for 2-3 weeks at home, children completed a battery of lab-based assessments measuring their numerical, social-cognitive, and domain general cognitive skills. We hypothesized that children who played *Find and Move* would show greater improvement in their numerical skills than children who played *Mind Hunter*. We also hypothesized that children who played *Mind Hunter* would show a greater improvement in their social-cognitive skills than children who played *Find and Move*. The methods and analyses were preregistered after data collection was completed but before looking at any data. See [https://osf.io/7p9q4/?view\\_only=c034e94d50d8467b9b28d85648db740e](https://osf.io/7p9q4/?view_only=c034e94d50d8467b9b28d85648db740e) for the preregistration.

## 4.4 Methods

### Participants

48 children (24 females, mean age= 5.6 years, range= 5.0 to 6.2 years) were included in our final sample of participants. Participants were recruited from the Greater Boston area by phone or email from a database of families at Harvard University who had expressed interest in participating in research studies. They received a small gift (e.g. a stuffed animal or a t-shirt) for participating. Our use of human subjects was approved by the Institutional Review Board at Harvard University. An additional 12 children were recruited but excluded due to failure to return for the second visit ( $n=7$ ), failure to play their assigned game at least 6 times ( $n=3$ ), lack of fluency in English ( $n=1$ ), or experimenter error ( $n=1$ ). The sample size and exclusions were specified before beginning data collection. There were equal numbers of children ( $n=24$ ) in the two treatment conditions. Most children were enrolled in Kindergarten. The remaining children,

recruited during the summer, had just finished Kindergarten and were scheduled to begin first grade in the fall.

## **Design**

This study consisted of two 90-minute lab sessions separated by 2-3 weeks of home-based game play. Children were assigned to one of the two games conditions--*Find and Move* or *Mind Hunter*—by a Research Assistant who knew the ages, genders and dates of testing of the children who had previously been tested in each condition but was unaware of which of the two conditions involved the number game. Assignments to conditions were made so as to match the participants in the two conditions by age, gender, and the time of year of their participation (a proxy for their amount of exposure to numerical concepts in school).

In the lab sessions, all children completed two versions of the assessment battery (A and B) measuring their numerical reasoning, mental state reasoning, and two domain-general cognitive skills (language learning and executive function). The two versions of the assessments were matched in difficulty but consisted of different items. Half the participants in each treatment condition received version A at pretest and version B at posttest; the others received the two versions in the reverse order. Assessments were administered by an experimenter who was naïve to the child's treatment condition.

In addition to completing the assessments, children also played one level of their assigned game in the lab, in the presence of an experimenter naïve to their performance on the assessment battery. Most children received the assessments first, followed by gameplay, but if a family had played the game less than 8 times at home, the order was reversed in the second lab session: they played the game before completing the assessments.

## Games

The number game, *Find and Move*, was designed to be played untimed and consisted of a number board, 10 decks of 16 cards each, and erasable colored markers (see Figure 4.1 a-d for board and example cards). The board had numbers ranging from 1-50 in five rows (See Figure 4.1a). The cards were of three types: (i) some displayed a single number using dots grouped by tens if the number was 10 or more (hereafter, “grouped dots”: see Figure 4.1b): (ii) some displayed addition problems where both the addends were depicted using grouped dots (See Figure 4.1c), and (iii) some displayed addition problems where the first number was depicted using grouped dots and the second number was depicted using Arabic numerals (See Figure 4.1d).

For the first type of cards (Find), children determined the number of dots and circled that numeral on the square of the board. For the second and third types of cards (Find and Move), children located the first number on the card (depicted using dots) on the board. Then children added the second number to it by moving on the board by the number of spaces that were depicted by the numerals or dot array that appeared after the plus-sign on the card. Finally, children circled the numerals on the square of the board at which they arrived. When the number of spaces to move was 10 or larger, children were taught that they could jump up the line by one space instead of moving across the board by ten spaces. See SI for a full description of the game.

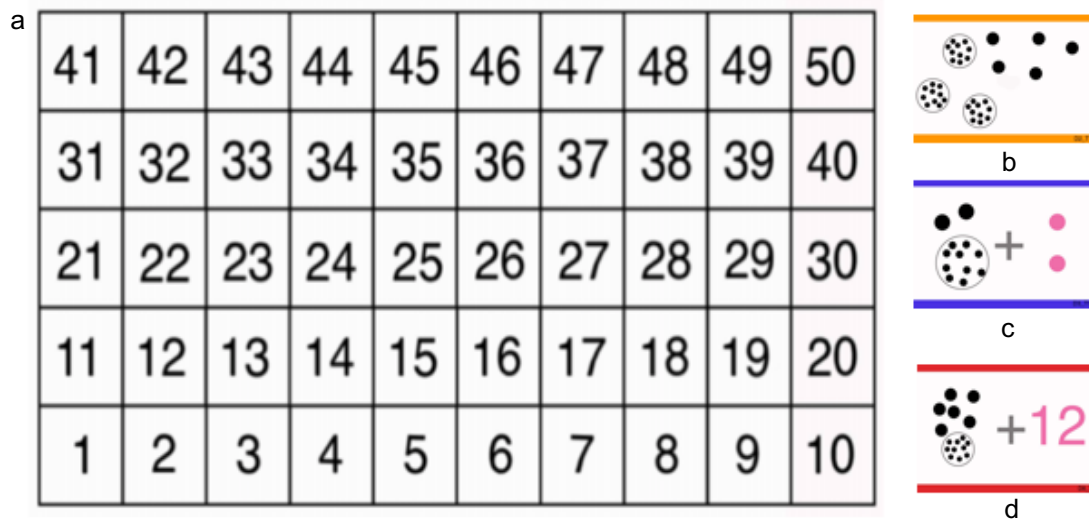


Figure 4.1: Number board (a) and examples of cards (b-d) from Find and Move. The card in (b) required children to determine the number of dots using both the 3 circles of ten and the 5 uncircled dots (35) and find that number on the board. The second card required children to first find the number in black (12) and then move by the number of dots in pink (2) to find the sum on the board (14). The third card required children to first find the number in black (16) and then move in accord with the symbolic number in pink (12) to find the sum on the board (28).

The social-cognitive game, *Mind Hunter*, was also designed to be played untimed. It consisted of 10 decks of 16 cards each and two trays (red and blue) into which the cards were sorted (see Figure 4.2 a-d for example cards). The top half of each card displayed one or more characters, depicted as having desires or goals, states of ignorance or belief, or states of perception or knowledge. The child's task was to reason about these states using depictions of events, of emotional expressions, and of visuo-spatial symbols in some combination. On the bottom half of the card, two images depicted two goal states, actions or outcomes, one drawn against a red background and the other drawn against a blue background. Children placed the card into the red or blue tray based upon the color of the right answer. To sort the cards correctly, children had to

infer the characters' mental states: their desires, beliefs, intentions, or perceptions of the surrounding environment. See SI for a full description of the game.

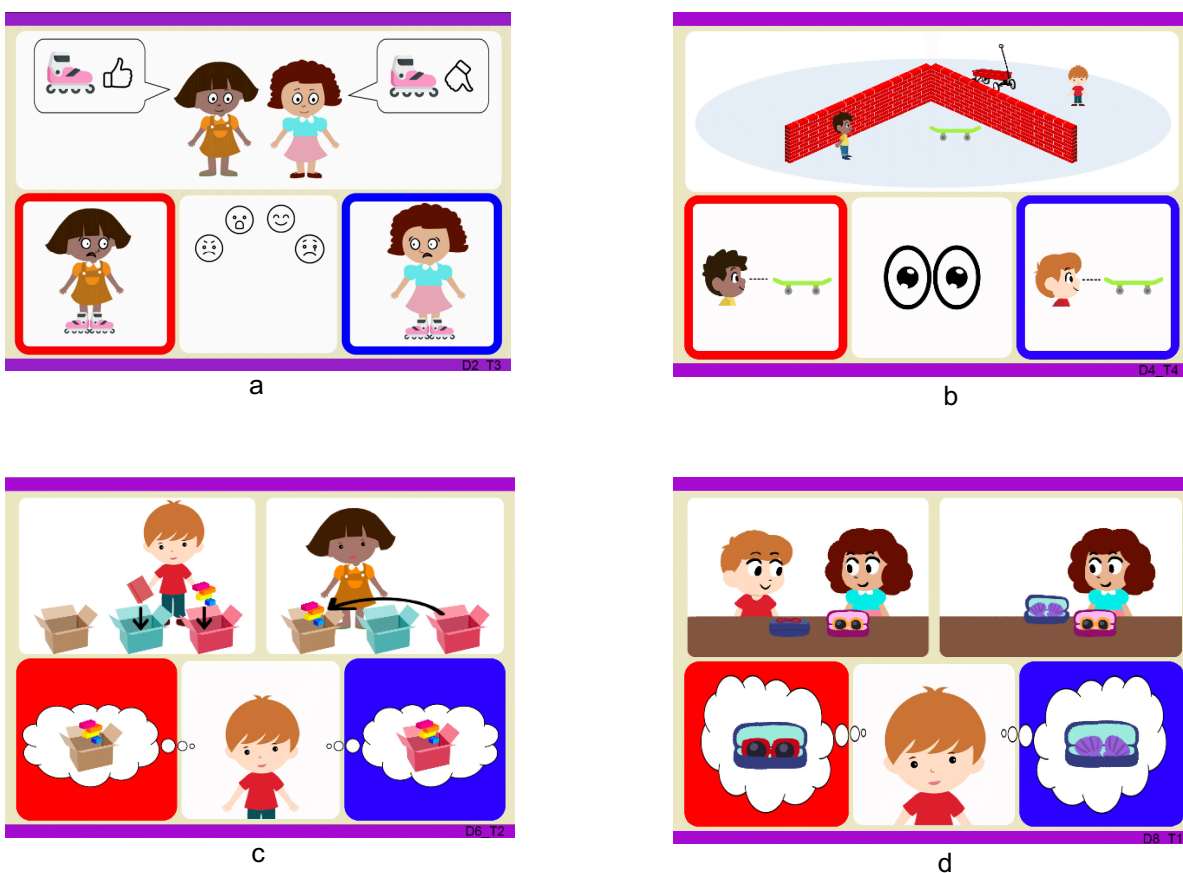


Figure 4.2. Example cards from Mind Hunter. On each card, the picture on top allows the child to infer a character's attitude or emotion (a), perception (b), or belief (c and d). Using the top picture, children must determine whether the character's mental state is represented by the red or blue picture on the bottom and sort the card into the bin of the same color. In (a), one child is unhappy skating (blue); in (b), one child can see the skateboard (red); in (c), the boy believes the object is in the box where he last saw it (blue); in (d) the boy believes the glasses case contains glasses rather than shells (red).

## Assessments

Two versions of the assessment battery (A and B), each consisting of 16 tests, were constructed (see [https://osf.io/7p9q4/?view\\_only=c034e94d50d8467b9b28d85648db740e](https://osf.io/7p9q4/?view_only=c034e94d50d8467b9b28d85648db740e) for a complete listing of all tests and questions). In each battery, 7 tests assessed children's numerical abilities, 7 tests assessed children's social-cognitive abilities, and 1 test each assessed children's executive functioning and vocabulary (for a detailed description of each test, see the SI. The order in which the assessments were presented is included in the SI. The two test batteries presented different problems across the same range of difficulty, so that children did not confront the same problems at posttest as at pretest; half the children received battery A at pretest and battery B at posttest and the half received the batteries in the opposite order. All tests were administered on a 16-inch computer screen.

The numerical tests measured children's approximate and exact numerical skills. Two tests focused on number comparison using ungrouped dots and on approximate symbolic addition, measured their approximate numerical skills. Four tests focused on exact symbolic number knowledge and required children to identify, name and compare written numbers, solve simple arithmetic problems, and map number symbols onto a linear number line. In some trials of the exact number tests, children were expected to succeed at higher rates if they understood base-10 compositionality. We highlight such trials in the SI.

The social-cognitive tests were of two kinds. Four tests measured children's abilities to interpret emotional expressions and directions of gaze in depicted faces and scenes. Three tests measured children's abilities to reason about characters' mental states: their desires, knowledge states, false beliefs, and perceptions of the environment.

The two domain-general tests of executive functioning and vocabulary were administered for two purposes: (1) to determine if these cognitive skills, at pretest, moderated the effects of games training on the numerical and social-cognitive tests; and (2) to determine if these cognitive skills were affected by training with either game, as indicated by an improvement in performance from pretest to post-test.

## **Procedure**

During the first lab session, an experimenter who was unaware of the child's condition administered one version (either A or B) of an assessment battery as a pretest. Then a second experimenter, unaware of the child's performance on the pretest, taught the child and their caregiver how to play their assigned game using one deck of the game (either *Find and Move* or *Mind Hunter*) that contained example cards from all levels of the game. Once children and parents understood how to play their game, they were encouraged to play independently with each other in the presence of the experimenter.

At the end of the first session, children and their caregivers were provided with 8 decks of the game materials to play at home for a 2-3-week period. They were asked to play each of the 8 decks of the game on 8 separate days during the home-based training period either by themselves or with a sibling. In addition to the games, they were provided with an informational brochure containing the game description and instructions.

During the home-based training period, a researcher checked in with the caregiver via email and asked them to fill out a brief survey to indicate which deck of the game they played, what time of the day they played, who played with the child, how many responses were correct, and how long it took. If it was found that families did not play the game at least 6 times within 3 weeks, they were given the option to reschedule their second lab session.

During the second lab session, an experimenter unaware of the child's condition administered a different version of the same assessment battery as a posttest. Finally, children played a deck of the game (similar to the one that they played in session 1) with their caregiver, in presence of the experimenter who had taught them their game. All the in-lab sessions were video recorded.

### **Data and analysis**

For all the tests except the Number Line Estimation, we calculated the percent correct by dividing the number of trials a child answered correctly by the number of trials presented to them (i.e., correct + incorrect + responses of "I do not know"). Practice trials and the trials that weren't presented to a child were not entered into this calculation. For the Number Line Estimation test, we calculated the scores in terms of the percent error using the following formula:  $((\text{Target Number} - \text{Child's Estimate}) / (\text{Range of number line})) \times 100$ . In order to reflect accuracy, we reverse coded the scores on each trial.

Each sub-test's score was transformed into Z-scores based on the mean and standard deviations of scores from both conditions at pretest and posttest. We then generated the two primary dependent variables: i) the number composite score, calculated by averaging the Z-scores of the 7 numerical tests and, ii) the social-cognition composite score, calculated by averaging the Z-scores of the 7 social-cognitive tests. If a child did not finish at least 50% of a test, we treated it as missing data. We used the lme4 package (Bates et al., 2015) in R (R core team, 2013) to implement linear mixed-effects models. All measures and analyses, unless specified otherwise, were preregistered prior to examining the data.

In addition to these primary measures, we coded videos of children's in-lab game play sessions, both in the initial practice session and in the final test session, to determine how much

time each child took to complete the game. We were unable to code for children's accuracy in playing the games, however, as the audio was lost due to a technical error.

## 4.5 Results

In order to determine whether playing *Find and Move* improved children's numerical concepts and skills, we conducted a preregistered analysis of the number composite scores as a function of time (pretest/posttest), condition (number/social cognition) and their interaction. We also included random intercepts for participants. We found a significant time by condition interaction ( $b = -0.33$ ,  $\beta = -0.47$ ,  $t = -2.88$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-0.80, -0.15]$ ) indicating that children who played *Find and Move* (the number game) showed a significantly greater increase in their number composite score from pretest to posttest than did the children who played *Mind Hunter* (the social-cognitive game).

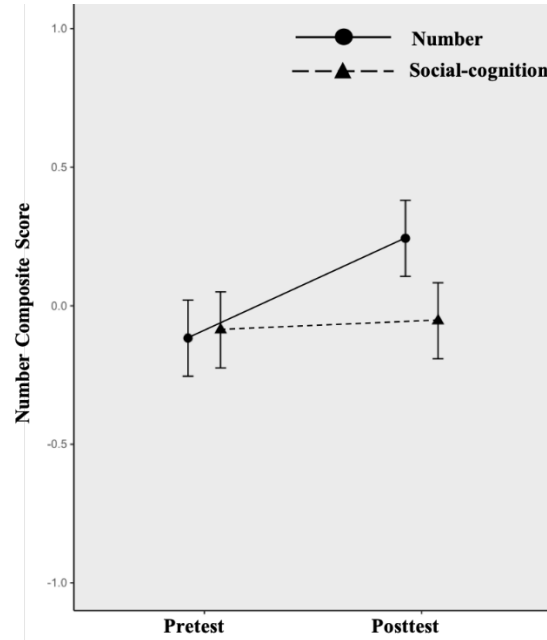


Figure 4.3: Estimates of the number composite score plotted by time (pretest/posttest) and condition (number/social-cognition). Error bars represent standard errors of the mean.

In order to determine the effectiveness of playing *Mind Hunter* towards improving children's social-cognitive concepts and skills, we conducted a preregistered analysis of the social-cognition composite scores as a function of time (pretest/posttest), condition (number/social-cognition) and their interaction. We also included random intercepts for participants. We did not find a significant time by condition interaction ( $b = -0.2$ ,  $\beta = -0.34$ ,  $t = -1.27$ ,  $p = 0.21$ , 95% CI of  $\beta$   $[-0.87, 0.19]$ ) indicating that children who played *Mind Hunter* (the social-cognitive game) showed no greater increase in their social-cognition composite score from pretest to posttest than did the children who played *Find and Move* (the number game).

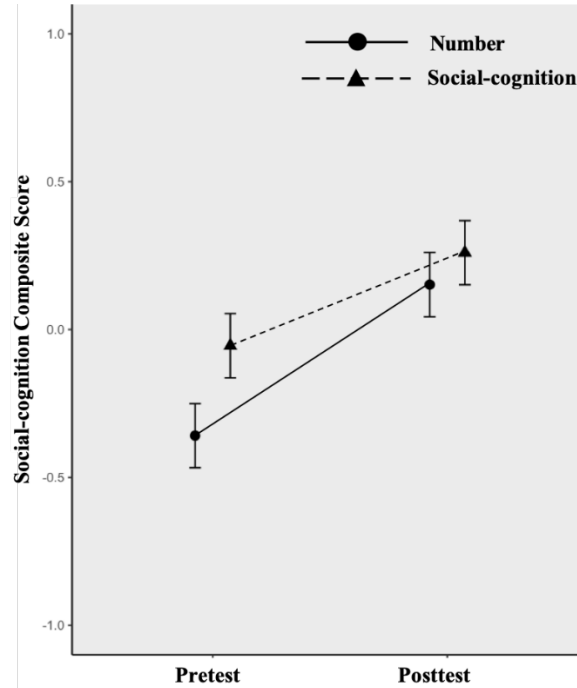


Figure 4.4: Estimates of the social-cognition composite score plotted by time (pretest/posttest) and condition (number/social-cognition). Error bars represent standard errors of the mean.

### Exploratory/non-preregistered analyses

To measure the efficiency of game play in each condition during the lab sessions, we performed dependent samples t-tests on the time (in minutes) taken to complete the practice decks. The average time declined from the first to the second session both for Find and Move (session 1:  $M = 18.58$  minutes,  $SD = 4.04$  minutes; session 2:  $M = 11.26$  minutes,  $SD = 5.52$ ;  $t(18) = 6.43$ ,  $p < 0.01$ ) and for Mind Hunter (session 1:  $M = 13.46$  minutes,  $SD = 2.92$ ; session 2:  $M = 9.42$  minutes,  $SD = 3.01$ ;  $t(17) = 5.35$ ,  $p < 0.01$ ). Therefore, children's efficiency at playing the two games improved to similar extents.

In order to determine the effects of playing the number game on children's understanding of base-10 compositionality, we created two additional composite measures. The *base-10 composite measure* included the trials from the exact numerical tests with double-digits for which

a child who lacked base-10 knowledge would be especially likely to err. The *non-base-10 composite measure* included all trials from the exact numerical tests containing only single-digit numbers and all the trials from the approximate numerical tests. *Number Line Estimation* did not enter into either composite measure. Although this analysis was not part of our initial preregistration, all decisions concerning which trials to include in the analysis were made prior to inspection of any trial-by-trial data from the experiment.

The analyses of the base-10 and non-base-10 composite scores (obtained by creating Z-scores of the average of percent correct on 29 trials and 37 trials, respectively) used the same linear mixed effects model as in the primary analyses. There was a significant time by condition interaction for the base-10 composite ( $b = -0.54$ ,  $\beta = -0.54$ ,  $t = -2.93$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-0.90, -0.17]$ ), but not for the non-base-10 composite ( $b = -0.43$ ,  $\beta = -0.43$ ,  $t = -1.76$ ,  $p = 0.08$ , 95% CI of  $\beta$   $[-0.91, 0.05]$ ). The children who played *Find and Move* showed a significantly greater increase only in their base-10 composite scores from pretest to posttest relative to the children who played *Mind Hunter*, though the results of the non-base-10 analysis trended in the direction favoring the children who played *Find and Move*.

In order to determine the differential effects of games training on the numerical and social cognitive tests separately, we analyzed the Z-scores for each of the tests using the same linear mixed effects model. Across all the numerical tests, we found a significant time by condition interaction only for one test, *Say the Number*, that required children to name numerals verbally ( $b = -0.44$ ,  $\beta = -0.44$ ,  $t = -2.64$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-0.78, -0.11]$ ): children who played *Find and Move* showed a significantly greater increase from pretest to posttest as compared to children who played *Mind Hunter*. The time by condition interaction was not significant for any of the other numerical tests (see SI for details about statistics), though the results trended in the direction

favoring the children who played *Find and Move*. Consistent with that trend, when the same analysis was applied to the composite score of all the numerical tests excluding *Say the Number*, it revealed a significant time by condition interaction ( $b = -0.31$ ,  $\beta = -0.46$ ,  $t = -2.48$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-0.82, -0.09]$ ).

The same analyses, applied to the social cognitive tests individually, revealed only one significant interaction ( $b = -1.02$ ,  $\beta = -1.02$ ,  $t = -3.2$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-1.66, -0.39]$ ), for a test that required children to reason about social scenarios verbally (see SI for details about statistics for the other social cognitive tests). This pattern was opposite to the predicted direction: the children who played *Find and Move* showed a significantly greater increase from pretest to posttest on this social cognitive test than did the children who played *Mind Hunter*.

## 4.6 Discussion

### A game to enhance children's readiness for learning mathematics

Given the importance of understanding base-10 compositionality for learning mathematics in school, we evaluated a number game, *Find and Move*, when played at home, on kindergarten children's numerical concepts and skills. Children who played the number game showed an improvement in their school-relevant numerical skills and base-10 understanding, relative to those who played a social cognitive game, *Mind Hunter*. This suggests that skills important for learning mathematics in school can be enhanced through games administered by children's families in their homes.

Since the improvements in numerical skills were specific to the children who played the number game, they are not attributable to the motivational or domain-general cognitive benefits of playing a game with a family member. However, the improvement could be attributable, in part,

to critical differences between the games. Only the number game was a board game. Therefore, it might have been more engaging to children than *Mind Hunter*, a card-sorting game.

Furthermore, it is possible that the caregivers who played the number game with their children were spontaneously engaging in practices, such as talking about the contents of the cards, due to their belief that numbers are directly relevant to the school's curriculum. In contrast, although the caregivers who played the social cognitive game with their children reported playing it for a similar amount of time as the number game (see SI), their beliefs in its direct relevance for children's learning in school might have been weaker. Future work might compare *Find and Move* to another board game whose content is likely to be believed by parents to be more directly relevant to the school's curriculum (such as reading or geometry).

In exploratory analyses, we found that children who played the number game showed an improvement on a composite measure of the tasks and trials that were more likely to be solved accurately if children used their understanding of base-10 compositionality. We did not find this improvement on a composite measure of the tasks and trials that were not likely to be solved using base-10 reasoning, either because they focused on approximate number representations or because they used only single-digit numbers. These findings thus provide preliminary evidence that the number game specifically improved children's understanding of base-10 compositionality. They raise the question of whether children who play this game in kindergarten will show more meaningful learning of mathematics in first grade, where they will encounter arithmetic algorithms that depend on the base-10 structure of numbers.

The present study also has several limitations. First, we trained children's base-10 compositional number concepts using multiple representations: a spatially organized board that aligned numbers by ones and tens, grouped dots, and number words. Each representation made

some aspect of the base-10 structure more transparent. Based on our study alone, it is not possible to determine if young children's understanding of base-10 compositionality benefits from all three representations. Future work might test the differential effects of each kind of representation on children's numerical understanding.

Second, due to our small sample size and large assessment battery, we lacked the power to detect effects on most of the individual numerical tests, except for the test that required children to recognize multi-digit numbers and name them verbally. Since our findings rest on composite measures comprised of multiple tests and trials a replication with a larger sample size might provide stronger evidence for the individual skills and concepts that such a game can improve.

Third, we tested the effectiveness of just one kind of spatial representation of numbers on a board. In our game, the numbers started from one and the tens (10, 20, 30, etc.) were all right-aligned, as in previous work that used board game interventions (Laski & Siegler, 2014). It is possible that children would benefit more from a spatial representation in which the tens are left-aligned. Such a board would start at 0 instead of 1; in order to find a number like 23, a child would begin at 0 and move 2 spaces upward to 20, and then move 3 spaces to the right to reach 23. This arrangement might help children better grasp the compositional structure of the numbers from 1 to 99, because the ten rows of the board will represent the number of tens in each number, whereas the ten columns will represent the number of remaining ones. Future testing could determine what kind of spatial representations best help children understand base-10 compositionality of numbers.

Finally, *Find and Move* focused only on the numbers from 1 to 50. It has been proposed, in accordance with the tolerance principle in linguistics (Yang, 2016) that English-speaking children may need to learn more than 50 number words before the productive rule for forming new number words will be apparent to them. This is because of the presence of number words, like

“eleven,” that obscure the base-10 structure. It is possible that training with a *Find and Move* game will be more effective if the board extends to higher numbers.

Given the brevity of this home-based intervention, the present study was not designed to improve children’s learning of math in the long term. Its findings complement the findings of large-scale educational interventions, including the two interventions for which the games tested here were developed. In the large-scale field experiment in India for which *Find and Move* game was developed, the math games intervention consisted of Find and Move and three other math games. Although that intervention had a positive impact on children’s math learning, it did not reveal which games contributed to this impact. Since the children in the present study played *Find and Move* by itself, the present experiment suggests that this game likely contributed to children’s math learning in the school context in India.

#### **A game to enhance children’s mental state reasoning**

In the field experiment in Brazil, for which the social cognitive game was developed, the social cognitive intervention consisted of two games: *Mind Hunter* and a second social-cognitive game, *Find the Emotions*, training children’s emotional intelligence. Because the intervention had a significant impact on Brazilian children’s social cognitive development (Andrade et al., in prep), the absence of an effect of *Mind Hunter* on children’s mental state reasoning in the present experiment raises the possibility that *Find the Emotions* played a larger role in enhancing Brazilian children’s mental state reasoning.

Our experiment does not reveal, however, whether *Mind Hunter* failed to impact on children’s mental state reasoning because of its content or its format, as a card game rather than a game played both with cards and with a board. Moreover, it is possible that children’s social cognitive abilities are less affected by short-term games-based interventions than are their

mathematical abilities, because children naturally are bathed in social experience, minimizing the effects of any 8-session intervention.

### **A Research Synergy Between Education and Developmental Cognitive Science**

In 1965, the great developmental psychologist, Eleanor J. Gibson, described her extensive research on reading (later published in Gibson & Levin, 1975) to the readers of *Science*: “[It is] my firm conviction that good pedagogy is based on a deep understanding of the discipline to be taught and the nature of the learning process involved....An analysis of the reading task—its discriminatory and decoding aspects as well as the semantic and syntactical aspects—tells us *what* must be learned. An analysis of the learning process tells us *how*. The consideration of formal instruction comes only after these steps, and its precepts should follow from them.” (Gibson, 1965, p. 1072). Randomized field experiments in education, based on basic research in developmental cognitive science, provide a concrete means both for testing the validity of Gibson’s convictions and for improving educational practices.

In cognitive science, however, no single method is perfect: Advances in this interdisciplinary field occur when different methods, with complementary strengths and limitations, are brought to bear on the task of evaluating and improving pedagogical practices. The present study focused on one problem that good teaching of mathematics must address: fostering children’s understanding of the base-10 compositional structure of the number system.

The present study aimed to address a limitation of any large-scale, randomized intervention in children’s schools: to be effective, such an intervention must be long in duration, because children come to master the logic of primary school mathematics slowly. If it is long, however, it also must be multifaceted, both because multifaceted interventions provide richer opportunities for learning and because they are better able to maintain children’s interest over the long term. When

a multifaceted, randomized intervention succeeds, therefore, its success does not reveal which components of the intervention were effective, or what specific effects they produced. As a result, it is not clear which aspects of such an intervention should be built on to improve mathematics education even further.

The present findings illustrate one manner in which large-scale, long, field experiments in education and small-scale, brief experiments in developmental cognitive science can synergize: lab-based research can test different components of a large-scale intervention individually and determine their effectiveness over the short term.

The present experiment drew on two large-scale educational interventions that successfully enhanced children's skills in two domains that are relevant to education: one in the domain of mathematics and the other in the domain of social-cognitive reasoning. The present experiment provides evidence that one of the four games used in the successful, school-based math intervention in India effectively enhanced children's understanding of the base-10 structure of the number system in the U.S. The experiment also suggests that one of the two games used in the successful, school-based social cognitive intervention in Brazil failed to enhance children's understanding of mental states of belief and desire when presented by itself in the U.S. In light of both these findings, future lab-based experiments, evaluating the effectiveness of the other games used in the large-scale interventions, may be especially rewarding.

### **Home-based Interventions to Foster Children's Learning in School**

The present study was undertaken to probe why previous school-based interventions were effective, so as to inform future school-based math curricula. In retrospect, however, we suggest that home-based interventions also might play an important role in enhancing children's learning in school. Although this research was planned and conducted prior to the onset of the school

closures that occurred in response to the COVID-19 pandemic, our findings have new relevance, as they show that making educational games available to families in their homes can help foster children's skills at the foundations of learning mathematics. Given this finding, such games may be useful in preventing the loss of mathematical skills that children are likely to suffer if future school closures occur. They may also serve to prevent declines in children's math skills over the summer months when schools are closed.

Children's home environments correlate with their school-relevant learning (Gunderson & Levine, 2011; Levine et al., 2010; Niklas & Schneider, 2014) and its importance was foreseen by Eleanor Gibson, as she worked to enhance children's learning to read. Three decades ago, she wrote of "... the wisdom of providing young children with an environment supplied with plenty of literature of the kind one wants them to be curious about, and letting them explore a variety of reading material with adults who are ready to answer questions and provide role models by reading themselves and to their children....One doesn't *teach* reading to someone, rather that person *learns* to read. We can only learn to read by engaging in reading, and we have to do it for ourselves. Motivating a child to...read is one place where we often fail." (Gibson, 1991, pp. 437-438).

Motivating a child to engage with mathematics is also a domain in which a richer home environment may foster children's school learning. The introduction of math games into children's homes, informed by research in developmental cognitive science, might provide children with more materials to explore and be curious about, and it might provide families with further means to answer their questions and aid their learning. If so, then educational interventions in children's homes may enhance their learning in school and help us establish the causal impacts of the home environment on their academic progress.

## Chapter 5

# Compositional Numerical and Reading Skills

Home-based games training compositional reasoning improved 6-7-year-old children's numerical skills but not their reading skills

### 5.1 Introduction

In order to learn math and reading efficiently in school, children must grasp the compositional systems that underlie them. In order to learn arithmetic for example, they must grasp how numbers combine to form larger numbers. The digits from 1-9, along with some power of the base number 10, can generate all the natural numbers: for example, “35” is “ $3 \times 10 + 5$ ”. Learning the combinatorial rules of numerical concepts can help children become efficient calculators as they use number in everyday life.

Similarly, in order to learn to read, children must grasp how speech sounds combine to form words, such that “bat” can be seen as a combination of “b” + “at”. Learning the combinatorial rules of written text can help children become efficient readers, who can combine symbols for the same speech sounds to represent different words, phrases, and sentences. In the present study, we evaluated two games, one focused on the base-10 compositional number system and the other focused on the compositionality of the English alphabet, in order to help 6-7-year-old children improve their school-relevant numerical and reading skills.

The base-10 number system and reading are both compositional systems that take years for children to master. They contain various symbols and rules for combining the symbols that might

not be explicit to children without clear instruction. Therefore, children might especially benefit from fun games that help them engage with and learn these concepts.

#### **Building on the numerical game *Find and Move* from Chapter 4**

Children’s learning of the numerical concepts that underlie the base-10 compositional number system benefitted from the game that we evaluated in an experiment in Chapter 4. In order to address some of the limitations of that experiment, we made changes to the *Find and Move* game that aimed to train children’s understanding of base-10 compositionality. We describe the changes below.

First, we included numbers up to 99 in the current version of the game. In Chapter 4, the game used numbers only until 50. Though learning larger numbers beyond 50 might pose challenges for children, there might be a benefit to getting exposed to them early on. This is because with larger numbers, children get exposed to many more number words that reveal the logic of base-10 compositionality more transparently. For example, the earliest double-digit number words in English (e.g., “eleven” and “twelve”) obscure their composition into a set of ten and ones. The earlier decades up to 50 (“twenty”, “thirty”, “forty”, and “fifty”) also do not transparently convey their composition with the single-digit number words (“two”, “three”, “four” and “five”) and the morpheme for ten (“ty”). Going beyond “fifty” will expose children to the more transparent decades in English (“sixty”, “seventy”, “eighty”, and “ninety”), and this might help them gain better insights into how sets of tens and ones compose to form double-digit numbers.

Second, we changed the structure of the numerical board in the current version of the game to make the spatial cues to base-10 compositionality more explicit. In Chapter 4, the game used a board that started with “1” and the decades (i.e., 20, 30, 40, and 50) were aligned to the right of

the board. The board's structure was similar to past work that used number boards to help children learn numerical concepts (e.g., Laski & Siegler, 2014). Aligning the tens to the right, however, means that numbers in the same decade (e.g., 20 and 21) would appear in different rows of the board. Therefore, in the current version of the game, we aligned the decades to the left and the board began from 0, though the symbol "0" itself was not present on the board (see Figure 5.2a). This arrangement of the board might better help children grasp the compositional structure of the numbers from 1 to 99, as all the numbers in a decade appear in the same row. The rows of the board represent the number of tens in each number, whereas the columns represent the number of remaining ones. Therefore, a board with left-aligned decades might help children more easily analyze a double-digit number into its constituent tens and ones.

Third, in the current version of the game, when presenting arithmetic problems, we represented one addend using grouped dots (i.e., a set of ten dots were enclosed inside a circle) and the second addend using Arabic numerals. In Chapter 4, the game used cards where half the time, the second addend was represented using grouped dots and the other half, it was represented using Arabic numerals. If the second addend is always represented using Arabic numerals, children will get more practice with decomposing number symbols into their constituent tens and ones.

Fourth, in the current version of the game, we aimed to teach children how to represent the same number (ten) in two different ways (*one ten* and *ten ones*) in order to aid faster mental calculation. Switching between these representations is also important for understanding the algorithms of arithmetic such as carrying and borrowing. In the current version of the game, in some advanced levels, when adding the number 9 (or any two-digit number with 9 in the ones place) to another number (e.g., 25), we show children the equivalence of moving by 9 spaces to the right on the board to reach the answer (i.e., 34) or moving up by ten (to 35) and then moving

one space to the left (to 34). This way, they might learn to represent ten as “one ten” when moving up the board to add ten and also as “ten ones” when subtracting one from ten. Practice with switching between multiple representations of ten might help children become efficient in making use of the base-10 compositional number system while performing arithmetic.

Finally, we designed the current version of *Find and Move* to be more fun and engaging by making it a two-player game whose winner is determined based on chance. It was important to make it a game of chance as most children played the game with their caregiver(s) who are already skilled in arithmetic. The current version of *Find and Move*, was designed in the format of a children’s game called War. In this game, children and their caregiver each get a pile of cards. They flip one card over at a time and solve the addition problem on it using the number board. The player with the higher number as the final answer wins the round and gets to keep both players’ cards. They repeat this procedure with all the cards in their initial piles and at the end, the player who accumulated the greatest number of cards wins.

Since we redesigned the game to help children grasp base-10 compositionality more readily, we also included assessments that more directly measured the base-10 compositional skills that children might learn while playing the game. For example, we included a measure of near transfer that assessed their grasp of the structure of the base-10 number board that we used in the game. Since we changed the structure of the board from previous studies and also from popular children’s games that use number boards (e.g., chutes and ladders) we wanted to make sure that children were still able to learn the structure of our new board. We also included other measures of far transfer that assessed their ability to represent two-digit numbers as sets of tens and ones and perform mental calculation using the tens efficiently (e.g.,  $25+20$  and  $25+29$ ).

### **A novel reading game: *Bingo***

Since reading is also taught in school and is crucial for children's academic success, we designed a novel reading game called *Bingo* that focuses on the compositionality of the English alphabet. The Bingo game helps children create mappings between spelling (graphemes) and sound (phonemes).

Learning to read proficiently, i.e., the ability to go from written text to meaning, takes years for children to master. Children go through three stages when learning to read (see Castles et al., 2018 for a review). The first stage is called pre-alphabetic, where children have no insight about how graphemes map on to phonemes. They are only able to “read” a few words by relying purely on visual cues and possibly rote memorizing a label that applies to a certain combination of letters (e.g., recognizing “Coca-Cola” as it is written on a can). Then they enter the partial-alphabetic stage where they begin to make links between graphemes and morphemes. Reading is still laborious in this stage as reading speed is determined by word length. Finally, they reach the full-alphabetic stage where they can decode unfamiliar spellings and word length does not play as big of a role in reading speed as familiarity with the words.

Our game, *Bingo*, is targeted at children who are in the partial-alphabetic phase. That is, they are starting to develop the insight of how graphemes represent the phonemes in their language. In order to help children map graphemes to phonemes, we designed several 5x5 reading boards. Each 5x5 board contained 25 words (real words and nonsense words) in them (see Figure 5.3a for an example of a board). Crucially, the boards were organized in manner where all the words within a column rhymed and all the words within a row had the same beginning sound. Children were required to find a word on the board and they could do so easily by analyzing each word into its beginning and ending sound.

Several past interventions have aimed to help children map sounds to spelling (e.g., Kyle et al., 2013; Potier Watkins et al., 2020; Potier Watkins & Dehaene, 2024). For example, one game called the Graphogame has multiple levels that help children map graphemes (e.g., “at” or “bat”) to their respective phonemes. However, to our knowledge, our game is the first to use a board to help children make use of the systematicity of rhyming words within a column and alliteration within a row when finding words.

*Bingo*, like the number game, was also directly modelled after a children’s game. Children and their caregivers each get their own Bingo board. The boards contain mostly similar words/nonsense words but in different positions. A word/nonsense word is played to them and they each need to find it on their board. The player who first gets 5 words in a straight line (either a row, column, or along the diagonal) wins.

### **Overview of the present experiment**

In the current experiment, we evaluated the effectiveness of the new version of *Find and Move* in improving children’s numerical skills by comparing it to a board game that focused on reading. Children were quasi-randomly assigned (equating for age, gender, and time during the school year) to play either *Find and Move* or *Bingo* at home with their caregiver. Before and after playing their assigned game for 2-3 weeks at home, children completed a battery of assessments measuring their numerical and reading skills via Zoom. We hypothesized that children who played *Find and Move* would show greater improvement in their numerical skills than children who played *Bingo*. We also hypothesized that children who played *Bingo* would show a greater improvement in their reading skills than children who played *Find and Move*. The methods and analyses were preregistered prior to complete data collection and examination of any data. See [https://osf.io/hr2ap/?view\\_only=448a6666e41c4185904ba43f216cf208](https://osf.io/hr2ap/?view_only=448a6666e41c4185904ba43f216cf208) for the preregistration.

## **5.2 Methods**

### **Participants**

64 children (mean age=6.5 years, range= 6.0-7.5 years; 36 females and 28 males) were included in our final sample of participants. Most participants were recruited by email from a database of families at Harvard University who had opted into being contacted to participate in research studies. Additional participants were also recruited via flyers posted on social media and in local libraries and via Children Helping Science. All participants resided in the United States of America. Participants received a 10\$ Amazon gift card for participating in the study. Our use of human subjects was approved by the Institutional Review Board at Harvard University.

An additional 35 children were recruited but excluded from the final sample due to failure to return for their second session ( $n=26$ ), failure to play their assigned game at least 6 times ( $n=8$ ), and lack of a video recording ( $n=1$ ). The sample size, exclusion criteria, and analyses were preregistered (see [https://osf.io/hr2ap/?view\\_only=448a6666e41c4185904ba43f216cf208](https://osf.io/hr2ap/?view_only=448a6666e41c4185904ba43f216cf208)) prior to examining any data. In addition to our primary analyses, exploratory intent to treat analyses were performed ( $n=72$ ) which included participants who completed both sessions but played the games fewer than 6 times at home.

### **Design**

This study involved two 90-minute long online sessions (facilitated by Zoom), separated by 2-3 weeks of at-home gameplay. Participants were assigned to one of two gameplay conditions, reading or number, prior to their first session by a research assistant with access to the child's age, grade, and gender in order to approximately equate these variables between the two conditions.

During both the online sessions all participants completed a battery of assessments designed to measure their pre and post intervention reading and numerical abilities. Two versions

(VA and VB) of an assessment battery, matched for difficulty, were used. Half of the participants in each condition received VA in the first session and VB in the second, and the other half received them in the reverse order. The experimenter that administered the assessments with the child was naive to the condition they were in. Another experimenter, naive to performance in the assessment battery, taught the child and their caregiver how to play their assigned game in session 1 after the assessments and facilitated one round of gameplay in session 2 prior to the assessments.

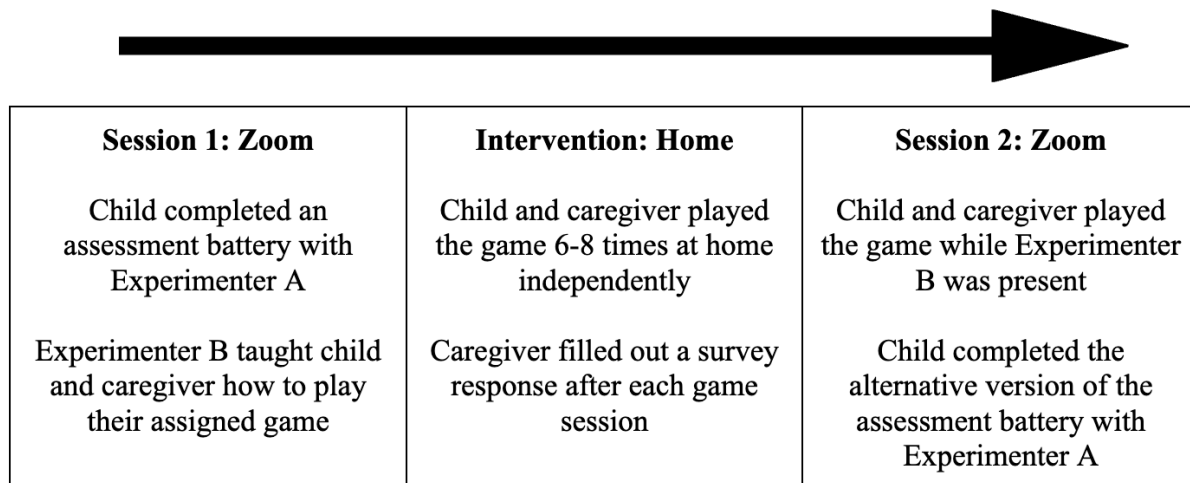


Figure 5.1. Overview of the study design

## Games

Participants in both conditions were mailed a box of materials to use for home-based gameplay prior to their participation in session 1. Each box contained the game materials (described in detail below), a sticker tracker sheet along with stickers, and an informational brochure with instructions.

## Number Game

Participants in the number condition played *Find and Move* at home, which included 3 number boards, 8 decks of cards with 30 cards each (along with 2 practice cards at the beginning

of each deck), and two game pieces (of different colors). Participants received 2 additional practice decks, which were played during the Zoom sessions in the presence of an experimenter.

There were two types of cards in the decks: (i) some displayed a single set of dots, which were ungrouped if the number was below 10. When the numbers were above 10, sets of 10 dots were grouped with a circle around them (hereafter, “grouped dots”) (See Figure 5.2b for an example) and (ii) some cards contained addition problems where one of the numbers was displayed as a set of grouped and ungrouped dots, and the second number was displayed as an Arabic numeral (See Figure 5.2c and 5.2d for examples). Decks 1-8 progressed in difficulty with each deck each introducing a new skill. Each deck included some cards requiring the use of the new skill introduced in that deck, and some cards requiring skills from previous decks, allowing participants to continue practicing previously learned skills while learning new skills. Both practice decks (played in the presence of an experimenter via Zoom) contained cards requiring each skill introduced in decks 1-8.

Each number board had spaces for numbers ranging from 0-99 in rows of 10, with the decades (e.g. 10, 20) aligned on the left side of the board (first square in each row), and the last number before a new decade (e.g. 19, 29) aligned on the right side of the board. The first square on the board (representing “0”) was left empty for the game tokens to be placed (see Figure 5.2a for an example). Of the 3 number boards, one had all the numbers filled in, one had 23 numbers throughout the board filled in with the rest of the squares being empty, and one had 15 numbers throughout the board filled in with the rest of the squares being empty.

*Find and Move* was played using similar rules to the card game called “War”. To play, children and the second player (caregiver/parent/sibling) each started with their game pieces on the blank “start” square on the board. Then they randomly distributed cards from a deck between

themselves. For each card, both players moved their game pieces to the number represented on their card. Whichever player's number was larger (and therefore had their game piece further along the board for that round) would keep both their own card and the other player's card and add them to their "win" pile. For cards depicting only a set of dots (e.g. in deck 1), players counted the number of dots on their card, and moved their game piece to that square on the board. The experimenter encouraged children to count and move "by tens" to find the right spot on the board, when applicable (e.g. to find 32, move the game piece up past 10 and 20 to get to 30, then over by 2 spaces). For the second type of card, players first counted the number depicted by a set of dots on the board. Then they added the second number by decomposing it into tens and ones, and then moving their piece by the amount of spaces corresponding to the second number. For all the cards, children wrote the numbers in the blank spaces at the bottom of the card.

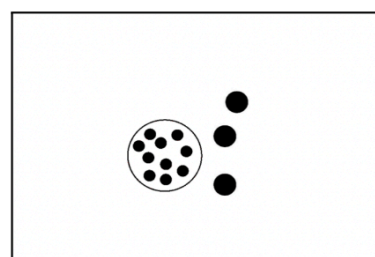
Deck 1 only contained a single set of grouped dots from 1-99 and required no addition. Decks 2-8 contained addition problems with one number represented as grouped dots (ranging from 1-98) and the second number represented as Arabic numerals. The sum was always below 99. The decks varied in difficulty based upon the type of Arabic Numerals used for the second number. Deck 2 contained single-digit numerals (1-9) as the second number. Children moved spaces to the right to add these numbers to the initial number.

Deck 3 contained multiples of 10 (e.g. 10, 20, 30, and so on), and children were taught that they can add tens by jumping up spaces on a line (e.g., solving  $32+20$  by finding 32 on the board, then moving their game piece up by two rows to get to 52). Deck 4 contained double-digit numbers that were not just multiples of tens but also had a ones digit (e.g. 21, 22, 23, and so on). This meant children had to jump rows to move by tens, then move their pieces to the right to add the ones.

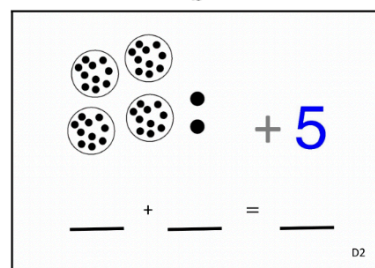
Deck 5 contained double-digit numbers that required “carrying”. For these cards, children had to jump rows to move by tens, then move to the right to add ones, and turn the corner to get to the next row when crossing decades. Deck 6 introduced cards where the second number ended with 9 in the ones place, and children were introduced to the strategy of finding the final number by moving by tens then backward by one (e.g. solving  $13+29$  by moving the game piece up 3 rows from 13 to get to 43, and then back one, to get to 42). Decks 7 and 8 included cards of the kinds featured in decks 2 through 6, but were each played on the two number boards with only some of the squares filled in.

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 |
| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
| 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
| 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|    | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |

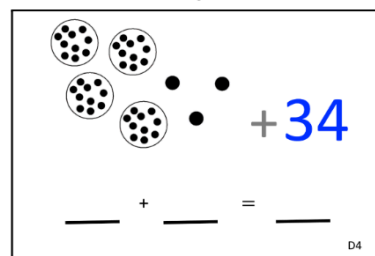
a



b



c



d

Figure 5.2. Number board (a) and examples of cards (b-d) from Find and Move. The card in (b)

required children to count the dots by combining the number of grouped dots (1 circle with 10 dots) with the ungrouped dots (3 dots) and then find the number (13) on the board. The second card (c) required children to count and find the first number depicted using dots on the board (42) and then move by the number of spaces depicted using the Arabic numeral (5) to find the sum (45) on the board. The third card (d) required children to count and find the first number depicted using dots on the board (43) and add the second number (34) by decomposing it into its respective tens and ones (three tens and four ones). To add 34, they moved their game piece up by three rows and then across by four spaces to find the sum (77).

### **Reading Game**

Participants in the reading condition played *Bingo*. The materials for *Bingo* included 8 decks of cards with 3 pairs of cards in each deck and audio recordings of words to accompany each card. Participants also received 2 practice decks that were to be played during the Zoom sessions in the presence of an experimenter. Each card had a *Bingo* board on one side, and two sentences on the other side. Each *Bingo* board was a 5x5 board with squares that either contained a real word/nonsense word or was blank (see Figure 5.3a for an example). To the left of the board, aligned with each row, was a column of “beginning sounds” (each was either one or two letters long). Below the board (aligned with each column), was a row with “word endings” (each was 2 to 5 letters long). This way all the words within a single column rhymed, and all the words within a single row contained the same beginning sound. On the back of each card, there was one sentence with words from the *Bingo* board, and another sentence without words from the board.

*Bingo* was played following similar rules to the classic Bingo game, where players crossed out spaces on their board when the word (real or nonsense) in a space was called via the audio provided to families. The first player to cross out all 5 spaces in a row, column, or diagonal said

“Bingo” and won the round. In this version of *Bingo*, each player began with a card in front of them, with the board side facing up. Then, the caregiver started playing the provided audio using their computer. First, the audio sounded out the “word endings”, and children and their caregivers pointed to each of them on the board to get familiar with the sounds. Then, the audio sounded out the words (real or nonsense) in a random order, and players looked for each word on their board. Children were encouraged to locate the target word by using the beginning sound and ending sound. When a word that was called wasn’t written on the board, children had to identify where it would go based on the beginning sound and ending sound and then write it before crossing the word out. After a player won the round, players turned their cards around to see the sentences on the back of the card. Children were then asked to identify which of the two sentences had words from the board they just played with and then read that sentence out loud. Families were also encouraged to make their own sentences with words from the board. Players repeated this procedure with every card in the deck.

The decks were designed such that, when played in order, they progressed in difficulty and introduced more complex words. The boards in decks 1 and 2 contained 3-letter words with short vowel sounds (e.g. “not” and “sep”). Deck 3 contained 4-letter words with long vowel sounds (e.g. “note” and “lice”). Deck 4 contained words with short vowel sounds and double consonants (e.g. “stutter” and “clutter”). Deck 5 contained 2 syllable words with long vowel sounds (e.g. “gater” and “slaper”). Deck 6 contained all types of words introduced in the previous decks. In deck 7, words containing combinations of vowels (e.g., “seam” and “pain”) were introduced. Finally, deck 8 introduced words with silent sounds (e.g., “thought” and “knit”). Like the number game, the reading game also included word structures from the previous decks in subsequent decks in order to help participants to continue practicing previously learned skills while learning new skills.

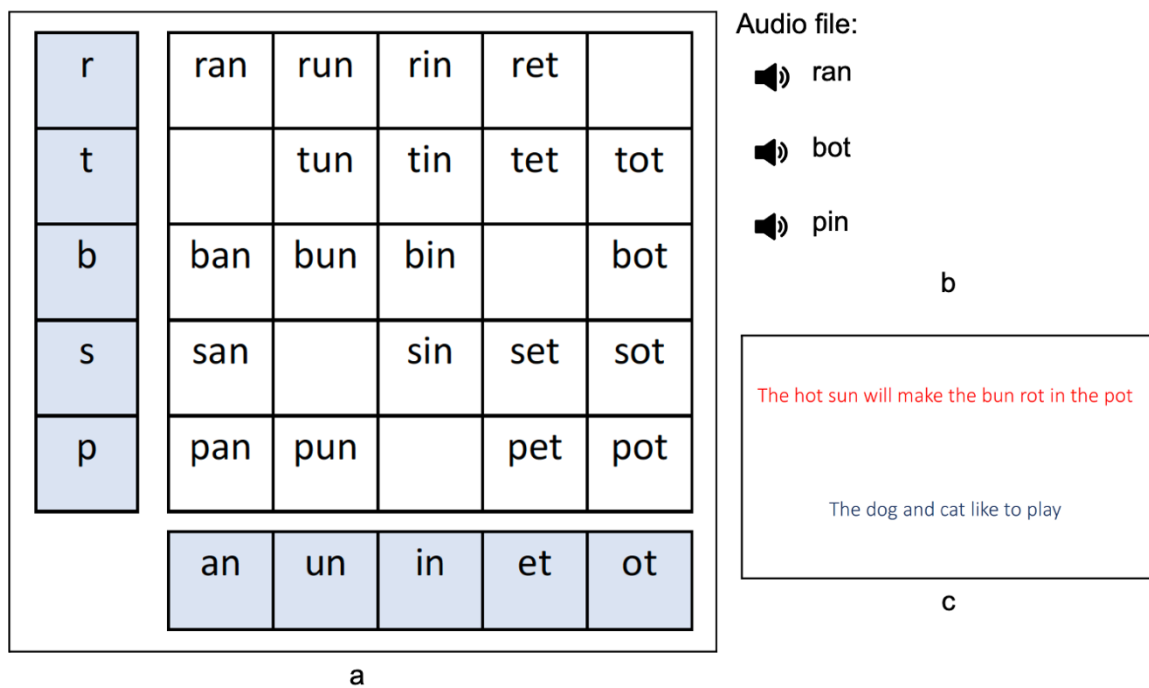


Figure 5.3. A Bingo board (a), example of audio files that accompanied the board (b) and two sentences on the reverse side of the Bingo board (c). The audio file was played and children had to locate the words on the board by utilizing the beginning sound of each word from the blue column on the left and the ending sound of each word from the blue row at the bottom. When the word was missing from the board (e.g., ‘pin’) children followed the same process to locate the square on the board where ‘pin’ should go and then wrote it in. Once complete, the reverse side of the card (c) required children to choose the sentence that contained words from the Bingo board (a) that they just played with and read it out loud.

### Assessments

An assessment battery was administered by an experimenter via Zoom using a Qualtrics survey before and after the home-based gameplay. The experimenter asked the questions verbally and participants indicated their answers either verbally, or by pointing with their mouse using the remote control feature on Zoom (numerical tests 2 and 5). Some participants were unable to use

remote control on Zoom due to their devices being incompatible with the feature, or the child being unable to control the mouse. In such cases, to complete numerical tests 2 and 5, the child indicated their answer by pointing to it with their finger, and the caregiver told the experimenter where the child was pointing.

Two versions of the assessment battery (A and B) measuring numerical and reading skills were constructed. The two versions presented similar tests across the same range of difficulty, but differed in the numbers and words used, to prevent children from seeing the exact same questions at pretest and posttest. In each condition, half of the participants were given version A at pretest and version B at posttest, and the other half completed the assessment versions in the opposite order. Each assessment version included 6 numerical assessments and 5 reading assessments. The numerical and reading assessments were presented in an alternating order.

All numerical tests were symbolic (i.e., they used number words and symbols). One test assessed participants' approximate arithmetic skills. The rest tested their understanding of symbolic numerals, exact arithmetic skills and base-10 understanding. There were 6 numerical assessments are described below:

1. Approximate Symbolic Addition (1 practice trial, and 16 test trials), which assessed children's ability to compare the sum of two numbers to a third number and identify which was larger approximately. Participants were shown a picture of a girl and a boy side by side, each with either one or two bags of candy below them. The bags each had an Arabic numeral which represented the number of candies inside the bag. Participants were asked to determine whether the girl or boy had more candy, by figuring out if the sum of candies in the bags belonging to the character with two bags, was more or less than the number of candies in the other character's single bag. The numbers on each bag were either single-

or-double-digits, and the total number of candies of the character with two bags was always below 100. The questions were forced-response (either the girl or the boy had more candy) and participants gave their answers verbally. This test was untimed.

2. Point to Number (12 trials), in which an experimenter either (a) said a number-word and asked the child to find the Arabic numeral on the screen that matched it (6 trials), or (b) asked the child to find the largest or smallest number on the screen (6 trials). In all trials, there were 6 numbers on the screen for children to choose from. Across trials, numbers up to 4-digits were used. Distractors were chosen such that participants could not rely on the number of digits or on just one of the digits to determine the correct answer. For example, if the target number was 53, there were other options with 5 as the first digit (e.g. 50 and 52) and other options with 3 as the second digit (e.g. 43). Children gave their responses either by pointing with their mouse, using Zoom's remote-control feature, or by pointing with their finger (in which case caregivers communicated the answer given by the child to the experimenter). This test was untimed.
3. Exact Oral Addition (1 practice trial, 10 test trials), which asked participants to choose an exact answer, out of four options, to arithmetic problems featuring one and two-digit numbers. In each trial, there was one arithmetic problem on the screen, and above each addend was an image of an object (within a problem, both addends had the same object. Objects differed across trials). An experimenter asked the child the arithmetic problem verbally (e.g., Suppose you have 51 oranges and I give you 10 more, How many oranges would you have now?) and read out each of the four potential options, which were also displayed on the screen. Participants responded verbally. Most of the trials included addends that were either multiples of tens or were double-digit numbers with "9" or "1" in

the ones place to assess children's ability to flexibly use the "tens" to aid mental calculation. This test was untimed.

4. Base-10 task (18 trials), which comprised of three trial types (described below) featuring individual cookies and boxes of cookies. Participants were introduced to a character named Bobo who was selling cookies. Participants were told that Bobo was selling cookies in boxes containing 10 cookies inside as well as individually. Then participants were given three sets of untimed test trials. Prior to each set of trials, experimenters asked participants how many cookies were inside a box and told them the answer if they answered incorrectly.

The three sets of trials are described below:

- a. Count (5 trials) in which participants were shown a set of cookies, represented both as boxes and individual cookies, and were asked to count the total number of cookies on the screen. In order to succeed, participants had to count each box as 10 cookies, despite the 10 cookies not being visible in the boxes and combine the number of cookies in the boxes with the number of individual cookies. In each display, there were between one and seven boxes, and between one and six individual cookies. This task was free-response and children gave their responses verbally.
- b. Compare (8 trials), where participants were told that there were now two Bobos; a blue Bobo, and a yellow Bobo, each of whom had some cookies (in boxes and individually). In each trial, children had to decide who had had more cookies. The two sets of cookies were designed such that children could not determine the correct answer by simply counting the elements (i.e. boxes and individual cookies)

displayed on the screen. Responses were forced-choice (either blue or yellow Bobo had more), and responses were given orally or by pointing.

- c. Create (1 practice trial, 5 test trials) in which participants were asked to determine how many boxes (tens) and individual cookies (ones) made up a double-digit number. In a practice trial, the experimenter said “If I want 21 cookies, no more or no less, how many boxes and individual cookies should Bobo give me?”. Children were given feedback only in the practice trial. In each of the five test trials, a cartoon character appeared on the screen next to Bobo’s shop. Participants were told how many cookies the character needed, then were asked to say how many boxes and individual cookies Bobo should give them. The amounts of cookies being asked for was always a two-digit number. These trials were free-response and responses were given verbally.
5. Fill in the board (1 practice trial, 10 test trials), which required participants to find a number on a board. The board contained numbers from 1 through 99, in rows of 10 (with the decades on the left, and the last number before a new decade (e.g. 19) on the right). On every trial, the numbers 1, 10, 99, and two additional numbers (48 and 74 in assessment version A; 35 and 82 in assessment version B) were already filled in, while the rest of the spaces on the board were empty. In a practice trial, participants were given feedback and shown how to use the filled-in numbers on the board to find the correct answer. In each of the 10 test trials, participants were asked to locate a given number (from 1-99) on the board. Their previous responses were not available on the next trial. Participants gave their responses either by pointing with their mouse, using Zoom’s remote-control feature, or by

pointing with their finger (in which case caregivers communicated the answer given by the child to the experimenter). This test was untimed.

6. Paper Addition, where children were given 2 minutes to solve and write the answers to as many single-and-double-digit addition problems as they could, on a paper with 24 problems. Each problem was free-response and participants gave their answers by writing them directly on the sheet of paper (which had been mailed to them in the box with the game materials). Caregivers were asked to take a picture of the sheet right after the child completed it and email it to the experimenters (either during the Zoom meeting or right after).

The reading assessments measured children's oral reading fluency, pronounceable words fluency, and comprehension. There were 5 reading assessments are described below:

1. Nonsense Word Fluency (DIBELS, 2021): Participants were asked to read nonsense words out loud from a list that contained 75 of them displayed on the screen. Participants were given two minutes to read from the list and were told to read as many words as they could within the time limit. As they read, the experimenter used their mouse to point to each word to help the child keep track of which word to read next. No feedback was provided during this assessment. There were two practice words that children received feedback on before starting the assessment.
2. Word Reading Fluency (DIBELS, 2021): Participants were asked to read real words out loud from a list that contained 105 of them displayed on the screen. Participants were given two minutes to read from the list and were told to read as many words as they could within the time limit. As they read, the experimenter used their mouse to point to each word to

help the child keep track of which word to read next. No feedback was provided during this assessment and there were no practice words.

3. Oral Reading Fluency (DIBELS, 2021): Participants were asked to read a short story out loud. They were given two minutes to read as much of the story as they could. As they read, the experimenter used their mouse to point to each word to help the child keep track of which word to read next. Only for this assessment, children were told that the experimenter would help them with any word they couldn't read or got stuck on so that they could continue reading as much of the story as possible. There was no practice for this assessment.
4. Comprehension: Participants were asked four free-response questions about the content of the short story that they had read for the assessment above. They responded verbally to the questions. All four questions were asked irrespective of whether or not the participant read the whole story. However, they were only scored on questions that were relevant to the parts of the story that they read.
5. Pronounceable Words Fluency: Participants were given a list of 75 nonsense words on a sheet of paper. About half of them were pronounceable (e.g. "wib") and the rest were not pronounceable (e.g. "kxi"). Participants were given two minutes to circle all nonsense words that they thought were pronounceable. Caregivers were asked to take a picture of the sheet right after the child completed it and email it to the experimenters (either during the Zoom meeting or right after). There were four practice words for this assessment and children were given feedback only during the practice.

## **Procedure**

Prior to the first online session, the caregiver of each participant was mailed a box of materials containing their assigned game, a brochure containing information and instructions about the game, and a sticker tracker sheet with stickers to track the progress of the game at home. The first online session started with an experimenter administering one battery of assessments with the child (either VA or VB). Once the assessment battery was complete, a second experimenter, naive to the child's performance on the assessments, took over the session. This experimenter taught the child and their caregiver how to play their assigned game using a practice deck. Once the rules of the game were understood, the child and their caregiver were asked to continue playing while the experimenter was still present on Zoom.

The first session concluded with the experimenter asking the child and their caregiver to play the game 6-8 times at home using the game materials that they had been sent. They were informed that each deck (1-8) of the game should ideally be played during separate sessions rather than all at once. They were also asked to reserve the second practice deck for the next Zoom session. Caregivers were encouraged to fill out a short online survey each time the game was played at home to say which deck had been played, how long it took, who played with the child and whether it was played cooperatively or competitively between the players. It also gave caregivers the opportunity to ask the experimenter any questions that they might have about the games. Children were encouraged to track their progress in the game by adding a sticker to their sticker tracker sheet each time they completed a session of the game.

After the home-based gameplay, children and their caregivers came back for the second online session. If in the online surveys, it was found that children had played the game less than 6 times, families were given the option to reschedule their second session. During the second session,

an experimenter (who had taught families how to play their assigned game) asked the child and their caregiver to play the second practice deck. Then, the second experimenter, naive to the child's condition, administered the alternate version of the assessment battery than what the participant had received in session one. All online sessions were recorded via Zoom.

### **Data and analysis**

For all the assessments, we calculated the percent correct by dividing the number of trials a child answered correctly by the number of trials presented to them (i.e., correct + incorrect + responses of “I do not know”). Practice trials and the trials that weren't presented to a child were not entered into this calculation.

Each sub-test's score was transformed into Z-scores based on the mean and standard deviations of scores from both conditions at pretest and posttest. We then generated the two primary dependent variables: (i) the number composite score, calculated by averaging the Z-scores of the 6 numerical tests and, (ii) the reading composite score, calculated by averaging the Z-scores of the 5 reading tests. If a child did not finish at least 50% of a test, we treated it as missing data for that test. We used the lme4 package (Bates et al., 2015) in R (R core team, 2013) to implement linear mixed-effects models. All measures and analyses, unless specified otherwise, were preregistered prior to examining the data.

### **5.3 Results**

In order to determine if playing the number game, *Find and Move*, improved children's numerical concepts and skills, we analyzed the number composite score as a function of time (pretest/posttest), condition (number/reading) and their interaction. We also included random intercepts for participants. We found a significant time by condition interaction ( $b = -0.27$ ,  $\beta = -0.36$ ,  $t = -2.85$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-0.61, -0.11]$ ) indicating that children who played *Find and*

*Move* showed a significantly greater increase in their number composite score from pretest to posttest than did the children who played *Bingo* (the reading game).

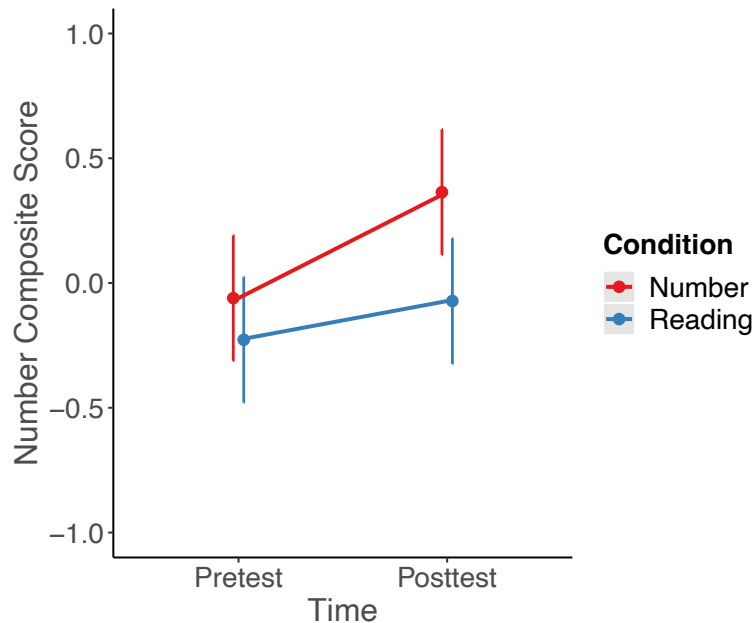


Figure 5.4. Estimates of the number composite score plotted by time and condition. Error bars represent standard errors of the mean.

In order to determine the effectiveness of playing the reading game, *Bingo*, towards improving children's reading skills, we analyzed the reading composite score as a function of time (pretest/posttest), condition (number/reading) and their interaction. We also included random intercepts for participants. We did not find a significant time by condition interaction ( $b = -0.03$ ,  $\beta = -0.04$ ,  $t = -0.33$ ,  $p = 0.74$ , 95% CI of  $\beta$   $[-0.27, 0.19]$ ) indicating that children who played *Bingo* showed no greater increase in their reading composite score from pretest to posttest than did the children who played *Find and Move* (the number game).

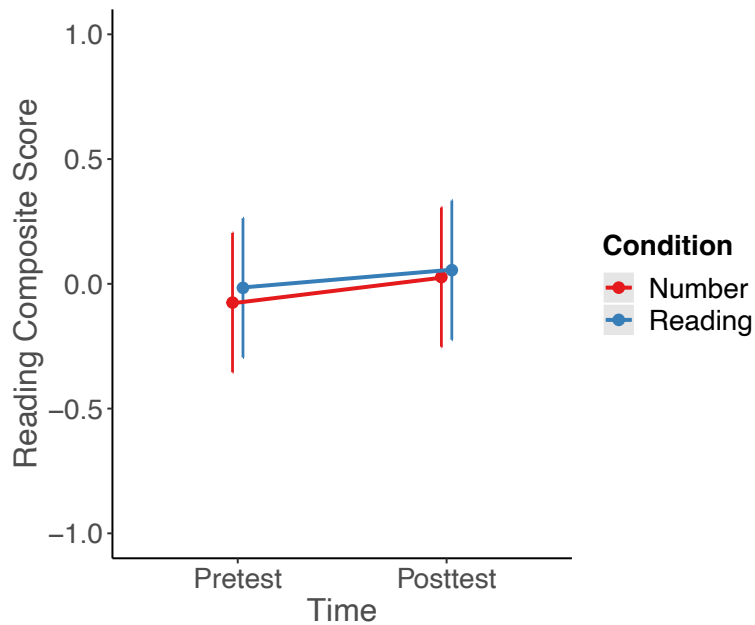


Figure 5.5. Estimates of the reading composite score plotted by time and condition. Error bars represent standard errors of the mean.

In order to determine the differential effects of games training on each numerical and reading test individually, we analyzed the Z-scores of each test using the same linear mixed effects model described above. Across all the numerical tests, we found a significant time by condition interaction only for one test named *Fill in the Board*, that required children to find numbers on a 10x10 board ( $b = -0.77$ ,  $\beta = -0.78$ ,  $t = -2.75$ ,  $p < 0.01$ , 95% CI of  $\beta$   $[-1.34, -0.22]$ ): children who played *Find and Move* showed a significantly greater increase from pretest to posttest as compared to children who played *Bingo*.

The time by condition interaction was not significant for any of the other approximate or exact numerical tests, though the results trended in the direction favoring the children who played *Find and Move* for the exact numerical tests (see Figure 5.6). Consistent with that trend, when the same analysis was applied to the composite score of all the numerical tests excluding the *Fill in*

*the Board* task and the *Approximate Symbolic Addition* task, it revealed a significant time by condition interaction ( $b = -0.26$ ,  $\beta = -0.33$ ,  $t = -2.14$ ,  $p < 0.05$ , 95% CI of  $\beta$   $[-0.63, -0.02]$ ). Therefore, children who played *Find and Move* showed a greater increase in a composite score measuring their exact numerical skills (even after excluding the *Fill in the Board* task) from pretest to posttest than did the children who played *Bingo*. None of the reading tests revealed a significant interaction when analyzed individually.

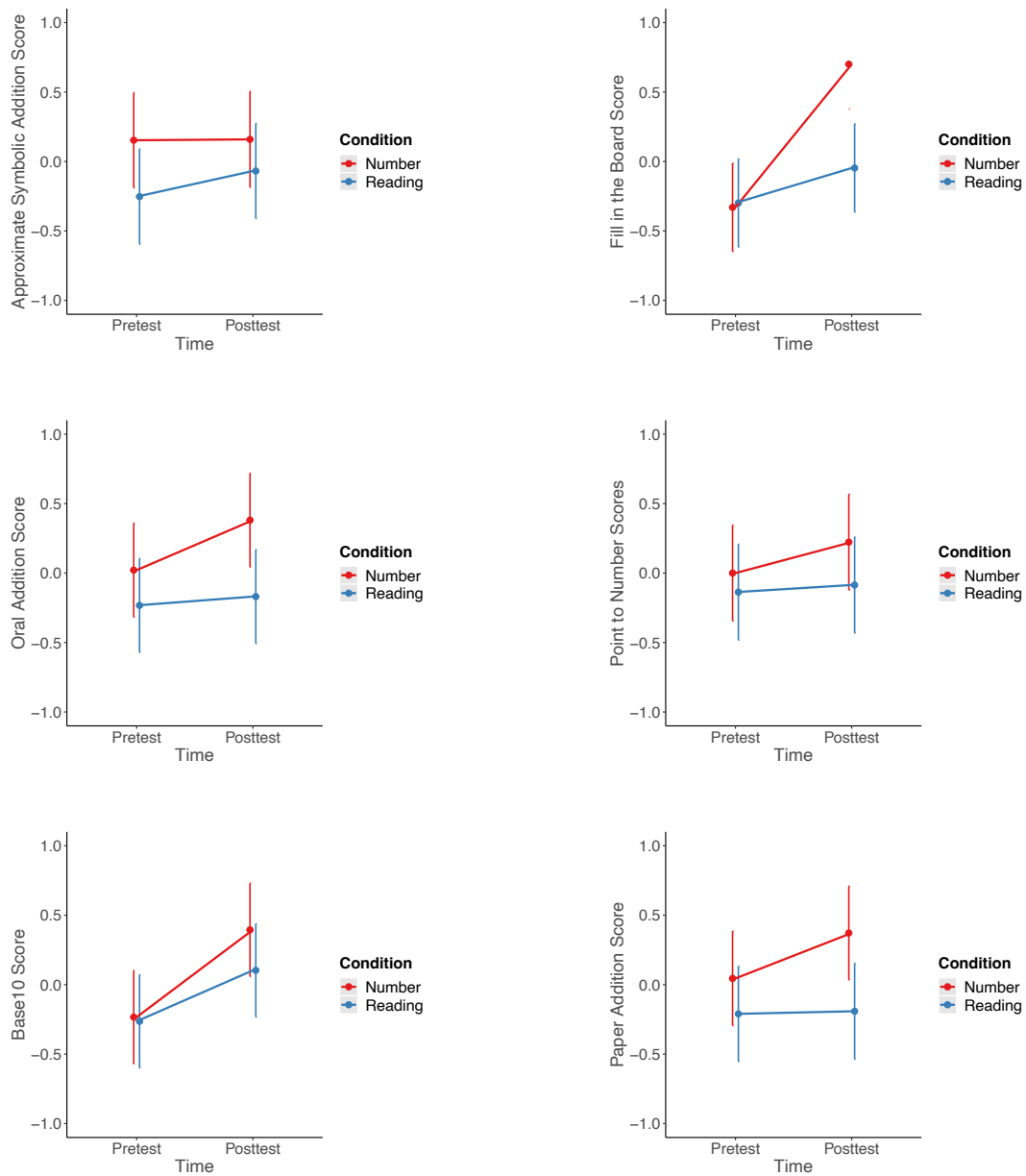


Figure 5.6. Estimates of the scores of different numerical assessments plotted individually by time and condition. Error bars represent standard errors of the mean.

## 5.4 Discussion

We evaluated an updated version of the numerical board game, *Find and Move*, by comparing it to a novel reading board game, *Bingo*, in improving 6-7-year-old children's numerical

concepts and skills. We found that children who played the numerical board game at home for 2-3 weeks with their caregiver(s) or sibling improved in their numerical concepts and skills relative to those who played the reading board game. This suggests that training children's developing understanding of base-10 compositionality enhances their school-relevant numerical concepts and skills in the short-term.

Both the numerical and reading board games were designed to enhance children's compositional reasoning in the domains of number and reading respectively. They also had a similar structure that required children to find numbers or words by utilizing the systematicity in the rows and columns of the boards. Since the improvements in numerical skills were specific to the children who played the numerical board game, they can be attributed to the numerical content of the game and not to the general features of the games or to the motivational benefits of playing educational games at home with a family member.

We found evidence that the effects on children's exact numerical skills were not just driven by increased familiarity with the materials of the game. Children who played the numerical board game improved in a composite score that excluded the near transfer measure of finding numbers on the board. Therefore, the game improved numerical skills that go beyond the familiarity with the numerical board that children played with at home.

We also found evidence that the new structure of the board (where decades are aligned to the left) did not pose any challenges to children's learning. Children improved significantly in the near transfer measure of finding numbers on the 10x10 board from pretest to posttest. Therefore, though children might be seeing this differently structured board for the first time in the intervention, they were still able to learn its structure in representing the numbers from 1-99. They got more accurate in finding numbers on it after the home-based game play. In order to test the

efficacy of the two different structures of the board (left or right aligned decades) in improving children's base-10 understanding, future work should directly compare the effects of training with both types of boards in the same intervention.

While we found significant improvements in a composite measuring children's exact numerical skills, we found no improvements (or even trends in the right direction) in children's approximate symbolic addition skills. Since the approximate symbolic addition test draws upon representations of the Approximate Number System to perform addition in a ratio-based manner, we find preliminary evidence that training children's exact numerical representations of the base-10 number system does not lead to improvements in their approximate numerical skills.

We did not find any positive effects of the reading game, *Bingo*, on children's reading skills. This might be due two reasons: (i) the assessments failed to capture what children learned in the short-term intervention or (ii) the game was ineffective in training children's reading skills. The reading assessments (DIBELS, 2021) that we used in our study are part of a standardized tool used in classrooms to monitor early reading skills and detect students who might be at risk for performing below grade level. They might not be well suited to capture the skills children can learn in a 2-3 week reading intervention. Such assessments may be better suited to capture improvements in reading that happen with instruction over several months in schools.

In order to determine whether children got better at playing the reading game, as a proxy to determine if they learned any skills from the game, our next step is to code video recordings of children's gameplay during the Zoom sessions before and after the home-based gameplay. We will code for two key measures: (i) time taken to finish playing with a reading board and (ii) accuracy in finding words independently on the board.

If children take less time and are more accurate in finding words on the board in the second Zoom session as compared to the first, then it would serve as evidence that they got better at playing the game. Since the words used in the practice decks in sessions 1 and 2 are similar in difficulty but different in terms of the exact words used, improvements in gameplay might be indicative of mastery with some skills that the game aims to train. If we find improvements in gameplay from session 1 to session 2, then the next step would be re-design the assessments to include near transfer measures.

### **Limitations and future directions**

Though we conceptually replicated the efficacy of *Find and Move* in improving children's numerical concepts and skills in the short-term, this experiment has several limitations. First, it focused only on short-term training and on a limited set of outcome measures. Future evaluations of the game with a larger sample size, longer duration of the intervention, and a measurement of school-based math learning in the long-term can help determine the efficacy of the game in robustly enhancing children's math learning.

Second, in order to determine whether there is an added advantage in using a game with numbers up to 99 in improving children's numerical understanding, future work should directly compare the effects of a number game going up to 50 to one that goes up to 99. If children are better able to learn the base-10 concepts underlying their numerical skills with a game using numbers up to 99, then it might be worth introducing larger numbers to children early on.

The efficacy of *Find and Move* can also be tested by comparing it with other numerical interventions that aim to train children's understanding of base-10 compositionality. This can serve two purposes: (i) by comparing it to other interventions in the literature (e.g., base-10 blocks) can help us determine if our game is more effective in improving children's numerical concepts and

(ii) by comparing it to different versions of the game that each include different representations of the base-10 system (i.e., numerical board, grouped dots by tens, or transparent number words) can help us which representations or features of the game are the most effective. Such studies can help us determine the efficacy of all three representations used in Find and Move in improving children's numerical understanding.

Finally, the reading game, *Bingo*, covered just one aspect of reading which involves mapping graphemes to phonemes. However, mapping text to meaning is a part of reading that future interventions should focus on. Moreover, compositional reasoning may be especially hard in the reading domain as, in English, there are 26 symbols (as opposed to 10 symbols in the number domain) that together can generate many words, but many words violate regular spelling. While our number game exposed children to all 99 numbers that featured in the assessments, the reading game exposed children only to a subset of speech sounds that featured in the assessments. Therefore, future reading interventions might also benefit from longer training durations.

In conclusion, we found positive effects of our numerical game on children's numerical concepts and skills. We show that introducing educational games in children's home has the potential to enhance the learning of math in school.

## **Chapter 6**

### **Conclusion**

The present dissertation investigated the challenges that children face in learning the compositional number system and evaluated game-based interventions, administered by children's caregivers in their own homes, to help them learn it and the numerical concepts it supports.

#### **6.1 Highlights of Chapters 2-5**

In Chapter 2, I found evidence that home-based games training children's innate numerical concepts do not support their learning of the symbolic numerical concepts typically taught in school. Though innate systems of numerical concepts continue to function throughout adulthood (Barth, Kanwisher, & Spelke, 2003; Barth et al., 2006; Hyde & Spelke, 2009; Piazza et al., 2004), become linked with symbolic numerical concepts (Gilmore, McCarthy, & Spelke, 2007) and are represented using similar brain regions as symbolic numerical concepts (Amalric & Dehaene, 2016; Cantlon et al., 2009; Piazza et al., 2007) they might not play a causal role in the development of symbolic numerical concepts.

Similar versions of the two numerical games from the paper in Chapter 2 were also evaluated as part of a large-scale school-based intervention that used a suite of four games training both numerical and geometry skills (Dillon et al., 2017). Chapter 2 provides evidence for the contexts in which training children's innate numerical representations helps. It suggests that it might work only with younger children and with longer training durations. Chapter 2 also highlights the features that made Dillon et al., (2017) successful and sheds lights on what future interventions should focus on. Children might need more scaffolding with the language and symbols that represent the compositional number system to learn math efficiently in school.

In Chapter 3, I found evidence that the degree of transparency of number words in conveying the logic of base-10 compositionality (i.e., “eleven” in English as opposed to its translation to “ten one” in Mandarin) affects children’s productive counting skills, such that children who speak more transparent languages outperform their age-matched peers who speak relatively less transparent languages. However, learning the logic of base-10 compositionality is hard and takes years even when children’s languages have number words that make it relatively more explicit. This highlights the need to develop and evaluate interventions both to deepen our understanding of how children learn compositional concepts and also to support their learning.

In Chapters 4 and 5, I found evidence that home-based games that use multiple representations (e.g., grouped dots by tens, numerical boards, and transparent numerical language), each that make some aspect of base-10 compositionality more explicit to children, help them learn it and the numerical concepts that it supports. The game from Chapter 4 was also evaluated as part of a large-scale school-based intervention that used a suite of four games among children in India (Dean et al., in prep). Chapter 4 provides evidence that the game is effective in a different context (i.e., among US children in their homes) and it also highlights that it was a crucial component in driving children’s learning in the multifaceted school-based intervention.

In Chapter 5, we built upon features of the game from Chapter 4 to make it more effective in helping children learn base-10 compositionality. We re-designed it to be a multi-player game, increased children’s exposure to larger numbers, challenged children to represent the same number in multiple ways, and introduced a novel spatial organization of number symbols on a board. We found that these features were effective in enhancing children’s compositional reasoning. This new version of the game is currently being tested in a large-scale school-based intervention to evaluate its efficacy in supporting children’s school learning in the long-term.

Taken together, this dissertation highlights two important features of home-based interventions that can advance our understanding of how children learn: (i) they help to establish causal impacts of the representations and specific experiences that combine to help children learn numerical concepts in their real-world environments and (ii) they help to pinpoint the effective features of large-scale, multifaceted interventions that future research can build upon to improve pedagogy. They help to determine whether every component of a large multifaceted intervention is necessary and enhances different but complementary aspects of children's learning.

Such research on children's learning of compositional concepts is at an early stage and this dissertation, therefore, leaves open several questions that I will briefly explore below.

## **6.2 What representations help children learn base-10 compositionality?**

Number words serve to represent the base-10 compositional number system. Since languages' number naming systems vary in the degree of transparency by which they make base-10 compositionality explicit, future research must test whether all children (who speak a variety of languages) learn base-10 compositionality using similar mechanisms. For example, do children who speak highly non-transparent languages like Hindi learn base-10 compositionality by relying more on number symbols as opposed to number words? Do children who speak transparent languages like Mandarin or Tamil learn base-10 compositionality primarily via their number words and independent of formal instruction with number symbols and arithmetic? Does formal instruction focusing on the compositional number system and arithmetic differentially impact the learning of speakers of different languages?

Relatedly, one line of research has found that children who sell produce in markets use the base-10 compositional number system efficiently to aid fast and flexible mental calculation (e.g., Nunes, Schliemann, & Carraher, 1993; Banerjee et al., in prep). They appear to do so not because

of formal instruction, but by observing other adults who sell in markets. How do these children acquire the compositional rules that underlie the number system and utilize it to perform arithmetic efficiently in market contexts? An analysis of their learning mechanisms can help inform school-based pedagogy that aims to impart the efficient use of the compositional number system to calculate.

In addition to number words and symbols, do children benefit from exposure to other representations such as physical models (e.g., abacus, base-10 blocks, 10x10 numerical boards, etc.) that serve to represent base-10 compositionality? When are such representations useful and what specific numerical concepts do they enhance? For example, Yuan et al., (2021) find that children do not benefit from instruction with physical models (e.g., abacus and base-10 blocks) when learning about how multidigit numbers are named and how they represent relative magnitudes. They found that children learn these skills better with direct exposure to the number words and symbols themselves.

However, we do not yet know the relative efficacy of physical models in helping children learn the compositional rules that underlie the number system. The present dissertation found that using multiple representations (e.g., grouped dots by tens, numerical boards, and transparent numerical language) together help children learn base-10 compositionality and the numerical concepts it supports. Future work must test the differential impact of each representation that aims to make some aspect of base-10 compositionality more explicit to children.

### **6.3 Do children rely on compositional reasoning to acquire the meanings of single-digit numbers?**

The present dissertation mainly focused on children's learning of large multidigit numbers and the compositional rules that underlie them. Does compositional reasoning also aid children's

learning of the meanings of single-digit numbers like “five” and “six”? How children acquire the earliest natural number concepts is still an area of active investigation.

One proposal suggests that compositional language, which highlights how bigger numbers are composed of smaller ones, may allow children to learn the meanings of the first few natural numbers beyond three or four (Spelke, 2017). In a separate line of work (not included in this dissertation) we are testing this proposal by evaluating the benefits of a short training paradigm, focused on compositional cues (i.e., linguistic and visual cues highlighting how a set of “five” can be made up of a set of “three” and a set of “two”; or how a set of “six” can be made up of a set of “three” and another set of “three”), on children’s understanding of “five” and “six” by comparing it to a control training condition without any compositional cues.

Such research can clarify whether children can use compositional cues at earlier ages and advance in their understanding of natural number concepts.

#### **6.4 Children’s compositional reasoning beyond the number system**

The present dissertation investigated children’s compositional reasoning in the domain of number. However, children are faced with the challenge of mastering other domains that also rely on compositional reasoning. For example, to learn to read, children must grasp how speech sounds combine to form words, how words combine to form phrases and sentences, and how these combinations are used to convey meaning. Does compositional reasoning in one domain predict it in another domain? Future work must test this question in order to examine whether there are any general features of compositional reasoning that are shared across domains.

Apart from reading and numerical concepts that most children come to learn formally in school, children also learn language, which is also a compositional system. Children learn their language within the first few years of their life without any formal instruction. Using language to

utter and understand infinite thoughts depends on our ability to combine finite vocabularies using a limited set of rules that we are not even consciously aware of. How do children master their language? Are the features that enable language learning shared with other domains such as reading and number? Future work must examine features of compositional reasoning (e.g., coordinating multiple representations of the same concept (*one ten* and *ten ones*) and representing part-whole relationships (*one ten* as composed of *ten ones*) that might be shared across multiple domains.

## **6.5 Final Conclusion**

The present dissertation advances our understanding of how children learn the compositional number system. More specifically, it suggests that (i) children's innate numerical concepts do not support their learning of the symbolic concepts that are supported by the compositional number system, (ii) base-10 compositionality is hard for children to learn even in languages that makes it more explicit, and (iii) multiple representations, that each make some aspect of base-10 compositionality more explicit, help children learn it. Several key questions remain concerning the efficacy of different representations in supporting the learning of base-10 compositionality, the role of compositionality in acquiring the earliest natural number concepts, and the relationship between compositional reasoning across multiple domains. Studying these questions can advance our understanding of how humans use compositional reasoning, a key feature of human learning that makes it efficient.

## Chapter 7

# Appendix

## 7.1 Supplementary Information (SI) for Chapter 3

### 7.1.1 Tasks

Table S7.1.1 The full trial structure for the base-10 task

|                         | Trial Type | Target # of cookies |                  |
|-------------------------|------------|---------------------|------------------|
|                         |            | Comparison Set 1    | Comparison Set 2 |
| Trial 1                 | Compare    | 6                   | 11               |
| Trial 2                 | Compare    | 23                  | 16               |
| Trial 3                 | Compare    | 30                  | 25               |
| Trial 4                 | Compare    | 28                  | 42               |
| Trial 5                 | Compare    | 13                  | 31               |
| Trial 6                 | Compare    | 74                  | 91               |
| Trial 7                 | Compare    | 85                  | 67               |
| <b>Practice Trial 1</b> | Count      | 21                  |                  |
| Trial 1                 | Count      | 13                  |                  |

|                         |        |    |
|-------------------------|--------|----|
| Trial 2                 | Count  | 30 |
| Trial 3                 | Count  | 25 |
| Trial 4                 | Count  | 44 |
| Trial 5                 | Count  | 60 |
| <b>Practice trial 1</b> | Create | 32 |
| Trial 1                 | Create | 23 |
| Trial 2                 | Create | 39 |
| Trial 3                 | Create | 16 |
| Trial 4                 | Create | 51 |
| Trial 5                 | Create | 77 |

### Give-N-Task

In this task, children were asked to “give” a certain number of objects (of the same kind) by placing them on a plate. They were asked to give 1-6 objects. This task was used measure whether or not a child knows how counting represents number (Wynn, 1992). We implemented a titrated version of this task, following the procedure used by Marchand & Barner (2020). For the Give-N task, caregivers were asked to gather small identical objects (e.g., pieces of pasta, grapes, pieces of Lego) and a flat plate from their own homes in study 1. In study 2, the experimenters provided the objects and flat plate as testing was done in-person.

## **Unit Task**

In this task, we measured children's understanding of the successor function, that is, the concept that when one object is added to a set, the number word that now applies to this set is the next word in the count list. This task was adapted from Sarnecka & Carey (2008) and involved children watching an animation that showed some objects (e.g., bananas) being dropped into a box. The children were then asked to say how many bananas there are in the box to confirm that they were paying attention and knew the correct initial number of objects in the container. Subsequently, 1 or 2 more objects of the same kind were added to the box and then the children were asked a two-choice question such as "How many bananas are there in the box now? N+1 or N+2?" or "How many bananas are there in the box now? N+1 or N+11?" or "How many bananas are there in the box now? N+1 or N+10?". There was a total of 12 test items in this task.

### **7.1.2 Math and language survey**

In addition to these tasks, parents of children from US, China, and Turkey completed a survey that asked them questions about the kinds of number-related activities their child engaged in at home and in school (if they are enrolled in one). It also asks them if the child is exposed to any other language apart from English.

### **7.1.3 Results from Study 1**

#### **Children's productive knowledge of counting**

In order to determine if there is a difference in the number of number words that children produce after being provided with the next number in their count range, based on their language, we conducted a one-way Anova. We found a statistically significant difference between the language groups ( $F(2, 212) = 3.17, p < 0.05$ ). English-speaking children, on an average produce about 5 number words after being provided with the next number word, Mandarin-speaking

children produce about 9 number words, and Turkish-speaking children produce about 9 number words. See table S7.1.2 for the average number of number words produced by age in the three languages.

Table S7.1.2. Average number of number words produced by age in years

|          | 4 years | 5 years | 6 years |
|----------|---------|---------|---------|
| English  | 4.61    | 5.68    | 3.18    |
| Mandarin | 8.63    | 8.96    | 6.9     |
| Turkish  | 6.56    | 8.06    | 14.93   |

### Effects of age, language, and count range on logic task performance

Before conducting any statistical modeling, we first plotted average logic task scores (i.e., the percent correct) by age in months (see figure S7.1.1) and then by count range (see figure S7.1.2) for each language group.

Trends from figure S7.1.1 indicate that English-speaking and Mandarin-speaking children have higher average logic task scores as compared to Turkish-speaking children, and Turkish-speaking children improve at a slower rate with age.

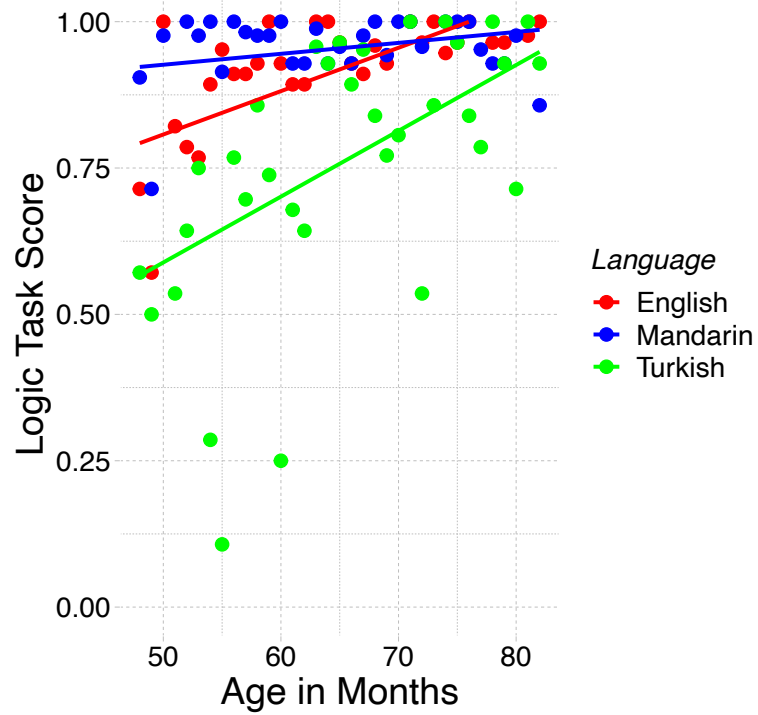


Figure S7.1.1. Average Logic Task Scores by Age by Language

Trends from figure S7.1.2 indicate that when matched for count range, English-speaking and Mandarin-speaking children have similar average logic task scores, but Turkish-speaking children have lower scores on average.

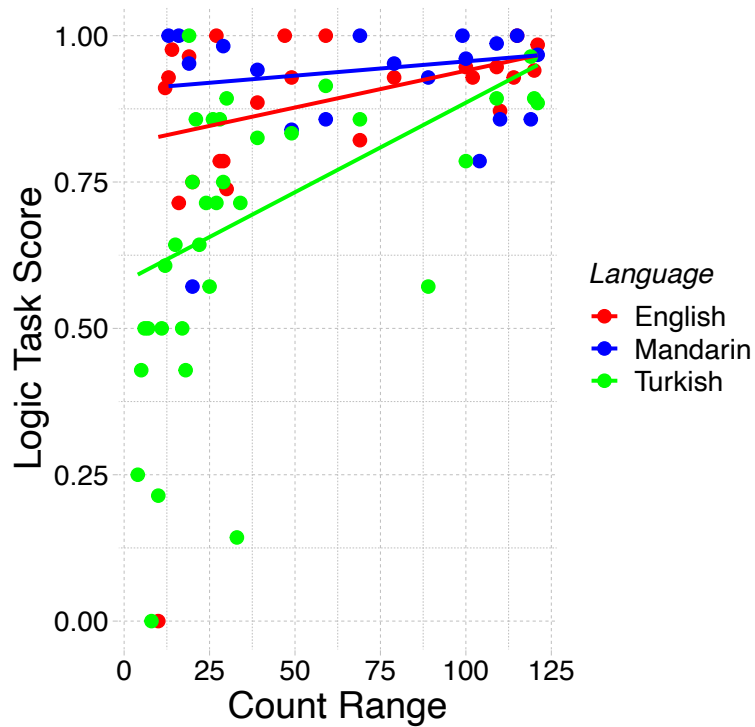


Figure S7.1.2. Average Logic Task Scores by Count Range by Language

In order to determine the effects of age, count range, and language on children's logic task performance, we built logistic mixed-effects models in R using the `glmer` function from the `lme4` package (Bates et al., 2015). We included subject and trial as random intercepts in all models. We first analyzed our dependent variable, logic task performance (coded as 0 or 1), as a function of age, and then sequentially added the predictors of language, and the interaction between age and language, in that order. As a final step, we added count range as a predictor to the same model. We then performed model comparisons using AIC and the ANOVA function to determine the model with best fit. See Table S7.1.3 for model specifications along with the results of the model comparisons.

From table S7.1.3, we found that model Model 4 was the best fitting model as it explained the data better than Model 3 ( $p < 0.01$ ). We found that there were significant effects of age ( $b = 0.63$ ,  $SE = 0.22$ ,  $z = 2.86$ ,  $p < 0.01$ ) and language for Turkish ( $b = -1.44$ ,  $SE = 0.28$ ,  $z = -5.22$ ,  $p < 0.01$ ) but not Mandarin ( $b = 0.37$ ,  $SE = 0.31$ ,  $z = 1.2$ ,  $p = 0.23$ ). We also found significant effects of count range ( $b = 0.59$ ,  $SE = 0.13$ ,  $z = 4.41$ ,  $p < 0.01$ ). This indicates that children's logic task performance improved with age and count range, and that Turkish-speaking children have lower scores than English-speaking children but there is no difference between Mandarin-speaking children and English-speaking children. In model Model 4, we also found a significant interaction between age and language for Mandarin ( $b = -0.59$ ,  $SE = 0.29$ ,  $z = -2.03$ ,  $p < 0.05$ ) but not for Turkish ( $b = -0.09$ ,  $SE = 0.27$ ,  $z = -0.34$ ,  $p = 0.73$ ). This indicates that the differences in the rates of improvement in logic task performance by age differed significantly between English-and-Mandarin-speaking children but not between English-and-Turkish speaking children, when controlling for count range. See figure S7.1.3 for the predicted probabilities of logic task scores by language.

Table S7.1.3. Model comparisons for the mixed-effects models predicting logic task performance

| Model Specifications                                                    | AIC     | $\chi^2$ | df | p      |
|-------------------------------------------------------------------------|---------|----------|----|--------|
| M1: Performance ~ Age + (1 Subject) + (1 Trial)                         | 2421.23 | -        | -  | -      |
| M2: Performance ~ Age + Language + (1 Subject) + (1 Trial)              | 2353.14 | 72.1     | 2  | <0.01* |
| M3: Performance ~ Age * Language + (1 Subject) + (1 Trial)              | 2352.79 | 4.34     | 2  | 0.12   |
| M4: Performance ~ Age * Language + CountRange + (1 Subject) + (1 Trial) | 2336.97 | 18.82    | 1  | <0.01* |

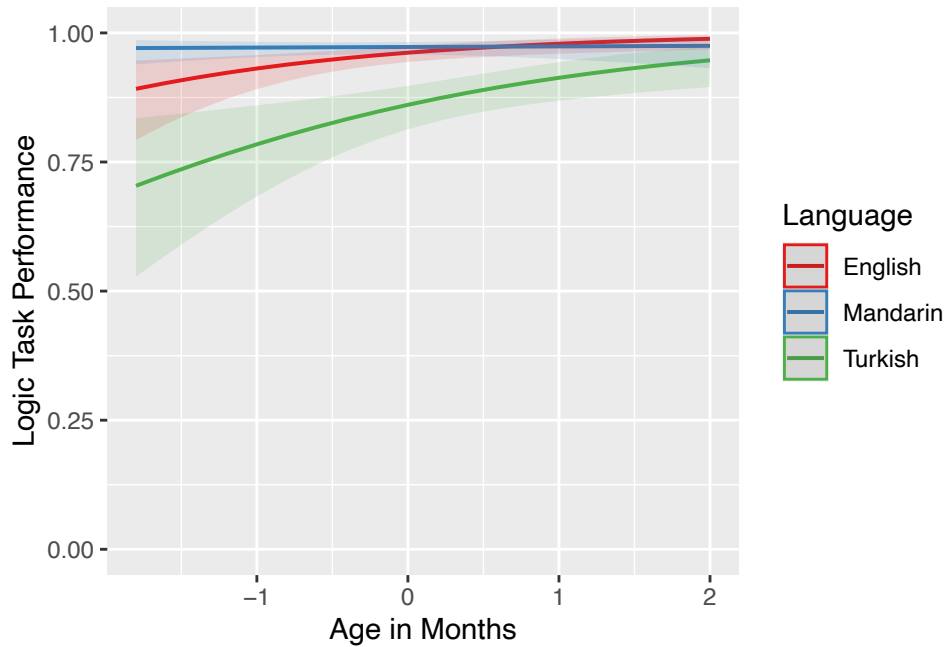


Figure S7.1.3. Predicted probabilities of Logic Task Scores from Model 4 by Age by Language.

Age in Months on the x-axis reflect scaled values

#### 7.1.4 Results from Study 2

##### Children's productive knowledge of counting

In order to determine if there is a difference in the number of number words that children produce after being provided with the next number in their count range, based on their language, we conducted an independent samples t-test. We found that Tamil-speaking children, on an average produce about 8 number words after being provided with the next number word and Hindi-speaking children produce about 2 number words. This difference is statistically significant ( $t(119) = -4.88, p < 0.01$ ). See table S7.1.4 for the average number of number words produced by age in Hindi and Tamil.

Table S7.1.4. Average number of number words produced by age in years

|       | 5 years | 6 years | 7 years |
|-------|---------|---------|---------|
| Hindi | 2.00    | 1.47    | 0.97    |
| Tamil | 2.74    | 11.18   | 9.90    |

### Effects of age, language, and count range on logic task performance

Before conducting any statistical modeling, we first plotted average logic task scores (i.e., the percent correct) by age in months (see figure S7.1.4) and then by count range (see figure S7.1.5) for each language group.

Trends from figure S7.1.4 indicate that Tamil-speaking children improved in their logic task scores at a faster rate with age as compared to Hindi-speaking children.

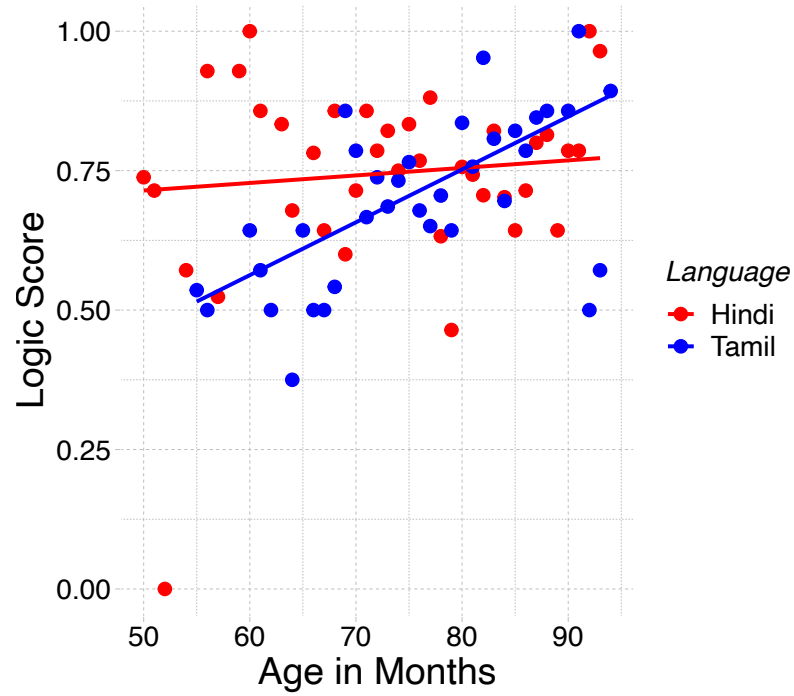


Figure S7.1.4. Average Logic Task Scores by Age by Language

Trends from figure S7.1.5 indicate that when matched for count range, Hindi-speaking children have higher average logic task scores as compared to Tamil-speaking children.

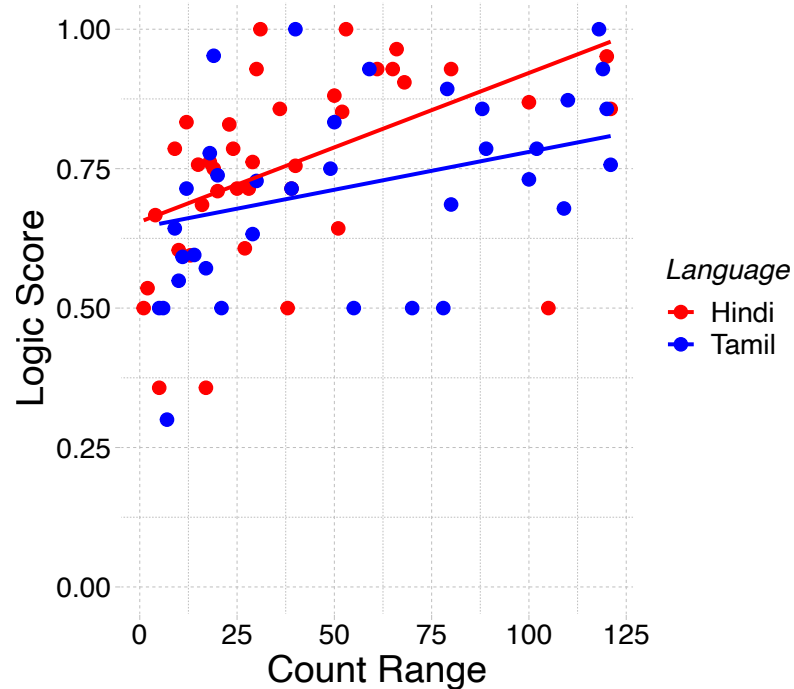


Figure S7.1.5. Average Logic Task Scores by Count Range by Language

In order to determine the effects of age, count range, and language on children's logic task performance, we built logistic mixed-effects models in R using the `glmer` function from the `lme4` package (Bates et al., 2015). We included subject and trial as the random intercepts in all models. We first analyzed our dependent variable, logic task performance (coded as 0 or 1), as a function of age, and then sequentially added the predictors of language, and the interaction between age and language, in that order. As a final step, we added count range as a predictor to the same model. We then performed model comparisons using AIC and the ANOVA function to determine the model with best fit. See Table S7.1.5 for model specifications along with the results of the model comparisons.

From table S7.1.5, we found that model Model 4 was the best fitting model as it explained the data better than Model 3 ( $p < 0.01$ ). We found that there was no significant effect of age ( $b = 0.09$ ,  $SE = 0.11$ ,  $z = 0.85$ ,  $p = 0.4$ ), but there were significant effects of language ( $b = -0.43$ ,  $SE = 0.18$ ,

$z = -2.38$ ,  $p < 0.05$ ), and count range ( $b = 0.42$ ,  $SE = 0.09$ ,  $z = 4.52$ ,  $p < 0.01$ ) on logic task performance. We did not find a significant interaction between age and language ( $b = 0.33$ ,  $SE = 0.17$ ,  $z = 1.92$ ,  $p = 0.06$ ) in this model. This indicates that the differences in the rates of improvement in logic task performance by age did not differ significantly by language when controlling for count range. See figure S7.1.6 for the predicted probabilities of logic task scores by language.

Table S7.1.5. Model comparisons for the mixed-effects models predicting logic task performance

| Model Specifications                                                    | AIC     | $\chi^2$ | df | p      |
|-------------------------------------------------------------------------|---------|----------|----|--------|
| M1: Performance ~ Age + (1 Subject) + (1 Trial)                         | 3058.47 | -        | -  | -      |
| M2: Performance ~ Age + Language + (1 Subject) + (1 Trial)              | 3059.55 | 0.92     | 1  | 0.34   |
| M3: Performance ~ Age * Language + (1 Subject) + (1 Trial)              | 3054.14 | 7.41     | 1  | <0.01* |
| M4: Performance ~ Age * Language + CountRange + (1 Subject) + (1 Trial) | 3036.35 | 19.8     | 1  | <0.01* |

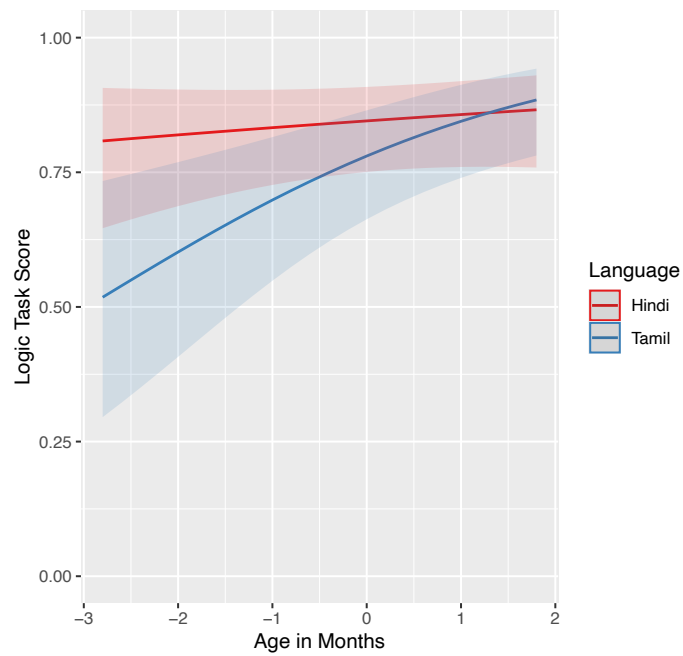


Figure S7.1.6. Predicted probabilities of Logic Task Performance from Model 4 by Age by Language. Age in Months on the x-axis reflect scaled values

## **7.2 Supplementary Information (SI) for Chapter 4**

### **7.2.1 Descriptions of the games**

#### **Find and Move**

The number board, which included numbers from 1 to 50 arranged in five rows, was printed on 8.3 x 11.7 inches cardstock white paper horizontally and laminated. Erasable colored markers were used to circle numbers on the board. All cards in 10 decks of the game were printed on 13.3 x 7.5 inches cardstock white paper and laminated. Decks 1-8 were played at home and families were encouraged to play them in order as they progressed in difficulty. The remaining 2 decks, called Practice 1 and Practice 2, were played in the lab.

Decks 1 and 2 introduced the “Find” operation in the absence of any movement on the board. Cards in these decks depicted a single number, presented as an array of dots grouped by tens. Children were tasked with finding and circling that number on the board. While Deck 1 had numbers only until 20, Deck 2 went up to 50.

Decks 3 to 8 introduced the “Move” operation in addition to “Find”, such that children first found a number, depicted using grouped dots, on the board and then added a second number to it by moving spaces on the board. The second number was either depicted using grouped dots or Arabic symbols. Deck 3 required children to find two numbers (either 2 or 12) depicted using dots, and move by a single-digit number depicted using dots. Deck 4 required children to find any single-digit number depicted using dots and move by a double-digit number depicted using dots. Deck 5 required children to find any number between 1-50 depicted using dots and move by a single-digit number represented as an Arabic symbol. Deck 6 required children to find any number between 1-50 depicted using dots and move by a number in the tens (10, 20, 30 or 40) represented as an Arabic symbol.

Deck 7 required children to find any number between 1-50 depicted using dots and move by double-digit numbers represented as an Arabic symbol, such that the final answer was 50 or below. Half the cards in this deck required children move by a number in the tens and half the cards required them to move by any double-digit number with both a tens and a units digit. Deck 8 required children to find any number between 1-50 depicted using dots and move by double-digit numbers represented as an Arabic symbol, such that the final answer was 50 or below. For decks 5-7, where the second addend was always represented using an Arabic symbol, the back of the card contained a depiction of it using dots. If children did not know how many tens there were in a number, they were encouraged to guess and/or look at the back of the card. Deck 8 contained no cues at the back of the cards.

The two practice decks that were played during the in-lab sessions contained example cards from all the 8 decks so that caregivers could learn how the game progresses. Practice 2 played in session 2 was similar in structure to Practice 1 but contained different cards. Caregivers were also provided with an answer sheet and were encouraged to check the answers either after every card or at the end of a deck.

### **Mind Hunter**

Like *Find and Move*, the cards in 10 decks of *Mind Hunter* were printed on 13.3 x 7.5 inches cardstock white paper and laminated. Decks 1-8 were played at home and families were encouraged to play them in order as they progressed in difficulty. The remaining 2 decks, called Practice 1 and Practice 2, were played in the lab.

Deck 1 required children to think about the preferences of characters depicted using symbols and determine which of two emotions or actions is more likely. Deck 2 consisted of cards half of which were similar to deck 1 but the other half required children to infer a character's

emotional state based upon social scenarios (for example, winning a race) and determine which of two emotions or actions is likely. Deck 3 required children to track knowledge states of different characters and determine what action they would take or what belief they would hold. Deck 4 required children to track the perceptual states of different characters and determine who can see what object(s).

Decks 5-8 required children to reason about the false beliefs of different characters. They were similar to the Sally-Anne (Baron-Cohen, Leslie and Frith, 1985) task where a character moves an object from one location to another in the absence of a second character and children need to reason about where the second character thinks the object will be. Deck 5 was easier as it only included two characters and two possible locations for the objects. Deck 6 also always included two characters but contained some cards in which there were three possible locations for the objects. Deck 7 contained some cards which included 3 characters along with three possible locations for the objects. Deck 8 consisted of cards half of which were similar to deck 7 but the other half had cards similar to the “Smarties” task (Gopnik & Astington, 1988) where children had to reason about a character’s alternative representations of the same object.

The two practice decks that were played during the in-lab sessions contained example cards from all the 8 decks so that caregivers could learn how the game progresses. Practice 2 played in session 2 was similar in structure to Practice 1 but contained different cards. The back of the cards in all decks contained the right answers which caregivers could use to determine if the cards were sorted into the right bins. In session 1, children and caregivers were introduced to a pretraining brochure before playing Practice 1 in order to familiarize them to the symbols used in the game.

### **7.2.2 Descriptions of the assessments**

#### **Numerical Assessments**

Below, we describe the numerical tests and indicate which ones were deemed to be diagnostic of base-10 compositional structure of the number system.

1. *Say the Number* (10 trials): A number symbol ranging between 2 to 210 was displayed on the screen and children needed to say the number word that corresponded to it. There was 1 single-digit trial, 7 double-digit trials, and 2 triple-digit trials. There were no practice trials for this task. The 9 trials that included double- or triple-digit numbers were more likely to be solved correctly if children used their understanding of base-10 compositionality.
2. *Point to Number* (8 trials): The experimenter verbally presented a number word and children pointed to the corresponding number symbol from six options displayed on the screen. The six test options presented single-digit numbers on two trials; double digit numbers on five trials, and a mixture of double- and triple-digit numbers on one trial. Some distractors contained the same digits as the right answer but in reverse order, others contained the same digit either in the tens (40) or units (55) place, and others were random. There were no practice trials for this task. The six trials presenting double- or triple-digit numbers were more likely to be solved correctly if children used their understanding of base-10 compositionality.
3. *Oral Math* (11 trials): The experimenter presented numerical problems verbally and children were required to produce verbal answers. The problems included numerical comparison (“can you tell me which is bigger, 16 or 12?”) and arithmetic (“I have 6 candies and get 2 more, how many candies do I have now?”). The arithmetic problems included both addition and subtraction. All the numbers used in this test were below 20. Children were allowed to use their fingers to produce the answers. There were no practice trials in this task. Two trials in this test involved two-digit numbers and were more likely to be solved correctly if children used their understanding of base-10 compositionality.

4. *Approximate Symbolic Addition* (12 trials): Two characters with bags of candy were displayed on the screen. One character had two bags and the other had one bag. The amount of candy in the bags was depicted using number symbols. Children needed to decide who had more candy by approximately adding the numbers on the pair of bags, comparing the sum to the number on the single bag, and indicating the character with more candy either verbally or by pointing (after Gilmore, McCarthy, & Spelke, 2007). All the numbers used in this task were below 100. There was one practice trial in this task. No trials in this task were deemed to be diagnostic of an understanding of base-ten compositionality as they required approximate accuracy based on the ratios between the numbers.
5. *Speed Symbolic Comparison* (12 trials): Two number symbols were displayed on the screen. Children pressed one of two keys to indicate the number symbol that they judged to be larger. The numbers used in this test were all double-digit with one exception where a double-digit number and a triple-digit number were presented. Pairs of numbers across the test were chosen in a manner that required an analysis of both digits within a number. For example, in the pair 32 vs 25, if a child only paid attention to the units place, they would get the answer wrong. They would however get it right if they only paid attention to the tens digit. Whereas, in the pair 13 vs 18, only paying attention to the tens digit would not help in arriving at the right answer. There was one practice trial in this task. All trials in this test were more likely to be solved correctly if children used their understanding of base-10 compositionality.
6. *Panamath* (12 trials): Two sets of dots were displayed on the screen, side by side. Without counting, children pressed one of two keys to indicate the set of dots that they judged to be more numerous (after Halberda et al., 2008). There were three practice trials in this task. Since

this test was purely non-symbolic and used no grouped dots, none of the trials were deemed to be diagnostic of an understanding of base-ten compositionality.

7. *Number Line Estimation* (35 trials): Number lines beginning from 0 with varying end points (10, 15, 25, 100) were displayed on the screen, each with a number symbol displayed above it. Children pointed to the location on the line where they judged that the number symbol belonged (after Laski & Siegler, 2014). 5 trials used a number line from 0-10. 10 trials used a number line from 0-15. 10 trials used a number line from 0-25. The last 10 trials used a number line from 0-100. The number lines had no markings except for the starting and ending points. There were no practice trials in this task. Although comparisons of number-line placements on different trials involving multi-digit numbers initially were considered to be diagnostic of base-ten understanding, analysis of the data suggested that many children used placement strategies that rendered such comparisons uninformative. Accordingly, performance the number line estimation was deemed not to be diagnostic of an understanding of base-ten compositionality.

### **Social-cognitive assessments**

Below, we describe the social-cognitive tests and indicate which ones were deemed to be diagnostic of understanding the compositional structure of mental states.

1. *Gaze Intruder* (18 trials): Six faces were displayed on the screen. Five faces looked in one direction, while the sixth looked in a different direction. Children were asked to point to the face looking in a different direction. Trials ranged from easy (e.g., five faces looking to the left while the “intruder” is looking to the right), to difficult (e.g., five faces looking to the left, while the “intruder” looks to the left and slightly downwards). Stimuli were drawn from the Radboud and NIMH face stimulus sets (Egger et al., 2011; Langner et al., 2010). There were

4 practice trials in this task. This task was not diagnostic of understanding the compositional structure of mental states.

2. *Emotion Intruder (18 trials)*: Six faces were displayed on the screen. Five faces expressed the same emotion, while the sixth expressed a different emotion. Children were asked to “point to the person who looks like they’re feeling something different from everybody else.” Faces were drawn from the Radboud and NIMH face stimulus sets (Egger et al., 2011; Langner et al., 2010). There were 4 practice trials in this task. This task was not diagnostic of understanding the compositional structure of mental states.
3. *Say the Emotion (9 trials)*: a single face was displayed on the screen and children were asked to verbally name the emotion that they thought the person was feeling. Faces were drawn from the Radboud and NIMH face stimulus sets (Egger et al., 2011; Langner et al., 2010). There were no practice trials in this task. This task was not diagnostic of understanding the compositional structure of mental states.
4. *Oral Social Scenarios (13 trials)*: The experimenter verbally presented short scenarios that involved one or more characters who expressed a preference or desire. Children were asked to pick the most likely action or outcome from two options based on the character’s preference or desire. Importantly, children were told that they could ask the experimenter to repeat the story if they wanted to hear it again before answering. There were no practice trials in this task. All trials on this task were deemed to be diagnostic of understanding the compositional structure of mental states as they related preferences or desires to intentions.
5. *Point to Emotion (9 trials)*: Six faces expressing different emotions were displayed on a screen. Children were asked to point to the emotion the experimenter verbally asked for (e.g. “point to the person who looks happy). Faces were drawn from the Radboud and NIMH face

stimulus sets (Egger et al., 2011; Langner et al., 2010). There were no practice trials in this task. This task was not diagnostic of understanding the compositional structure of mental states.

6. *Theory of Mind (17 trials)*: Various theory of mind problems using pictures and verbal descriptions were presented to children. Some of them were similar to the Sally-Anne task (Baron-Cohen, Leslie and Frith, 1985) where children had to reason about a character's false beliefs about an object and its location. Some of them required children to engage in perspective taking where the same object appears different from different perspectives (Flavell et al., 1981). Some of them were similar to the "Smarties" task (Gopnik & Astington, 1988) where children had to reason about alternative representations of the same object. Finally, some of them asked participants to infer an agent's preference based on the amount of effort an agent put in order to reach a goal (adapted from Jara-Ettinger et al, 2015). Children were asked to produce verbal answers. There were no practice trials in this task. All trials on this task were deemed to be diagnostic of understanding the compositional structure of mental states.
7. *Park Play (4 trials)*: A map of a park and zoo was displayed on the screen. Children were told that there are three activities displayed in each of the maps but only two were available on a given day. An agent is naïve to which two they were. The agent's path from one location to the other was described by an experimenter and children were asked to guess what the agent's favorite activity was based upon on their path. In the park map the agent expressed a negative emotion (disappointment) when one activity was available over another, as an additional cue to the correct answer. In the zoo map, the agent expressed no emotions based on the availability of different activities. There were no practice trials in this task and two questions

per map were asked. All trials on this task were deemed to be diagnostic of understanding the compositional structure of mental states.

### **Domain general cognitive measures**

*Executive Function (30 trials organized into 3 rounds of 10 of increasing difficulty):* Two faces with different colors (green or purple) were displayed on the screen. Children were asked to press a key on the same side or the opposite side relative to where the green and purple faces appeared. For the first round, children were asked to press the key on the same side as the purple face. For the second round, children were asked to press the key on the opposite side as the green face. For the final round, either the green or the purple face appeared at random requiring children to alternate between pressing the key on the same side versus the opposite side of the face. Children were asked to keep their index fingers on both keys at all times during this test. There were 8 practice trials in this task and before each round, the relevant practice trials were presented. This test was based on Dillon et al. (2017).

*Vocabulary (30 trials):* Four pictures were displayed on the screen and the experimenter verbally presented the name of one of them. Children were asked to point to the picture that they think best fits with the word. This task was adapted from the PPVT (Dunn, 1997) with words varying in difficulty from the level of a 5-year-old's vocabulary to a 12-year-old's vocabulary.

### **Assessment order**

The assessments were presented in the following order to children in both versions: Say the Number, Gaze Intruder, Point to Number, Emotion Intruder, Oral Math, Say the Emotion, Approximate Symbolic Comparison, Oral Social Scenarios, Speed Symbolic Comparison, Point to Emotion, Panamath, Theory of Mind, Number Line Estimation, Park Play, Vocabulary, and Executive Function.

### **7.2.3 Analyses**

#### **Base-10 and non-Base-10 composites**

The Base-10 composite was obtained by creating Z-scores of the average of percent correct on 29 trials from different numerical tests that were deemed to require some understanding of base-10 compositionality. The non-Base-10 composite was obtained by creating Z-scores of the average of percent correct on 37 trials from different numerical tests that were deemed to not require base-10 understanding as they included single-digit numbers or required approximate accuracy. All such trials are flagged in the descriptions of the numerical assessments in section 2.1.

#### **Caregiver reports of home-based gameplay**

From caregiver reports of home-based gameplay, we found that majority of the children played the game with their caregiver at home. Only 5 children played with a sibling on some days during the home-based gameplay period. The final sample only includes children who played their assigned game at least 6 times based on parent reports of gameplay. Based on descriptive statistics, we found that children in the social-cognitive condition took longer than the children in the number condition to play their game in the beginning, but the differences in time became smaller as the days progressed. The graph below depicts these trends.

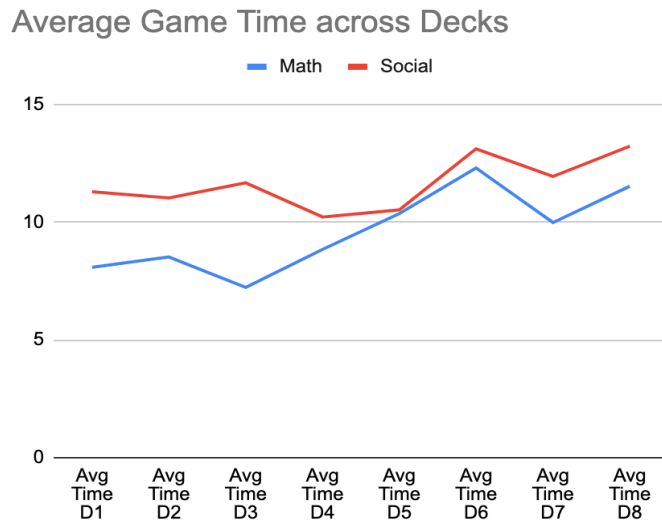


Figure S7.2.1. Average time (in minutes) taken to play each deck of *Find and Move* (in blue) and *Mind Hunter* (in red) based on parent reports of gameplay at home.

### Differences in pretest scores

In order to determine whether the pretest composite scores varied by condition, we conducted independent samples t-tests. There were no significant differences in the pretest number composite scores between the children in the number ( $M = -0.12$ ,  $SD = 0.66$ ) and those in the social-cognition condition ( $M = -0.09$ ,  $SD = 0.78$ ),  $t(47) = -0.14$ ,  $p = 0.89$ . However, children in the social-cognition condition had higher pretest social-cognitive scores ( $M = -0.05$ ,  $SD = 0.5$ ) than those in the number condition ( $M = -0.36$ ,  $SD = 0.51$ ),  $t(47) = -2.08$ ,  $p < 0.05$ .

### Test-retest reliability of the assessment battery

In order to determine the effectiveness of the assessment battery, we computed the test-retest reliability for the number and social-cognition composite scores as well as the individual tests in the assessment battery. We did so by correlating the composite scores of the assessment tests at pretest and posttest. We found a positive correlation between the number composite pretest

and posttest scores,  $r(46)=0.81$ ,  $p< 0.01$ . We found a positive correlation between the social-cognition pretest and posttest scores but it was lower than the number composite score,  $r(46)=0.5$ ,  $p< 0.01$ . The table below summarizes the correlation coefficients between the pretest and posttest scores of each test in the assessment battery separately. In general, we found that the correlations for the numerical tests were higher than the social-cognitive tests.

Table S7.2.1. Correlation coefficients between pretest and posttest scores of each test in the assessment battery

| S.No | Test Name                            | Correlation coefficient (r) |
|------|--------------------------------------|-----------------------------|
| 1    | Oral Math (OM)                       | 0.56                        |
| 2    | Panamath                             | 0.2                         |
| 3    | Point to Number (PtN)                | 0.49                        |
| 4    | Say the Number (StN)                 | 0.81                        |
| 5    | Speed Symbolic Comparison (SSC)      | 0.64                        |
| 6    | Approximate Symbolic Comparison (VM) | 0.44                        |
| 7    | Number Line Estimation (NLE)         | 0.63                        |
| 8    | Gaze Intruder (GI)                   | 0.53                        |
| 9    | Emotion Intruder (EI)                | 0.5                         |
| 10   | Oral Social Scenarios (OSS)          | 0.27                        |
| 11   | Park Play (PP)                       | 0.23                        |
| 12   | Point to Emotion (PtE)               | 0.14                        |
| 13   | Theory of Mind (TM)                  | 0.26                        |
| 14   | Say the Emotion (StE)                | 0.19                        |

|    |                         |      |
|----|-------------------------|------|
| 15 | Executive Function (EF) | 0.46 |
| 16 | Vocabulary (Voc)        | 0.25 |

### **Correlations within and across domain in the assessment battery**

In order to further evaluate the assessment battery, we computed correlations within and between domains (i.e., number and social cognition) both at pretest and posttest. The figures below depict the results.

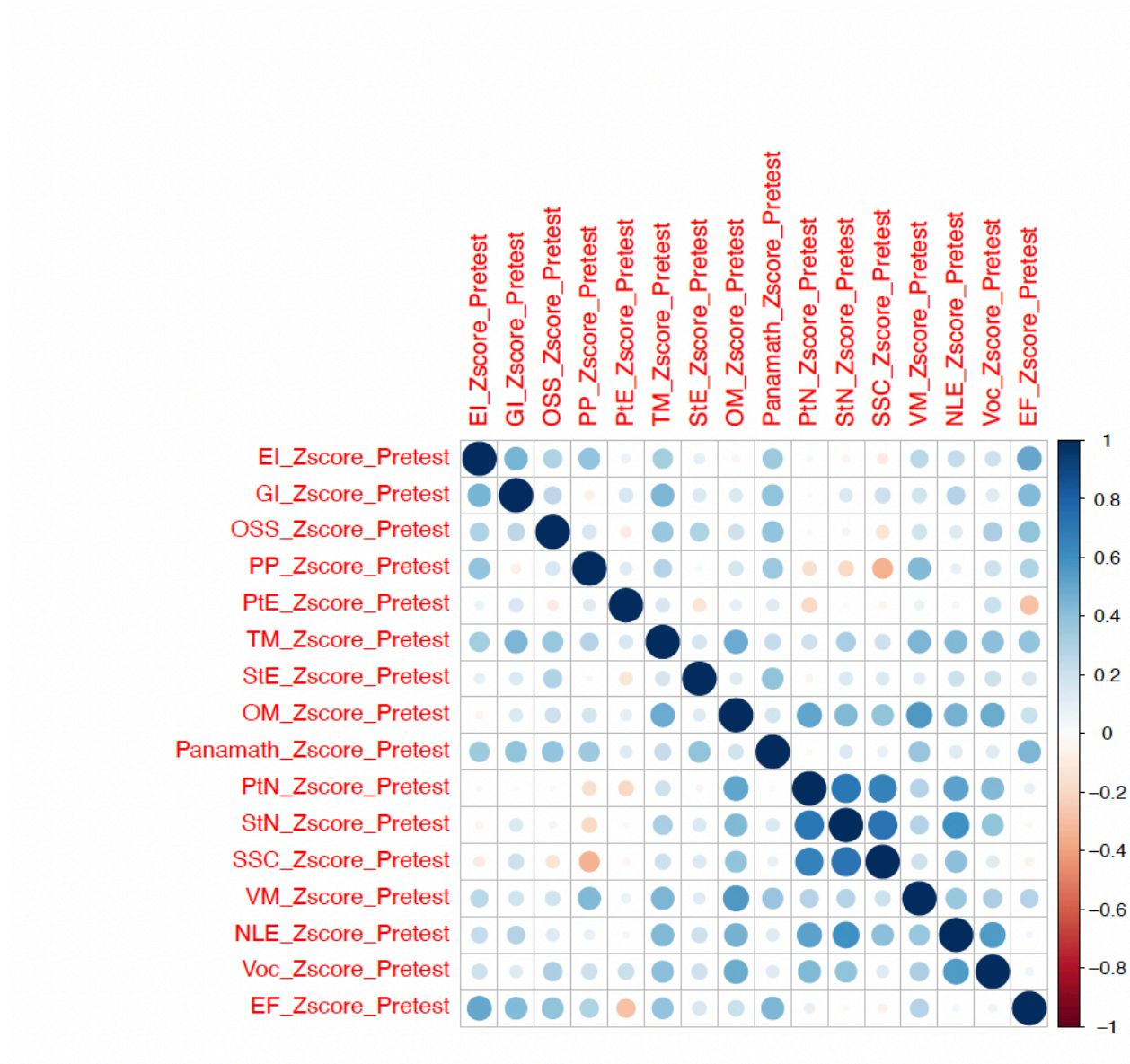


Figure S7.2.3. Correlation plots of pretest scores of each test in the assessment battery. The colored bar on the right depicts the values of the correlation coefficients.

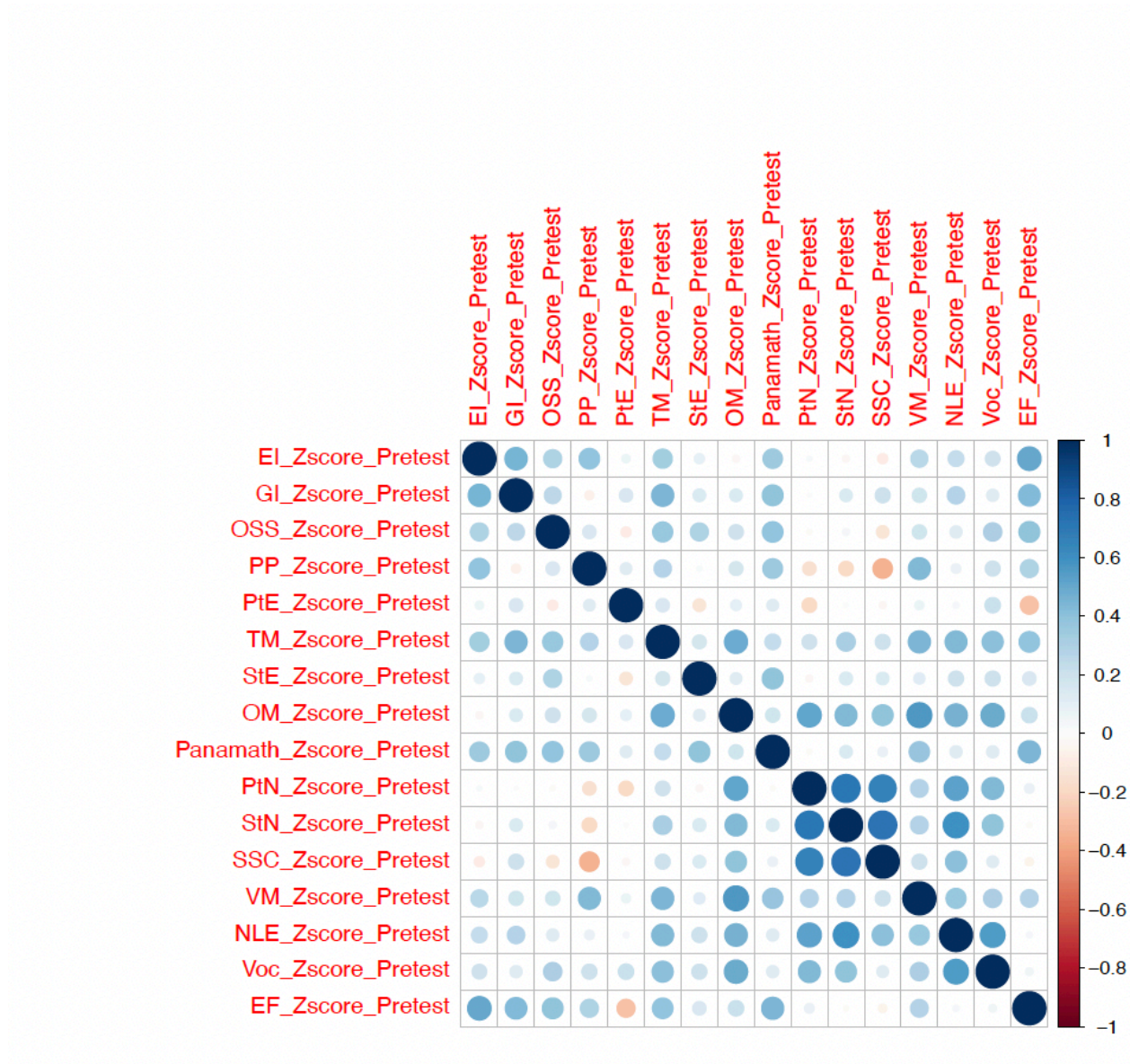


Figure S7.2.4. Correlation plots of posttest scores of each test in the assessment battery. The colored bar on the right depicts the values of the correlation coefficients.

### Number Line Estimation Analyses

In order to determine if children used their understanding of multidigit numbers in the *Number Line Estimation* test, we created pairs of trials from the same number line where the larger number in the pair had a smaller digit in the units place as compared to the smaller number (e.g.,

24 and 19). We created 8 such pairs in VA and 7 such pairs in VB of the test. For each pair, we coded whether children placed the larger number to the right of the smaller number on the number line. We then calculated the percentage of trials for which this was true for each child. We regressed these scores as a function of time, condition, and their interaction (including random intercepts for participants) and found no significant time by condition interaction ( $b = 11.94$ ,  $\beta = 0.50$ ,  $t = 1.75$ ,  $p = 0.09$ , 95% CI of  $\beta$   $[-0.07, 1.08]$ ). Therefore, children in the number condition were not more likely to correctly place a multidigit number on the number line as compared to children in the social-cognition condition. If anything, the results were trending in a direction where children in the number condition did worse at posttest as compared to children in the social-cognition condition.

We also applied the same analysis to pairs of numbers where the larger number in the pair had a larger digit in the units place as compared to the smaller number (e.g., 20 and 33) and found no significant time by condition interaction ( $b = 5.19$ ,  $\beta = 0.25$ ,  $t = 0.72$ ,  $p = 0.48$ , 95% CI of  $\beta$   $[-0.45, 0.96]$ ). The results were trending in a direction where children from both conditions did worse at posttest as compared to the pretest.

Moreover, while analyzing the trial by trial data for each child, we discovered that most children's placement strategies on the number line were influenced by their estimate on the previous trial, even though the markings from a previous trial were never available on the next trial. Therefore, we did not include this test while creating the composites to measure children's understanding of base-10 compositionality.

### **Moderating effects of domain general cognitive skills on children's performance**

In order to determine if the effect of *Find and Move* on children's performance on the number composite score was moderated by their Executive Function scores, we conducted a three-

way interaction of time, condition, and Executive Function scores (including random intercepts for participants). We found no significant time by condition interaction ( $b = 0.02$ ,  $\beta = -0.27$ ,  $t = 0.1$ ,  $p = 0.92$ , 95% CI of  $\beta$   $[-0.77, 0.23]$ ). A similar analysis with the social cognition composite scores also revealed no significant time by condition interaction ( $b = -0.32$ ,  $\beta = -0.57$ ,  $t = -1.92$ ,  $p = 0.06$ , 95% CI of  $\beta$   $[-1.17, 0.02]$ ).

Similar analyses with the Vocabulary scores also revealed no significant time by condition interaction, for the number ( $b = -0.06$ ,  $\beta = -0.09$ ,  $t = -0.41$ ,  $p = 0.68$ , 95% CI of  $\beta$   $[-0.53, 0.35]$ ) or social-cognition composite ( $b = 0.02$ ,  $\beta = 0.03$ ,  $t = 0.1$ ,  $p = 0.92$ , 95% CI of  $\beta$   $[-0.61, 0.68]$ ), indicating that children's domain general cognitive skills did not moderate their performance on the numerical assessments.

## References

- Baron-Cohen, S. Leslie, A. & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition* 21, 37–46.
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, 98(3), 199–222. <https://doi.org/10.1016/j.cognition.2004.09.011>
- Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences*, 102(39), 14116–14121. <https://doi.org/10.1073/pnas.0505512102>
- Bartsch, & Wellman, H. M. (1995). *Children talk about the mind*. Oxford University Press.
- Basargekar, A., & Lillard, A. S. (2021). Math achievement outcomes associated with Montessori education. *Early Child Development and Care*, 191(7–8), 1207–1218. <https://doi.org/10.1080/03004430.2020.1860955>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, 83(3), 223–240. [https://doi.org/10.1016/S0010-0277\(02\)00005-7](https://doi.org/10.1016/S0010-0277(02)00005-7)
- Bell, G. (1990). Language and counting: Some recent results. *Mathematics Education Research Journal*, 2, 1–14.
- Bugden, S., Szkudlarek, E., & Brannon, E. M. (2021). Approximate arithmetic training does not improve symbolic math in third and fourth grade children. *Trends in Neuroscience and Education*, 22, 100149. <https://doi.org/10.1016/j.tine.2021.100149>
- Byrge, L., Smith, L. B., & Mix, K. S. (2014). Beginnings of Place Value: How Preschoolers Write Three-Digit Numbers. *Child Development*, 85(2), 437–443. <https://doi.org/10.1111/cdev.12162>
- Cankaya, O., LeFevre, J.-A., & Dunbar, K. (2014). The role of number naming systems and numeracy experiences in children's rote counting: Evidence from Turkish and Canadian children. *Learning and Individual Differences*, 32, 238–245. <https://doi.org/10.1016/j.lindif.2014.03.016>
- Carey, S. (2009). *The Origin of Concepts*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367638.001.0001>

- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological science in the public interest*, 19(1), 5-51.
- Ceylan, M., & Aslan, D. (2018). Cardinal Number Acquisition of Turkish Children. *Journal of Education and E-Learning Research*, 5(4), 217–224. <https://doi.org/10.20448/journal.509.2018.54.217.224>
- Chan, W. W. L. (2014). Strategic counting: A novel assessment of place-value understanding. *Learning and Instruction*, 17.
- Chan, W. W. L., Au, T. K., & Tang, J. (2014). Strategic counting: A novel assessment of place-value understanding. *Learning and instruction*, 29, 78-94.
- Cheng, Y.-L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development*, 15(1), 2–11. <https://doi.org/10.1080/15248372.2012.725186>
- Cheung, P., & Ansari, D. (2021). Cracking the code of place value: The relationship between place and value takes years to master. *Developmental psychology*, 57(2), 227.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.
- Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core Knowledge of Geometry in an Amazonian Indigene Group. *Science*, 311(5759), 381–384. <https://doi.org/10.1126/science.1121739>
- DeLoache, J. (1987). Rapid change in the symbolic functioning of very young children. *Science*, 238(4833), 1556–1557. <https://doi.org/10.1126/science.2446392>
- Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. *Science*, 357(6346), 47-55.
- Dowker, A., Bala, S., & Lloyd, D. (2008). Linguistic Influences on Mathematical Development: How Important Is the Transparency of the Counting System? *Philosophical Psychology*, 21(4), 523–538. <https://doi.org/10.1080/09515080802285511>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology*, 43(6), 1428.
- Dunn, Lloyd M. PPVT-III : Peabody Picture Vocabulary Test. *Circle Pines, MN :American Guidance Service*, 1997.
- Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of

- children's facial emotion stimuli. *International Journal of Methods in Psychiatric Research*, 20(3), 145–156. <https://doi.org/10.1002/mpr.343>
- Evans, D. K. & Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3), 532-540.
- Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3), 295–313. <https://doi.org/10.1016/j.cognition.2004.09.010>
- 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>
- Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, 91(2), 173-190.
- Flavell, J. H. (1986). The development of children's knowledge about the appearance–reality distinction. *American Psychologist*, 41(4), 418.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive psychology*, 15(1), 95-120.
- Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Fuson, K. C. (1990). Conceptual Structures for Multiunit Numbers: Implications for Learning and Teaching Multidigit Addition, Subtraction, and Place Value. *Cognition and Instruction*, 7(4), 343–403. [https://doi.org/10.1207/s1532690xci0704\\_4](https://doi.org/10.1207/s1532690xci0704_4)
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag.
- Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 21(3), 180–206. <https://doi.org/10.2307/749373>
- Fuson K.C., & Kwon, Y. (1992). Korean children's understanding of multidigit addition and subtraction. *Child Development*, 63(2), 491-506.
- Fuson, K. C., & Li, Y. (2009). Cross-cultural issues in linguistic, visual-quantitative, and written-numeric supports for mathematical thinking. *ZDM*, 41(6), 793–808. <https://doi.org/10.1007/s11858-009-0183-7>
- Gelman, R., & Gallistel, C. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

- Gibson, E. J. (1965). Learning to read. *Science*, 148, 1066-1072.
- Gibson, E. J. (1991). *An Odyssey in Learning and Perception*. Cambridge, MA: MIT Press.
- Gibson, E. J. & Levin, H. (1975). *The Psychology of Reading*. Cambridge, MA: MIT Press.
- Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, 447(7144), 589–591. <https://doi.org/10.1038/nature05850>
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child development*, 26-37.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306(5695), 496–499. doi:10.1126/science.1094492
- Griffin, S., Case, R., & Siegler, R. S. (1994). Classroom lessons: Integrating cognitive theory and classroom practice. In K. McGilly (Ed.), *Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure* (pp. 25–50). MIT Press.
- Guerrero, D., Hwang, J., Boutin, B., Roeper, T., & Park, J. (2020). Is thirty-two three tens and two ones? The embedded structure of cardinal numbers. *Cognition*, 203, 104331. <https://doi.org/10.1016/j.cognition.2020.104331>
- Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge: Types of parent number talk. *Developmental Science*, 14(5), 1021–1032. <https://doi.org/10.1111/j.1467-7687.2011.01050.x>
- Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, 25(10), 896–910. <https://doi.org/10.1016/j.tics.2021.07.008>
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455(7213), 665–668. <https://doi.org/10.1038/nature07246>
- Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review*, 27(3), 465–482. <https://doi.org/10.3758/s13423-019-01694-7>
- Ho, C. S.-H., & Fuson, K. C. (1998). Children's knowledge of teen quantities as tens and ones: Comparisons of Chinese, British, and American kindergartners. *Journal of Educational Psychology*, 90(3), 536–544.

- Huttenlocher, J., Newcombe, N., & Vasilyeva, M. (1999). Spatial Scaling in Young Children. *Psychological Science*, 10(5), 393–398. <https://doi.org/10.1111/1467-9280.00175>
- Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, 131(1), 92–107. <https://doi.org/10.1016/j.cognition.2013.12.007>
- Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2011). Flexible intuitions of Euclidean geometry in an Amazonian indigene group. *Proceedings of the National Academy of Sciences*, 108(24), 9782-9787.
- Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour*, 5(11), 1548–1554. <https://doi.org/10.1038/s41562-021-01118-4>
- Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., & Hyde, D. C. (2016). Effects of Non-Symbolic Approximate Number Practice on Symbolic Numerical Abilities in Pakistani Children. *PLOS ONE*, 11(10), e0164436. <https://doi.org/10.1371/journal.pone.0164436>
- Kheu, N. M., & Nunes, T. (2023). Learning Numbers in Different Languages: Chinese versus Malay. *Journal of Mathematics Education*, 10(2), 30-47.
- Klein, E., Bahnmüller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2013). Language influences on numerical development—Inversion effects on multi-digit number processing. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00480>
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly*, 48(1), 61-76.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40, e253.
- Laski, E. V., Ermakova, A., & Vasilyeva, M. (2014). Early use of decomposition for addition and its relation to base-10 knowledge. *Journal of Applied Developmental Psychology*, 35(5), 444–454. <https://doi.org/10.1016/j.appdev.2014.07.002>
- Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology*, 50(3), 853–864. <https://doi.org/10.1037/a0034321>
- Laski, E. V., Vasilyeva, M., & Schiffman, J. (2016). *Longitudinal Comparison of Place-Value and Arithmetic Knowledge in Montessori and Non-Montessori Students*. 2, 15.
- Lauer, J. E., & Lourenco, S. F. (2016). Spatial Processing in Infancy Predicts Both Spatial and Mathematical Aptitude in Childhood. *Psychological Science*, 27(10), 1291–1298. <https://doi.org/10.1177/0956797616655977>

- Lê, M.-L. T., & Noël, M.-P. (2020). Transparent number-naming system gives only limited advantage for preschooler's numerical development: Comparisons of Vietnamese and French-speaking children. *PLOS ONE*, 15(12), e0243472. <https://doi.org/10.1371/journal.pone.0243472>
- Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology*, 46(5), 1309–1319. <https://doi.org/10.1037/a0019671>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability: Approximate number system and math abilities. *Developmental Science*, 14(6), 1292–1300. <https://doi.org/10.1111/j.1467-7687.2011.01080.x>
- Lipton, J. S., & Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition*, 98(3), B57-B66.
- Mann, A., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2012). On the development of Arabic three-digit number processing in primary school children. *Journal of Experimental Child Psychology*, 113(4), 594–601. <https://doi.org/10.1016/j.jecp.2012.08.002>
- Marcruz, O. Y. L., Carrie, H. K. L., Manabu, K., Mayumi, T., & Kumpei, M. (2022). Understanding of base-10 concept and its application: A cross-cultural comparison between Japan and Singapore. *International Journal of Early Years Education*, 30(4), 766–780. <https://doi.org/10.1080/09669760.2020.1848525>
- Marcus, G. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.
- McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, 116(2), 204–216. <https://doi.org/10.1016/j.cognition.2010.05.003>
- McCrink, K., & Wynn, K. (2004). Large-Number Addition and Subtraction by 9-Month-Old Infants. *Psychological Science*, 15(11), 776–781. <https://doi.org/10.1111/j.0956-7976.2004.00755.x>
- Miller, K. F., Kelly, M., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163–177). Psychology Press.
- Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, 6, 56–60.
- Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development*, 2(3), 279–305.

- Miura, I. T., Kim, C. C., Chang, C., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development*, 59, 1445-1450.
- Miura, I., & Okamoto, Y. (2003). Language supports for mathematics understanding and performance. In A. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills* (pp. 229–242). Mahwah, NJ: Erlbaum.
- Miura, I. T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., & Han, J. H. (1999). Language supports for children's understanding of numerical fractions: Cross-national comparisons. *Journal of Experimental Child Psychology*, 74(4), 356–365.
- Miura, I., Okamoto, Y., Kim, C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: cross-cultural comparisons – France, Japan, Korea, Sweden and the United States. *Journal of Educational Psychology*, 85, 24–30.
- Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C. M., Steere, M., & Fayol, M. (1994). Comparisons of children's cognitive representation of number: China, France, Japan, Korea, Sweden, and the United States. *International Journal of Behavioral Development*, 17(3), 401-411.
- Mix, K. S., Bower, C. A., Yuan, L., Hancock, G. R., & Smith, L. B. (2023). Predictive relations between early place value understanding and multidigit calculation: approximate versus syntactic measures. *Educational Psychology*, 43(7), 795-813.
- Mix, K. S., Bower, C. A., Hancock, G. R., Yuan, L., & Smith, L. B. (2022). The development of place value concepts: Approximation before principles. *Child development*, 93(3), 778-793.
- Mix, K. S., & Cheng, Y.-L. (2012). The Relation Between Space and Math. In *Advances in Child Development and Behavior* (Vol. 42, pp. 197–243). Elsevier. <https://doi.org/10.1016/B978-0-12-394388-0.00006-X>
- Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General*, 145(9), 1206–1227. <https://doi.org/10.1037/xge0000182>
- Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young Children's Interpretation of Multidigit Number Names: From Emerging Competence to Mastery. *Child Development*, 85(3), 1306–1319. <https://doi.org/10.1111/cdev.12197>
- Mix, K. S., Smith, L. B., Stockton, J. D., Cheng, Y.-L., & Barterian, J. A. (2017). Grounding the Symbols for Place Value: Evidence From Training and Long-Term Exposure to Base-10 Models. *Journal of Cognition and Development*, 18(1), 129–151. <https://doi.org/10.1080/15248372.2016.1180296>

- Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities*, 32(5), 1837–1851. <https://doi.org/10.1016/j.ridd.2011.03.012>
- Newman, J., Bidjerano, T., Ozdogru, A. A., Kao, C., Ozkose-Biyik, C., & Johnson, J. J. (2007). What do they usually do after school? A comparative analysis of fourth-grade children in Bulgaria, Taiwan, and the United States. *Journal of Early Adolescence*, 27, 431–456.
- Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. Cambridge University Press.
- Niklas, F., & Schneider, W. (2014). Casting the die before the die is cast: The importance of the home numeracy environment for preschool children. *European Journal of Psychology of Education*, 29(3), 327–345. <https://doi.org/10.1007/s10212-013-0201-6>
- Olkun, S., Reçber, H., Ata, A., Özer, E., & Çelebi, Ö. (2011). Türk Çocuklarının Bilişsel Sayı Temsilinin Çinli, Fransız, Japon, Koreli, İsveçli ve Amerikalı Yaşlılarıyla Karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 079–090. [https://doi.org/10.1501/Egifak\\_0000001216](https://doi.org/10.1501/Egifak_0000001216)
- Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, 152, 278–293. <https://doi.org/10.1016/j.jecp.2016.07.011>
- Park, J., & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science*, 24(10), 2013–2019. <https://doi.org/10.1177/0956797613482944>
- Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, 133(1), 188–200. <https://doi.org/10.1016/j.cognition.2014.06.011>
- Perry, M. (2000). Explanations of mathematical concepts in Japanese, Chinese, and U.S. first and fifth grade classrooms. *Cognition and Instruction*, 18, 181–207.
- Piaget, J. (1952). The child's conception of number. London, UK: Routledge.
- Piantadosi, S. T. (2011). Learning and the language of thought. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. *Psychological Science*, 24(6), 1037–1043. <https://doi.org/10.1177/0956797612464057>

- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, 306(5695), 499–503. doi:10.1126/science.1102085
- Potier Watkins, C., Caporal, J., Merville, C., Kouider, S., & Dehaene, S. (2020). Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training. *Journal of Computers in Education*, 7(2), 183-212.
- Potier Watkins, C., & Dehaene, S. (2024). Can a Tablet Game That Boosts Kindergarten Phonics Advance 1st Grade Reading?. *The Journal of Experimental Education*, 92(1), 32-55.
- Ramani, G. B., & Siegler, R. S. (2008). Promoting Broad and Stable Improvements in Low-Income Children's Numerical Knowledge Through Playing Number Board Games. *Child Development*, 79(2), 375–394. <https://doi.org/10.1111/j.1467-8624.2007.01131.x>
- Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, 32(3), 146–159. <https://doi.org/10.1016/j.appdev.2011.02.005>
- Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental science*, 16(4), 610-621.
- Ross, S. (1986). The development of children's place-value numeration concepts in grades two through five. Presented at the annual meeting of the American Educational Research Association.
- Ross, S., & Sunflower, E. (1995). Place-Value: Problem-solving and written assessment using digit-correspondence tasks. Presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108(3), 662-674.
- Schleppenbach, M., Perry, M., Miller, K. F., Sims, L., & Fang, G. (2007). The answer is only the beginning: Extended discourse in Chinese and US mathematics classrooms. *Journal of Educational Psychology*, 99(2), 380.
- Schneider, R. M., Sullivan, J., Marušič, F., Žaucer, R., Biswas, P., Mišmaš, P., Plesničar, V., & Barner, D. (2020). Do children use language structure to discover the recursive rules of counting? *Cognitive Psychology*, 117, 101263. <https://doi.org/10.1016/j.cogpsych.2019.101263>
- Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—But not circular ones—Improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, 101(3), 545–560. <https://doi.org/10.1037/a0014239>

- Song, M., & Ginsburg, H. P. (1988). The effect of the Korean number system on children's counting: A natural experiment in numerical bilingualism. *International Journal of Psychology*, 23, 275–302.
- Spelke, E. S. (2011). Natural Number and Natural Geometry. In *Space, Time and Number in the Brain* (pp. 287–317). Elsevier. <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.
- Spelke, E. S. (2022). *What Babies Know: Core Knowledge and Composition Volume 1* (Vol. 1). Oxford University Press.
- Stigler, J. W., & Stevenson, H. W. (1992). The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education. New York: Summit Books.
- Szkudlarek, E., Park, J., & Brannon, E. M. (2021). Failure to replicate the benefit of approximate arithmetic training for symbolic arithmetic fluency in adults. *Cognition*, 207, 104521. <https://doi.org/10.1016/j.cognition.2020.104521>
- Towse, J., & Saxton, M. (1998). Mathematics across national boundaries. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 129–150). Hove, UK: Psychology Press.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402. <https://doi.org/10.1037/a0028446>
- Vasilyeva, M., Laski, E. V., Ermakova, A., Lai, W.-F., Jeong, Y., & Hachigian, A. (2015). Reexamining the language account of cross-national differences in base-10 number representations. *Journal of Experimental Child Psychology*, 129, 12–25. <https://doi.org/10.1016/j.jecp.2014.08.004>
- Vasilyeva, M., Laski, E., Veraksa, A. N., & Shen, C. (2016). Development of children's early understanding of numeric structure. *Psychology in Russia: State of the Art*, 9(3), 76–94. <https://doi.org/10.11621/pir.2016.0305>
- Wang, J., Lin, E., Tanase, M., & Sas, M. (2008). Revisiting the Influence of Numerical Language Characteristics on Mathematics Achievement: Comparison among China, Romania, and U.S. *International Electronic Journal of Mathematics Education*, 3(1), 24–46. <https://doi.org/10.29333/iejme/216>
- Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, 44(2), 588–596. <https://doi.org/10.1037/0012-1649.44.2.588>

- Winkler-Rhoades, N., Carey, S. C., & Spelke, E. S. (2013). Two-year-old children interpret abstract, purely geometric maps. *Developmental Science*, 16(3), 365–376. <https://doi.org/10.1111/desc.12038>
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155–193. doi:10.1016/0010-0277(90)90003-3
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220–251. doi:10.1016/0010-0285(92)90008-P
- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), B15–B25. [https://doi.org/10.1016/S0010-0277\(03\)00050-7](https://doi.org/10.1016/S0010-0277(03)00050-7)
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1–B11. [https://doi.org/10.1016/S0010-0277\(99\)00066-9](https://doi.org/10.1016/S0010-0277(99)00066-9)
- Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8(1), 88–101. <https://doi.org/10.1111/j.1467-7687.2005.00395.x>
- Yang, C. (2016). *The Linguistic Origin of the Next Number*. <https://ling.auf.net/lingbuzz/003824>
- Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition*, 189, 89–104. <https://doi.org/10.1016/j.cognition.2019.03.013>
- Yuan, L., Prather, R., Mix, K., & Smith, L. (2021). The First Step to Learning Place Value: A Role for Physical Models? *Frontiers in Education*, 6, 683424. <https://doi.org/10.3389/educ.2021.683424>