Title
The Roles of Language and Executive Function on Early Mathematics Among Emergent Bilinguals

Author
Lin, Grace C.

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

The Roles of Language and Executive Function on Early Mathematics Among Emergent Bilinguals

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Grace C. Lin

Dissertation Committee:
Associate Professor Susanne M. Jaeggi, Chair
Associate Professor Joshua Lawrence
Associate Professor Drew Bailey

2018

DEDICATION

To

my *Baba*, *Mama*, *Gege*, and *Meimei*s

for their unquantifiable, wordless support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ACKNOWLEDGMENTS

First, I would like to thank my parents and siblings. Thank you for understanding. Thank you for allowing me to selfishly pursue this at times irrational path and for giving me the courage, spirit, and—may I say—stubbornness to tackle the demons born out of my curiosity. Your quiet and gentle support is what carried me through.

From my real life family to my academic ones:
I want to thank my academic advisor and committee chair, Susanne Jaeggi. I thank my lucky stars every day that you have decided to come to Irvine. Thank you for all of your support and guidance—from pushing me to think more deeply in our scientific discussions to how to tackle health-related issues *scientifically*, to even your high recommendations for mobile games—you are the definition of a true mentor. I often think back to all the seemingly unrelated areas I was interested in researching six years ago, and I have been able to pursue them all because of you. Thank you.

I want to also thank my committee member and unofficial co-advisor, Josh Lawrence. I just realized that in all the craziness of last year's holiday season, my thank you holiday card never made its way to Oslo (and it was a card with word stuff!). So please allow me to express my gratitude here. Thank you for believing in me. Thank you for allowing me to be a part of the Bifrost project and to learn from you. Thank you for all of your kind words and encouragement. It really means a lot.

This dissertation and journey could not have been possible without Drew Bailey, Jamal Abedi, Julio Torres, and Penelope Collins. To Drew Bailey, thank you for providing me with such insightful feedback and for being the devil's advocate when I go too far down one line of research thinking. Thank you also for organizing the math discussion group and for involving me in its planning. To Jamal Abedi, thank you for the opportunity to learn from you and for entrusting me as your GSR for the three projects during my second year. Thank you for inspiring me to engage with measurements more critically. To Julio Torres, thank you for all your thoughtful feedback and suggestions regarding heritage language learners, language backgrounds, and language dominance. They have been influential in how I think and design my language-related projects. Finally, to Penelope Collins, thank you thank you thank you. None of this would have been possible if it weren't for you taking a chance on me as a Ph.D. student. Thank you also for your wisdom and emotional support.

Now a shout out to the Working Memory and Plasticity Lab. I am grateful to have met all the wonderful souls in the lab family from our sassy, quirky, fun lab managers Chelsea Pelleriti and Austin Moon to the hardworking RAs who are always playing musical chair due to our space limitation. I would like to give special thanks to a few members. To Masha Jones, thanks for always being so level-headed whenever I was in or was heading toward a downward spiral. Your energy, perspective, and absolute work-life balance (or should I say blend?) have been absolutely inspiring. To Nancy Tsai, thanks for accompanying me on all those off-campus work sessions. This dissertation process would have been isolating if it weren't for the occasional parallel working/writing sessions we had at tea or coffee shops. To Jacky Au, thanks for pretty much letting me take over your desk. I sometimes feel guilty that I slowly started to put materials into the drawers on the left side of the desk meant for you… To Shafee Mohammed, thanks for being like a little brother, not to mention all the parking hours! Sharing an office with you has been enjoyable. There is nothing that alleviates stress like some anime playing in the background (and you know I'm saying this sincerely). To Emily Sumner, thank you for inspiring me to start thinking about Bayes and for being such a fun collaborator. To Sara Burt, thank you for your artistic rendering for the words and math problems. The EVT and lightbulb questions would not have been possible without you.

To my M2/B10/F1/FIT/C4/C5/M&M teams. First, yes, you all deserve a separate paragraph, especially the splendid, superhuman Snigdha Kamarsu. *Thank you* How many graduate students can say that they have a lab manager taking care of the day-to-day operations of the projects? Because of you, because I know that the RAs and the projects are in good hands, I was able to focus on my data and dissertation. Thank you. I would, of course, like to also acknowledge my RAs. Not only would the data not have been collected or scored if not for them, but they have also taught me how to be a research mentor. So thank you to Dorreen Sun, Chris Ngov, Alyssa Ford, Ratiana Karapet, Rachel Smith, Gabriela Hernandez, Daisy Jaimes, Ling Chen, Astrid Bartolo, Celena Alvarez, Viviana Garcia, Lizzie Lin, Patrick Liao, Tina (Ting Han) Yang, Christine Constantino, Cristal Martinez, Ellen Oum, Jacqueline Ngo, Eva (Ran) Yi, Priscilla Ng, Winnie Xu, Charlie Li, Haoyue Chang, Angelica Sheen, Aunnika Short, Ben Bui, Cheryl Huynh, Diego Valenzuela, Elaine Vo, Julia (Huiling) Zhang, Kimia Akhavein, Krismay Morena, Robert Chen, Susana Reyes, and Joseph Cachapero.

I would also like to extend my thanks to the team at the University of Maryland, College Park. To Geetha Ramani, thank you for always being so supportive and thoughtful. I learned a lot from how you devise the tests/games and from your graceful considerations as the project evolves. To Emily Daubert, thanks for being so fun to work with! Your lighthearted seriousness is something I strive to achieve. To Alaina Wodzinski, thank you for your insights into what testing conditions were like over there and changes that might be needed.

Thank you also to Jin Kyoung Hwang, Neil Young, Joyce Lin, Soobin Yim, Marcela Reyes, and Bianca Cung for your feedback and support on this dissertation and other aspects of graduate school life.

Last but not least, I would also like to express my gratitude to my friends "in real life" – Jenny Wang, Jenny Lee, Cathy Chai, Lucy Cheng, Lianne Yu, Tiffany Tsai, and Trina Osothprarop. Thank you for always reminding me that there is life outside of school. Thank you for keeping me balanced. To the first five of you, thank you for the frequent update phone calls, pictures, and/or videos of Angeline, Celestine, Ivan, Cason, Audrey, and Claire. You have no idea how therapeutic "ba-na-na," the simplest smiles from your babies, or the sound of your voices are. To Tiffany, thank you for introducing me to hot yoga. The time taken to work on my physical and mental wellness has been a tremendous asset to finishing this dissertation. To Trina, thank you for taking the time to hang out even though you live so far away and for sending me hearts on that game when we can't manage to meet up!

Portions of the research in this dissertation was supported by NSF Collaborative Research grants (1561404 & 1561447) to Susanne Jaeggi and Geetha Ramani.

CURRICULUM VITAE

GRACE C. LIN

3200 Education, School of Education
University of California, Irvine
Irvine, CA 92697-5500
Email: graceCL@uci.edu

EDUCATION

2018 **PH.D., Education**
University of California, Irvine
Specialization in Learning, Teaching, Cognition, and Development *and* Language, Literacy, and Technology
Dissertation: The Roles of Language and Executive Function on Early Mathematics Among Emergent Bilinguals
Committee members: Susanne M. Jaeggi (chair), Joshua F. Lawrence, Drew Bailey

2016 **M.A., Education**
University of California, Irvine

2011 **ED.M., Mind, Brain, and Education**
Harvard University Graduate School of Education

2006 **B.A., Psychology,** *High Honors*, with a minor in Pre-Business, *summa cum laude*
New York University
College of Arts and Science

PUBLICATIONS

Peer-reviewed articles

Lawrence, J. F., Hagen, A., Hwang, J., **Lin, G.C.** and Melby-Lervag, M. (in press). Academic vocabulary and reading comprehension: Exploring the relationships across measures of vocabulary knowledge. *Reading and Writing*.

Reich, S. M., Kay, J. S., & **Lin, G. C.** (2015). Nourishing a partnership to improve middle school lunch options: A community-based participatory research project. *Fam Community Health, 38*(1), 77-86. doi: 10.1097/FCH.0000000000000055.

Ishak, S., Adolph, K. E., & **Lin, G. C.** (2008). Perceiving affordances for fitting through apertures. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 1501-1514.

Manuscripts submitted or in preparation:

Lawrence, J. F., **Lin**, **G. C**., Jaeggi, S., Krueger, N., Hwang, J., & Hagen, A. (under review). Making sense of words: Testing lexical processing hypotheses with measures for 62,954 words in English

Jones, M. R., **Lin, G. C.**, & Jaeggi, S. M. (under review). Self-generation of test materials: A study technique to promote active learning.

Lin, G. C., & Jaeggi, S. M. (in revision). Linguistic and nonlinguistic codes: A modified model of exact symbolic arithmetic.

Collins, P., & **Lin, G. C.** (in revision). The role of lexical features in vocabulary acquisition for young, Spanish-speaking English language learners.

Lawrence, J. F., Hwang, J., Hagen, A., and **Lin, G. C**. (in revision). What makes an academic word difficult to know?: Exploring lexical dimensions across novel measures of word knowledge.

Lin, G. C., & Jaeggi, S. M. (in preparation). Linguistic and domain-general cognitive influences on numeracy in emergent bilingual kindergartners.

Lin, G. C., & Jaeggi, S. M. (in preparation). Mandarin at a young age or typical English instruction: A Bayesian comparison of kindergartners' numerical cognition.

HONORS AND AWARDS

02/2018	Mentoring for Achievement and Excellence Recognition Office of Inclusive Excellence, University of California, Irvine
05/2017	Michael E. Martinez Prize for Outstanding Educational Research and Service Honorary Mention ($100) University of California, Irvine, School of Education
05/2017	Certificate of Teaching Excellence by Center for Engaged Instruction
04/2017	Certificate of Completion of UCI CIRTL by UCI Graduate Division
2016-2017	Pedagogical Fellowship ($3,050) UC Irvine Center for Engaged Instruction (Division of Teaching Excellence and Innovation)

MENTORSHIP

Undergraduate Research Opportunity Program (UROP)

2016-2018	Six advised projects awarded to 12 mentees: Rachel N. Smith; Tzu-Hsin (Kelsy) Chou and Sharmin Shanur; Dorreen (Shengzi) Sun, Priscilla Ng, and Ran (Eva) Yi; Gabriela Hernandez, Daisy Jaimes, and Dorreen Sun; Ling Chen	Totaling $6,421

Advised Honors Theses

2017-2018	Mentee: Joseph D. Cachapero. Tentative Title: *Individual Differences in Metamemory*
2016-2017	Mentee: Ling Chen. *Vocabulary Size of Emergent English-Mandarin Speakers*

SELECTED CONFERENCE PRESENTATIONS

Lin, G. C.*, Sumner, E. S., Kalinowski, R., Kamarsu, S., Wu, M., Parlett, C., Mohammed, S. (2017, October). *Media and Technology for Cognitive Enhancement throughout the Lifespan*. Panel session presented at the Digital Media and Learning Conference, Irvine, CA.

Lin, G. C., Kamarsu, S., Daubert, E., Wodzinski, A., Ramani, G., & Jaeggi, S. M. (2017, September). *Can Playing Memory Games Improve Executive Function Skills?* Poster session presented at Flux Congress, Portland, OR.

Lin, G. C., & Jaeggi, S. M. (2017, September). *Language x Working Memory = Numeracy Development? Investigating Numerical Knowledge among Emergent Bilinguals*. Poster session presented at the Workshop on Linguistic and Cognitive Influences on Numerical Cognition, Tuebingen, Germany.

Lin, G. C., Kamarsu, S., & Jaeggi, S. M. (2017, August). *Alvin the Alien to the Rescue! Child-friendly Games to Improve Children's Math and Memory*. Poster presented at the Games for Change Festival, New York, NY.

Sumner, E. S., **Lin, G. C.**, Melcon, R., Cooper, M., Kent, H., Lopez, A., & Somoza, A. (2017, August). *SHIMA: A Virtual Reality Game to Measure Risk Propensity*. Invited presentation at the VR for Change Summit in the Games for Change Festival, New York, NY.

Lawrence, J., Hwang, J. K., **Lin, G. C**., & Hagen, A. (2016, December). *What are Important Academic Words for Reading?* In session paper presented in the Symposium on Support for Academic Language Across Grade Levels and Domains at the Literacy Research Association Conference, Nashville, TN.

Lin, G. C., Lawrence, J., Jaeggi, S. M., Krueger, N., Hwang, J., & Hagen, A. (2016, November). *Polysemy and Semantic Precision: New Semantic Measures Extracted from WordNet.* Poster session presented at the Psychonomic Society's 57th Annual Meeting, Boston, MA.

Lin, G. C., Cung, B., & Wu, M. (2016, October). Interactive Tales. A demonstration presented at the Society for Research in Child Development – Technology and Media in Child Development Special Topic Meeting at Irvine, CA.

Lin, G. C., Lawrence, J., Jaeggi, S. M., Krueger, N., Hwang, J., & Hagen, A. (2016, October). *Polysemy and Semantic Precision: New Semantic Measures Extracted from WordNet.* Poster session presented at the Mental Lexicon Conference, Ottawa, ON, Canada.

Lin, G.C., & Cung, B. (2016, October 6). *Interactive Tales.* Technology showcase presented at the Digital Media and Learning Conference, Irvine, CA.

Lin, G. C., Cung, B., & Wu, M. (2016, June). Interactive Tales: A choose-your-own-adventure story game for learning languages. Talk presented at the Games for Engaged Learning (GEL) Network Meeting on Sensorimotor, Perceptual Learning and Training: SPLAT! at Boston, MA.

Lin, G. C., Collins, P., Liu, J., & Estrella, G. (2015, July). Vocabulary acquisition among Spanish-speaking English language learners. Poster session presented at the Society for the Scientific Study of Reading's 22nd Annual Meeting, The Big Island, Hawaii.

RESEARCH EXPERIENCE

Graduate Student Researcher April 2016 – July 2018

NSF-EHR funded project "*Domain-General and Domain-Specific Training to Improve Children's Mathematics*."

PIs: Susanne M. Jaeggi and Geetha Ramani. Irvine, CA and College Park, MD

The NSF-EHR funded project examines the effects of mathematics and working training tablet games on the mathematics skills of young children from low SES backgrounds. My contributions include manage, train, supervise team of coordinator and research assistants, coordinate research activities between research sites, play test tablet games, analyze data and prepare reports, and present findings to various stakeholders.

Project Researcher September 2015 – July 2018

B10 Project: How languages can affect mathematics learning

This project examines the effects of children's language backgrounds and the learning of more than one language on children's base-10 mathematics skills. It also aims to explore how the language-math relationship may be influenced by children's working memory capacity. I am the lead researcher on this project.

Project Researcher July 2015 – July 2018

Bifrost Project: New Semantic Measures. PI: Joshua Lawrence

Irvine, CA and Oslo, Norway

This is a computational linguistics project focused on new semantic measures extracted from WordNet. My contributions include gathering, cleaning, and merging word feature databases, analyzing relation of new vocabulary features with existing ones, and assisting with manuscript preparation.

Project Researcher 2013 - 2014

Vocabulary Project: Vocabulary Acquisition in ELLs. PI: Penelope Collins. Irvine, CA

This vocabulary project examined different types of vocabulary language skills in Spanish-speaking English language learners. I administered the tasks, cleaned, coded, and analyzed data, and wrote the research manuscript.

Graduate Student Researcher and Project Manager 2013 – 2014

Impact of the WRITE Program on English Learner Achievement and Teacher Instructional Practice

PIs: Eric Hass, West Ed; Jamal Abedi, UC Davis

Funding Agency: Institute of Education Sciences

Formative Assessment in Mathematics

PI: Jamal Abedi, UC Davis

Funding Agency: National Science Foundation

A Comprehensive Research-Based Computer Assessment and Accommodation System for ELL Students

PI: Jama Abedi, UC Davis

Funding Agency: Institute of Education Sciences

Project Researcher 2013

Youth Participatory Action Research (YPAR) at a Middle School.

PI: Stephanie Reich, UC Irvine, CA

Project Researcher 2013

iSelfControl. PIs: Kimberley Lakes, Sabrina Schuck, Natasha Emmerson,

UC Irvine Child Development Center

Research Assistant 2010 – 2011

Blurry and Fuzzy Black Hole Experiments. PI: Matthew Schneps

Harvard-Smithsonian Center for Astrophysics,

Laboratory for Visual Learning & Science Media Group

Undergraduate Research Assistant 2004 – 2006
Aperture, Sitting, and Weighted Vest Lean. PI: Karen Adolph
New York University Infant Action Lab

TEACHING EXPERIENCE

Post-Secondary Settings

Pedagogical Fellow 2016 – 2018
University of California, Irvine
Irvine, CA

- Design and lead pedagogical training seminars and advised new teaching assistants on strategies for fostering inclusive, equitable classroom environment and mentoring undergraduate students
- Pioneered the quarter-long sequence of the Teaching Assistant Professional Development Program (TAPDP) at the School of Education in Fall 2016
- Led one of the two-day TAPDP workshops for Engineering students in September 2017

Invited Guest Lectures

Lin, G. C. (2016, March). *Media and Child Development*. UC Irvine, School of Education.
Lin, G. C. (2015, November). *Educational Technology*. UC Irvine, School of Education.
Lin, G. C. (2015, January). *Media and Globalization*. UC Irvine, School of Education.
Lin, G. C. (2013, June). *Motivation and Media*. UC Irvine, School of Education.
Lin, G. C. (2012, November). *Motivation and Media*. UC Irvine, School of Education.

Teaching Assistant
University of California, Irvine

Education 50: Origins, Purposes, and Central Issues in K-12 Education	Fall 2015
Education 107: Child Development in Education	Fall 2012; Spring 2013; Winter 2016
Education 124/Chicano Latino Studies 183: Multicultural Education	Winter 2013, Summer 2013, Spring 2015
Education 352: Creating a Supportive & Healthy Environment for Student Learning in Secondary Classrooms	Spring 2016

PROFESSIONAL ACTIVITIES

Workshops Attended

From Good Intentions to Real Outcomes: Equity by Design in Learning Technologies
Digital Media & Learning Conference October 2017
Linguistic and Cognitive Influences on Numerical Cognition,
University of Tuebginen, Germany September 2017
VR Brain Jam,
Games for Change Festival July–August 2017
Unpacking Teacher Effectiveness in the Context of IES-Funded Proposals,
University of Virginia June 2017

SERVICE

Service to Profession

Journal Reviewer: Intelligence, AERA Open, Neuropsychologia

Conference Reviewer: Connected Learning Summit, UC Irvine Associate Graduate Student Symposium, Harvard Graduate School of Education Student Research Conference

University Service

Pedagogical Fellow	UC Irvine	2016-2018
Cohort Representative	Associated Doctoral Students of Education	2016-2017
School of Education Representative	UC Irvine Discover Science Initiative	2016-2017
AGS Council Member	UC Irvine Associate Graduate Students	2013-2014
Voting Member	UC Irvine Student Health Insurance Advisory Committee	2013-2014

PROFESSIONAL AFFILIATION

- American Association of Applied Linguistics (AAAL)
- Association for Psychological Science (APS)
- Psychonomic Society
- Society for Research on Child Development (SRCD), Asian Caucus

RESEARCH SKILLS

Research Methods: Ordinary Least Squares & Logistic Regression, Analysis of Variance, Bayesian ANOVA and Regression, Structural Equation Modeling, Item Response Theory, Discourse Analysis, Content Analysis

Research Software: DataVyu, PsychoPy, E-Prime

Statistical Software: SPSS, STATA, JASP, jamovi, JMP

LANGUAGES

- Native level in English and Mandarin Chinese (speaking, reading, writing)
- Native level comprehension and near-native level speaking in Taiwanese (Southern Min Dialect)
- Basic level in Japanese

ABSTRACT OF THE DISSERTATION

The Roles of Language and Executive Function on Early Mathematics

Among Emergent Bilinguals

By

Grace C. Lin

Doctor of Philosophy in Education

University of California, Irvine, 2018

Associate Professor Susanne M. Jaeggi, Chair

There is growing recognition that language and executive function may independently affect numerical cognition, but few studies have examined all three constructs simultaneously. In two studies, I examined the relationship between language and math while considering executive function in emergent bilingual kindergartners in mainstream English classes and Mandarin immersion classes.

The 11 measures—three for language, four for executive function, and four for mathematics—used in this dissertation were adequately reliable and valid for children attending typical English instruction classes. However, the measures showed a distinct structural pattern for children in the Mandarin immersion classes. Adjustments were therefore made for the studies in this dissertation. Implications of such differences were also discussed.

In Study 1, I investigated the relationship between linguistic, executive function, and mathematical skills in predominantly low-SES emergent bilingual kindergartners. Using structural equation modeling, I showed that language has a direct effect on executive function, and that executive function, in turn, has a direct effect on overall mathematics such that the effect

of language proficiency on mathematics appeared to be mediated through its effect on executive function.

In Study 2, I examined how learning a language with a base-10 transparent number system may or may not influence kindergartners' different mathematics performance. Using multiple regressions supplemented by Bayesian analyses, I demonstrated that being exposed to the base-10 transparent Chinese number system in Mandarin immersion classes is associated with higher performance in counting, comparing magnitude, and estimating the position of a number on a number line even after accounting for children's executive function skills.

Findings from this dissertation underscore the importance of considering interrelated constructs in cognitive research, as the story that emerged may broaden our understanding of the mechanism of human cognition. Furthermore, they tentatively show that monolingual children may be able to benefit from dual-language immersion programs beyond learning a second language.

CHAPTER 1

Introduction

Most children growing up worldwide, and increasingly more children growing up in the United States, are exposed to more than one language. Some speak a heritage language at home while learning and speaking English at school in their classrooms and with their peers. Others are monolingual native English speakers whose parents enroll them in language classes (whether afterschool or immersion language programs) in hopes that their children will grow up to be global citizens with abilities to communicate and think in different languages. Politically speaking, with the passage of bilingual education ("California Proposition 58, Non-English Languages Allowed in Public Education," 2016) in November 2016, the number of emergent bilingual[1] children will undoubtedly increase. How does the changing linguistic and learning landscape affect children's cognition?

Additionally, while it may be logical that the changing linguistic environment may have influences on children's linguistic development (for one thing, they would be learning to speak more than one language), does it affect children's cognition in other domains, such as mathematics?

This dissertation seeks to take a stab at answering the broad question by investigating the relations between language, mathematics, and executive function among emergent bilingual children experiencing this linguistic environment. In the following sections, I lay down the theoretical foundations for connecting the three constructs, discuss the rationale for studying emergent bilingual children, and provide an overview for the rest of the dissertation chapters.

[1] Note that the term "emergent bilinguals" is coined as an alternative to "English learners," which may convey a more deficit point of view (García, 2009). In this dissertation, I use the term "emergent bilinguals" to encompass all children who are learning and/or using a nonnative language at school.

Language Influence on Math

In the realm of mathematics and numerical cognition, one of the most influential theories is the triple-code model (Dehaene, 2011; Dehaene & Cohen, 1997). The model specifies that for numerical processing, there are three interrelated constructs or codes: Arabic code where people read and interpret the Arabic numerals, verbal code where the phonology of numbers and certain arithmetic facts (e.g., 4 times 2 equals 8) are stored, and the quantity code most in line with the approximate number system and is in charge of quantitative mathematical skills such as magnitude representation and comparison. The three codes work in parallel anytime a person tries to tackle a math problem. Dehaene and colleagues have been able to distinguish the three codes and differentiate the contribution of quantity and linguistic influences by examining the neural networks that were involved in different tasks. For example, for the purely mathematical, quantity code, neuroimaging studies have implicated the posterior superior parietal lobule (PSPL) and the horizontal segment of the bilateral intraparietal sulcus (hIPS) to be responsible for quantity representation and manipulation, regardless of the form or notation of the numbers (for a review, see Dehaene, Molko, Cohen, & Wilson, 2004). In contrast, neuroimaging and neuropsychological evidence suggests that regions of the left hemisphere responsible for language comprehension are also implicated in arithmetic processes (e.g., Baldo & Dronkers, 2007). The triple-code model thus contributes to our understanding of the differentiated mechanism through which language may influence math processing.

The model has gained increasing popularity in the past few decades with the surging interest in cross-domain and interdisciplinary work, as it draws attention to how language is inseparable from numerical processing.

Since then, work relating language and math has expanded to finer-grained aspects of varying, distinct constructs within mathematics, language, or other general domains in attempts to better elucidate the mechanism of numerical cognition. For instance, the Pathways to Mathematics Model (LeFevre et al., 2010) specifies three pathways of linguistic, quantitative, and spatial attention, each independently contributing to distinct math outcomes. According to the pathways model, the linguistic skills would contribute most to acquiring and understanding the symbolic number system, which in turn would most strongly predict math skills such as number line estimation and calculation. On the other hand, the pathway for quantitative cognitive skills would lead to numerical magnitude processes, which would most strongly predict math performance in magnitude comparison (this pathway is very similar to the triple code's "quantity code"). Because different mathematical outcomes appear to be predicted by different cognitive skills, in this dissertation, I use multiple measures of mathematics (see Chapter 2 for details).

Similar to how researchers focused on mathematics break math outcomes down into multiple math skills, researchers focused on the linguistic influences of mathematics have delineated the roles of various language properties. Departing from the somewhat limited view of the verbal code, which focused mainly on phonological processing, the various language properties now include syntax, phonology, reading direction, and lexical features of particular languages; they also sometimes consider individual abilities related to languages such as verbal working memory as a type of linguistic properties (see special issue edited by Dowker & Nuerk, 2016). In this expanded view, the grammatical structure of different languages, for example—some with singularity/plurality markers and others without—has been shown to influence young children's acquisition of the cardinality principle (Sarnecka, Kamenskaya, Yamana, Ogura,

&Yudovina, 2007). It is therefore important to consider and distinguish the kind of linguistic property a study chooses to investigate.

In line with previous interdisciplinary efforts to connect language and mathematics, this dissertation aims to examine linguistic influences on math from two differing language perspectives:

1) language in terms of proficiency in language of instruction, i.e., language at an individual level, and
2) language in terms of its lexical structure and how its structure may influence math even among new learners of the language, i.e., language at the broader lexical level.

Executive Function and (Emergent) Bilingualism

The relationship between math and language has been explored via multiple studies that sprang from theories such as the triple-code model. However, the picture is not complete without considering the multiple demand, general cognitive system of executive function. To be specific, I use the term "executive function" (EF) in this dissertation as a canvas term for different domain-general cognitive control skills, as they have been coined differently by researchers following different traditions and lines of research. For example, Miyake and Friedman (2012) consider executive function to consist of three components of updating, shifting (or switching), and inhibition. Other researchers consider this domain general cognitive skill as working memory (Baddeley, 1992; Engle, 2002) or even executive attention (Engle, 2002, 2010). Despite the differing terminology, there is a general consensus for the contribution of EF to other skills. The general EF processes regulate and control individuals' attention, selecting relevant information, suppressing irrelevant ones, and allowing individuals to carry out tasks with multiple demands (Engle, 2002, 2010). As mathematical tasks are inherently complex—even basic tasks like counting would be complex for young children first learning the numbers—executive function skills are required. Similarly, linguistic processing would also require EF. Yet, with the exception of the Pathways model (LeFevre et al., 2010) or Geary and Hoard's (2005) framework for approaching the study of mathematical disabilities, few have attempted to unify all three constructs of language, mathematics, and executive function. Even in these models, the three constructs are not equally or fully represented. In the Pathways model, for example, the third cognitive skills domain was only one aspect of executive function, "spatial attention." Geary and Hoard's (2005) model, on the other hand, emphasized executive function and considered language only as a support system. Considering how both math and linguistic processing would require EF, examining all three constructs together is essential.

Disentangling the relations among the three constructs will be no easy feat. Examining the connection between two of the constructs—EF and language—already reveals a complex situation further complicated by the language history or background of individuals. Specifically, a hotly debated field of research focuses on the idea of bilingual advantage when it comes to EF. To examine linguistic influences on math while considering EF, it is therefore essential to take individuals' language background into account.

Bilingual Advantage? As mentioned previously, a growing population of children in the United States consists of (emergent) bilinguals, but how does bilingualism affect learning? The language background of participants may be of particular importance because previous research has established different patterns of language learning and cognitive effects in bilingual individuals as compared with monolinguals (Kroll, Bobb, & Hoshino, 2014). Indeed, neuroimaging studies have suggested that learning a second language, even in adults, is associated with structural differences in multiple brain regions implicated in language processing such as the left inferior fronto-occipital fasciculus (IFOF), left superior longitudinal fasciculus (SLF), and corpus callosum, and executive function such as the anterior thalamic radiation, forceps minor, or cingulum-hippocampus (e.g., Kuhl et al., 2016). Additionally, the neural correlates for suppressing interference appear to be different between monolingual and bilingual individuals (Luk, Anderson, Craik, Grady, &Bialystok, 2010).

However, even with converging evidence of structural differences, differences between bilingual and monolingual individuals in behavioral studies are not always consistent. One camp of researchers claims a bilingual advantage, especially in EF tasks such as task switching, inhibition, interference suppression, and selective attention (Barac & Bialystok, 2012; Bialystok & Viswanathan, 2009; Esposito, Baker-Ward, & Mueller, 2013; Kroll & Bialystok, 2013; Morales, Calvo, & Bialystok, 2013; Poulin-dubois, Blaye, Coutya, & Bialystok, 2011). For example, a comparison of 40 Portuguese-Luxembourgish bilingual children with 40 monolingual Luxembourgish children matched on multiple characteristics including age, classroom size, socioeconomic status, and even body mass index has demonstrated that bilingual children outperform the monolingual children in attention control (i.e., Sky Search and flanker) tasks (Engel de Abreu, Cruz-Santos, Tourinho, Martin, &Bialystok, 2012). The combined results from these studies suggest that there might really be a bilingual advantage.

Yet, another line of research challenges the stance of a “bilingual advantage.” For instance, bilingual individuals tend to underperform on tasks of verbal working memory (e.g., Fernandes, Craik, Bialystok, & Kreuger, 2007) and have lower vocabulary in their two languages compared with their monolingual peers (Bialystok & Luk, 2012; Hoff, Core, Place, & Rumiche, 2012). Other studies have shown no difference between bilingual and monolingual groups, though the null effects could possibly be attributed to favorable language learning environments where both groups of individuals were well supported (e.g., Thordardottir, 2011). Even in the components where the "advantages" are most often found, i.e., switching and inhibition of EF, the findings have not always been consistent. For example, the bilingual advantage findings appear to be associated with the sample size of the various studies (Paap, Johnson, & Sawi, 2014). Moreover, using classic tasks for inhibition and switching such as Simon, flanker, and antisaccade tasks, some researchers were not able to find any difference between monolingual and bilingual adults; in fact, for some measures, there were monolingual advantage (Paap & Greenberg, 2013; Paap & Sawi, 2014). With conflicting findings even with the control component of EF, more research is needed to explore the existence of potential behavioral differences that may be consequences of bilingualism and the reasons behind the conflicts.

Thus far, some studies have been able to pinpoint few potential explanations for the discrepancies in studies exploring the existence of bilingual advantage. For example, differential performance on verbal tasks (e.g., receptive vocabulary test, Wugs test, sentence construction test) appears to depend on the similarity between the two languages the learners speak (Barac & Bialystok, 2012; Bialystok, Majumder, & Martin, 2003). For example, French-English, Hebrew-English, and Spanish-English (all alphabetic languages) bilinguals exhibit the bilingual advantage for phonological awareness, but not Chinese-English (one logographic, one alphabetic language) bilinguals (Bialystok, Luk, & Kwan, 2005; Bialystok et al., 2003). Yet others point out working memory as the main contributor to differences in executive control, not bilingualism (Namazi & Thordardottir, 2010). Children's pattern recall performance was predictive of their performance on the Simon task, and the language groups were equally distributed among the low and high performing pattern recall groups (Namazi & Thordardottir, 2010).

With the discrepancies in findings on bilingualism and EF, particularly executive control, and the suggested role of working memory performance on executive control, it is imperative that EF—particularly the working memory components—be considered in investigating linguistic influences on math. Additionally, even if we suppose that cognitive changes do occur for bilingual individuals, it is still unclear how early this cognitive change occurs for monolingual children developing into bilingual individuals, hence “emergent” bilinguals. Few recent longitudinal studies have suggested that learning a second language in an immersion setting may enhance executive functions of young children after a few years (Nicolay & Poncelet, 2013, 2015). However, the effects may be limited to only certain aspects of executive function (e.g., attention) and it is yet unknown how replicable the findings could be, how long the children would have to be learning the second language for an effect to take place, and whether learning a second language starting at a young age has any effect on children’s learning of other cognitive domains.

Though numerous studies have been devoted to the cognitive effects of bilingualism on both executive function and language, how bilingualism may affect the relation of EF and language to *mathematics* performance is less clear. Drawing from Paivio’s (2014) bilingual dual coding theory (DCT) of logogens and imagens, bilingual individuals may experience separate sets of logogens in their two languages. These logogens are language-specific linguistic units (e.g., a word) that are connected to the nonverbal imagens where semantic information are stored (as an image, scene, or even event). Combining Paivio’s DCT with the triple-code model (see Figure 1.1), we can see how there may be two potential separate lines of entries for the verbal processing of numerical information for bilingual individuals. Although combined models such as the one presented in Figure 1.1 may provide insight into mechanism for bilingual math processing, research on linguistic influences on mathematics have not yet investigated such specific processing.

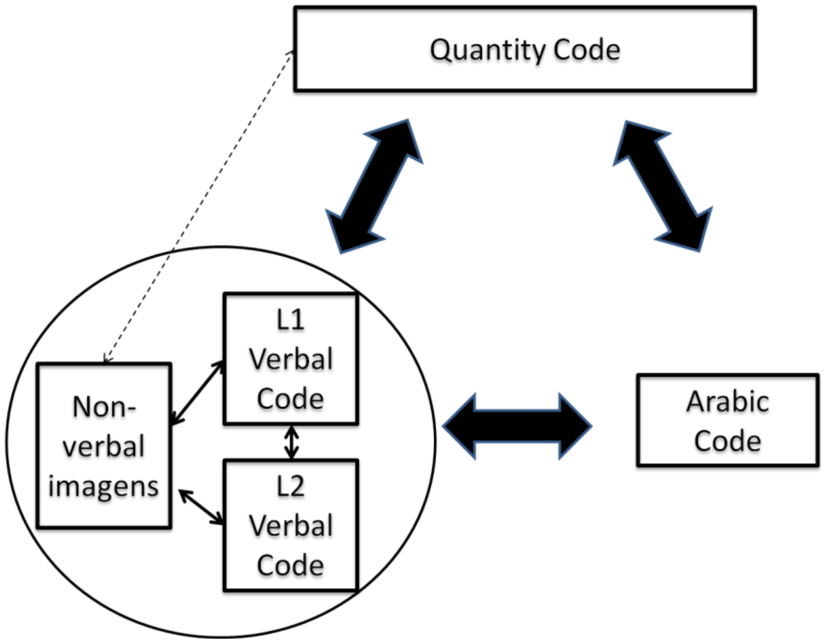


Figure 1.1. An integrative model of the triple-code model (Dehaene & Cohen, 1997) and the dual coding theory (Paivio, 2014).

Most research on bilingual individuals or emergent bilingual children focus on one domain-specific area (language) or the domain-general executive processing cognition. Working memory (i.e., updating) is actually one less studied executive control process among emergent bilinguals (Barac, Bialystok, Castro, & Sanchez, 2014, p. 711). Additionally, research that has been conducted on content areas such as mathematics tends to focus on one specific group of emergent bilinguals, English learners, especially on how to help English learners improve their mathematics performance (Bunch, 2013; Doabler et al., 2016; Robinson, 2010). It is less clear how native English-speaking emergent bilinguals (i.e., those who are receiving instruction in a non-English language) would fare on language, working memory, and mathematics performance.

To recap, a number of studies connect math with language (triple-code model, pathways model). Executive function is also heavily studied in conjunction with language, particularly in bilingual individuals. However, how the three constructs of language, math, and executive function interrelate as a whole is less explored.

Goals and Aims

Considering the links between language, math, executive function, and bilingual individuals, merely painting a picture of linguistic influences on math or considering language as a part of executive function (e.g., the role phonological loop plays) is incomplete. Therefore, this dissertation aims to answer aforementioned two types of linguistic influences on mathematics while also taking into account children's domain general executive function skills.

Overview of the Chapters

Chapter 2. From counting sheep to touching base: Reliability and validity of the mathematics, linguistic, and executive function measures

Chapter 2 serves as a separate methods paper that focuses on the psychometric properties of the tasks used in this dissertation. The in-depth examination of the measures describes the ten tasks[2] used in my studies, provides an overview of their administration procedures, and lays out the psychometric properties (i.e. reliability and validity) of these measures, as most are non-standardized tasks developed or modified by me, the Working Memory & Plasticity Lab at University of California, Irvine, and/or the Early Childhood Interaction Lab at the University of Maryland, College Park. I also briefly outline decisions for the subsequent studies based on the reliability and validity of the measures.

Chapter 3. Study 1. How do linguistic, executive function, and mathematics skills relate to each other in emergent bilingual kindergartners? A SEM analysis

[2] There are 10 tasks that make up 11 measures because the visuospatial working memory task is split into two measures (forward and backward).

In the first study, I explored the relationship between language, executive function, and mathematics among kindergartners attending regular English classes in three southern Californian schools. The majority of the participants are emergent bilingual students who speak another language at home. I created latent factors for the three constructs from 11 different measures and employed structural equation modeling (SEM) to investigate the direct and indirect effects of language and executive function on mathematics. Of particular interest in this study is whether language or executive function would act as a mediator without breaking down the three constructs into their individual components. In particular, can we draw any general conclusion regarding the mechanism of how language or EF affect math?

Chapter 4. Study 2. Learning Mandarin as a second language in an immersion setting vs. receiving mainstream English instructions: A Bayesian comparison of kindergartners' numerical cognition

The second study of my dissertation investigates the differences in mathematics skills between kindergartners enrolled in typical English instruction classes and those enrolled in Mandarin immersion classes. All children are reported to have English as their home language. To ensure that the language proficiency would not confound the results, I first compare their performance on vocabulary tasks. In contrast to Study 1, here I examine the effects on different math skills—number line estimation, counting, magnitude comparison, and number identification—separately. All analyses (t-tests and regression) are performed from a frequentist perspective and supplemented with Bayesian statistics. In the multiple regressions, I employ Bayes factors to inform the best model for each math skill. To examine the unique contribution of each predictor variables, I also report semipartial correlations along with their corresponding Bayes factor. The findings from this study would provide insight onto whether learning mathematics in a language where the base-10 place value system is transparent (i.e., 11 is ten-one, 12 is ten-two, 20 is two-ten) would have any effect on English speakers' various mathematical skills. I hypothesize that learning Mandarin may facilitate children's math learning even after controlling for students' EF performance or maternal education.

Structure of the Dissertation

Rather than delving directly into the two studies, I open with an in-depth examination of the measures used in the studies. The two studies that follow are presented as separate manuscripts, each with its own literature review, research questions, method, results, discussion, and conclusion. I end by providing a general summary and discussion of the studies, their implications, and the direction for future work.

CHAPTER 2

From counting sheep to touching base:

Reliability and validity of the mathematics, linguistic, and executive function measures

The objective of this chapter is to describe in detail the measures used in this dissertation—their administration and scoring—and provide the psychometric properties of these measures. In particular, I highlight the reliability and validity of the measures. In this chapter, I first provide a brief overview of reliability and validity, pointing out the type of reliability and validity test I employ for the measures. I then provide a detailed description of each of the tasks along with the reliabilities of these tasks for my test samples. Finally, I present the results of and discuss the findings from the validity tests.

The concept of reliability comes from classical test theory (CTT), which stipulates that an individual's test score is composed of her true score plus measurement error. Reliability of a test is then a proportion of the true score to the error; a test with a reliability of 0 means that everything was from measurement error, and a reliability of 1 indicates a perfect test with no measurement error. The reliability of the measures is therefore important, as low reliability means that any outcome variables derived from the measures would be indicative of measurement errors rather than individuals' true skills or abilities. There are multiple ways to estimate test reliability, such as test-retest, parallel form, split-half, and internal consistency reliability. Internal consistency reliability serves to check whether all test items measure the same construct as the total test measures; in essence, it tells the strength of the correlations among each individual item. I therefore choose to report internal consistency reliability in this chapter. As the individual measures used in this dissertation are unidimensional (see Sijtsma 2009 suggestion for internal consistency when tests are *not* unidimensional), I report internal consistency as measured by Cronbach's alpha.

Validity, on the other hand, tells us how appropriate an interpretation and use of a test may be (Messick, 1979). In other words, a test is valid if it appropriately measures what the user intended for it to measure (National Council on Measurement in Education, 2018). It is therefore essential to establish the validity of my 11 measures because my analyses and interpretations in subsequent chapters will be moot if, for example, magnitude comparison is not a valid measure of children's math skills. As with reliability, there are multiple forms of validity, such as criterion validity that focuses on how well measures correlate with similar measures (this is further broken down into predictive and concurrent validity) and construct validity, which focuses on whether a test can be interpreted as a measure of a construct, or a "postulated attribute of people" (Cronbach & Meehl, 1955, p. 178).

In this chapter, I examine my measures' validity by exploring both concurrent validity and construct validity. My 11 measures from 10 tasks came with a priori conjectures as to the constructs (language, math, or executive function) they are measuring. Therefore, measures within the same construct categories serve as each other's external criteria for concurrent validity, and the strength of this validity is measured via their correlation estimates. At the same time, these correlations provide information for convergent and discriminant validity based on the Multitrait-Multimethod (MTMM) matrix approach (Campbell & Fiske, 1959). (Details provided in the later Validity Section). To further explore construct validity of the measures, I employ factor analyses (both principal components analysis and confirmatory factor analysis) to illustrate if the measures indeed belong to their assigned constructs. Because the study involved participants from different types of schools who followed different testing procedures, I perform correlational and factor analyses separately for the different samples.

The Measures

Language

Bilingual Picture Vocabulary Test (PVT). To estimate children's language ability, I administered the bilingual version of the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin, 2013). For each question on the PVT, a child saw a set of four pictures while hearing a word stimulus. The experimenter asked the child to point to the picture that best matched the word he/she heard. The established reliability coefficients among 5- and 6-year-olds was 0.97 and 0.94, respectively, for the Spanish-English bilingual version (Martin, 2013). In order to increase testing consistency, we digitized the study using PsychoPy and recorded all word stimuli in English and Spanish. In order to assess children's acquisition of the target language, I manipulated the program such that children will hear stimuli in the language of instruction, English, first. If the children missed the word in English or indicated that they would like to hear the word in Spanish, the test administrator would click the "Spanish" button to play the word in Spanish. The reliability in our sample of 86 kindergartners who attend regular English instruction classes was 0.96.

For the children attending the Mandarin immersion classes, I devised a modified version of the PVT. All of the words from the Spanish-English version were first translated to Mandarin Chinese. The items without a direct translation equivalent were then removed from the overall test. For example, the last item on the test, spinet, means a small upright piano. Instead of having an equivalent word in Chinese, its translation is literally, "small upright piano." I removed the item when the translation of the word is its literal definition or may include confounding characters that would bias the answer (e.g., "barking" is translated into a two-character word with the first character meaning "dog"). In the end, 11 items (#47, 56, 75, 81, 96, 108, 153, 157, 165, 167, 180) were removed. The Chinese translations were then audio recorded using Audacity and saved with the Mandarin-English version of the PVT PsychoPy program. The Mandarin-English program was manipulated such that the children would hear stimuli in the language of instruction, Mandarin, first. Just like the other version, children were able to choose to hear the word in the other language if they so desire. Cronbach's alpha for the sample of 33 students attending the Mandarin immersion classes was 0.95.

The program—for both bilingual versions—started with a questionnaire asking for participants' age and language use. Based on the reported age, the program jumped to the corresponding question (e.g., item 25 for age 5 to age 6-11). A basal was established when a child answered eight consecutive questions correctly. The child reached ceiling when they answered four out of six consecutive questions incorrectly, and the test was then terminated. Typical of standardized vocabulary measures, all items below the basal were assumed to be correctly answered (score of 1 point each) and all items above ceiling were assumed to be incorrect (score of 0 each). The total raw score was calculated by subtracting the number of incorrect items from the highest number reached (e.g., if a child reached question 100, but got 8 questions wrong along the way, she would have a raw score of 92).

Expressive Vocabulary Test (EVT). Items from a picture-naming task developed for kindergartners taking English classes in Taiwan (Lin & Johnson, 2016) were translated into English. We created new pictures to go with the words and modified the expressions used to elicit children's response (see Figure 2.1 for an example). For each item, children saw the picture while an experimenter asked a question or gave an expression. For example, a picture of a bear teaching the alphabets is accompanied by the expression, “this bear is a ___ (teacher).” There were a total of 40 questions on the EVT, and each question was scored as 0 (incorrect), 0.5 (partial credit), or 1 (correct) point. An example of partial credit was when the item is trying to elicit the word “father” but the child gives “dad,” “daddy,” or “papa” as the response. (See Appendix I for a full set of questions and corresponding images.) The score reported in this paper is the raw points accumulated for all 40 questions. Even though the original task was designed to measure kindergartners’ Chinese vocabulary, we were able to use this EVT for English speakers as well. Cronbach’s alpha for the English version of this test with 75 participants attending English instruction classes was 0.85.

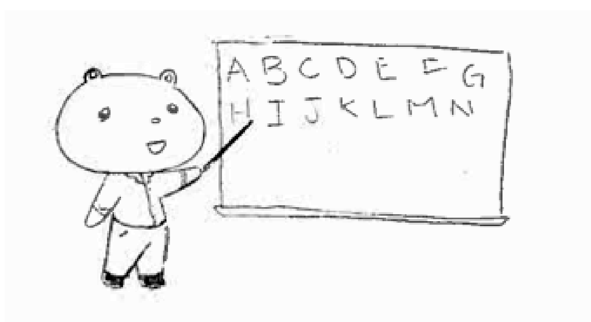
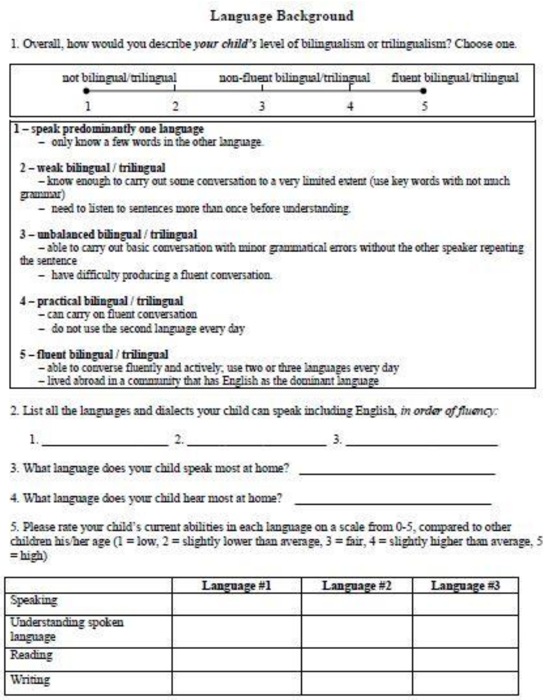


Figure 2.1. Teacher image corresponding to the question of “this bear is a ___ (teacher)” from the modified Expressive Vocabulary Test

For students attending the Mandarin immersion program, the test was first conducted in Chinese with the option of answering in either Mandarin or English. When the child did not understand the question in Mandarin, the experimenter asked again in English. Their responses in Mandarin and/or English were scored similarly with 0, 0.5, or 1 point per question. Perhaps because of the language switch involved, Cronbach's alpha with 34 kindergartners attending Mandarin immersion class was only 0.57 for these children's combined conceptual vocabulary, whereas Cronbach's alpha for their Mandarin responses was 0.95. For comparison purposes in this dissertation (see Study 2), only the conceptual vocabulary score (English responses were accepted) was used.

Parent Questionnaire. All parents filled out a demographics and child language/math background questionnaire. A full questionnaire can be seen in Appendix II. Of particular interest to this dissertation are the questions regarding child language proficiency, which was adapted from the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, &Kaushanskaya, 2007). Parents rated their children's understanding, speaking, reading, and writing in each language compared to other children their age on a scale of 1 (low) to 5 (high). If a child is exposed to more than one language, the parent can fill out the information for the additional language(s). The language background section of the questionnaire is displayed as Figure 2.2.



Language Background

1. Overall, how would you describe *your child's* level of bilingualism or trilingualism? Choose one.

not bilingual/trilingual — non-fluent bilingual/trilingual — fluent bilingual/trilingual

1 2 3 4 5

1 – speak predominantly one language
– only know a few words in the other language.

2 – weak bilingual / trilingual
– know enough to carry out some conversation to a very limited extent (use key words with not much grammar)
– need to listen to sentences more than once before understanding.

3 – unbalanced bilingual / trilingual
– able to carry out basic conversation with minor grammatical errors without the other speaker repeating the sentence
– have difficulty producing a fluent conversation.

4 – practical bilingual / trilingual
– can carry on fluent conversation
– do not use the second language every day

5 – fluent bilingual / trilingual
– able to converse fluently and actively; use two or three languages every day
– lived abroad in a community that has English as the dominant language

2. List all the languages and dialects your child can speak including English, *in order of fluency*:

1. ______________ 2. ______________ 3. ______________

3. What language does your child speak most at home? ______________

4. What language does your child hear most at home? ______________

5. Please rate your child's current abilities in each language on a scale from 0-5, compared to other children his/her age (1 = low, 2 = slightly lower than average, 3 = fair, 4 = slightly higher than average, 5 = high)

	Language #1	Language #2	Language #3
Speaking			
Understanding spoken language			
Reading			
Writing			

Figure 2.2. The language background page of the parent questionnaire

Mathematics

Number Line Task. The number line task is a robust measure of children's numerical representation and development, with older children shifting from a logarithmic representation to a linear one (Booth & Siegler, 2006; Siegler & Opfer, 2003). The task has been digitized so that children can indicate the position of 26 numbers between 0 and 100 on a line displayed on a tablet (the 26 numbers can be found in Appendix III). The numbers are displayed one at a time in a random order; each number appears on the top of the screen while the line with 0 on the left side and 100 on the right stays in the middle of the screen. The child can touch the line to move the marker around. When the final placement of the marker is decided, the child touches the "next" button to move on to the next trial. After all 26 numbers have been displayed, the child sees the “Finished!” screen. I calculated the average percent absolute error for each individual. Percent absolute error (PAE) was calculated using the following formula.

$$PAE = \frac{abs(Number - Marked\ number)}{100}$$

We have tested the digitized number line task with 215 kindergartners from 2016 to 2017, and Cronbach’s alpha for the measure was 0.91.

Common Core Counting Questions. The Common Core State Standards for Mathematics has a standard specific for Number and Operations in Base Ten (NBT; CCSS-M, 2010). For kindergartners, the NBT standard is to be able to work with numbers 11-19 to gain foundations for place value. For first graders, NBT standards include extending the counting sequence to 120 (CCSS-M, 2010). Two sets of questions were derived from the Hawaii Unified School District’s sample Common Core assessment question pools (“Standards Toolkit - Common Core Mathematics Assessments,” n.d.). The first set asked children to count the number of objects in two separate bags. Figure 2.3 shows the sample objects. The second set assessed the children’s rote counting ability to count from 0-25, 28-40 (or 38-50), 87-100, and 100-120. A total score of 14 was possible (3 points each for counting the items in the two bags correctly, and 2 points for each of the rote counting categories). The outcome variables derived from this measure were total correct or percentage correct. Students’ percentage correct is calculated as total points earned divided by 14.

Figure 2.3. Stars counted by the students in the first section of the counting task

Kindergartners in the Mandarin immersion classes received an extended version of the NBT questions. Specifically, after counting objects from the two bags, they were asked to add the numbers together. They also answered number pattern questions, a word problem requesting children to circle group of 10s together, and a set of two base-10 decomposition questions (for full set, see Appendix IV). Additionally, for the rote counting section, the experimenter asked the participants to count first in Mandarin. After all four segments were completed, they counted in English. However, for comparison purposes (see Study 2), only responses to the questions that overlapped with the ones received by the English instruction classes (the first two sections: mystery bags without the addition and rote counting) and the English rote counting responses were analyzed. The overall reliability of the measure from 120 kindergartners was 0.74, with Cronbach's alpha being 0.72 and 0.74 for the 86 kindergartners attending English instruction classes and 34 kindergartners in Mandarin immersion classes, respectively.

Number Identification. Twenty-four single- and double-digit numbers were printed on 4”x4” laminated index cards. With the exception of numbers below 20, two numbers from each decade were selected for this task. The numbers were 1, 5, 8, 10, 11, 13, 15, 18, 20, 26, 31, 37, 41, 44, 53, 59, 62, 64, 75, 78, 83, 86, 92, and 95. The experimenter shuffled the cards before the session and presented the numerals one card at a time to the participant and asked the participant to name the number. Each question was worth 1 point, and the questions were scored as either correct (1) or incorrect (0).

Children in the Mandarin immersion program did one round of number identification in Chinese and another round in English. For comparison purposes (see Study 2), only the English responses were reported. A total of 120 kindergartners completed the task, and Cronbach’s alpha for the task was 0.96. Cronbach’s alpha was 0.96 for the 86 children attending English classes, and 0.97 for the 34 attending Mandarin immersion classes.

Magnitude Comparison. For magnitude comparison, 24 number pairs were shown to the participants on a ringed binder. The experimenter read the numbers and asked the participant to identify the number that was bigger. The number pairs were selected such that the ratio of the smaller number to the bigger number ranged from 0.32 to 0.79. The bigger number is displayed on the left side for half of the questions. The numbers pairs and the ratios are included in the appendix (see Appendix V for the number pairs). Each question was worth 1 point. The reliabilities (Cronbach's alpha) were 0.85 for all kindergartners (n = 116), 0.82 and 0.85 for the children in the English instruction and Mandarin immersion classes, respectively.

Executive Function

Verbal Working Memory: Following Instructions (FI). We assessed kindergartners' verbal working memory by using a hands-on task called "Following Instructions" developed by Gathercole, Durling, Evans, Jeffcock, and Stone (2008). Previous study with the items specified by Gathercole and colleagues (2008) yielded a low test-retest reliability of .09 to .26 (Ramani et al., 2017). Therefore, our modified Following Instructions task used up to five distinct child-friendly items (i.e., box, cup, three small toy counter items such as fish, car, and plane) with three colors (blue, red, and yellow) for each item. To keep the item word length consistent, the Mandarin Chinese version of Following Instructions uses rabbit (兔子, tu zi), bird (小鳥, xiao niao), and train (火車, huo che) as the toy counter items; all words were two characters. Figure 2.4 displays a sample of the setup for the English version.

Figure 2.4. Following Instructions setup

The procedure for administering the task was as follows. The experimenter gave the child a verbal instruction (e.g., “touch the blue fish”), and the child followed suit. The task became increasingly harder with increasing number of actions that the participant would have to keep in memory before acting out the instruction. Each difficulty set contained four questions. A child must perform three out of the four questions correctly in order to get to the next difficulty round. For example, if a child correctly performed three of the two-action sequences (e.g., “pick up the red car and put it in the blue cup”), the experimenter would proceed to the next round (four questions of three-action sequences such as “touch the yellow plane and put the blue car in the blue box”). I used the total questions correct for the analyses in this dissertation.

With each question scored as 0 for incorrect or 1 for correct, Cronbach’s alpha of the modified English version was 0.83. See Appendix VI for sample questions from Form A of the English version.

However, the experimenters encountered unexpected complications when testing FI at the Mandarin immersion program. The test was designed to be administered using the participants’ language of instruction. Unfortunately, many of the children did not understand Mandarin Chinese enough to complete the task in Chinese. The experimenter had to teach some of the words (e.g., train, cup, box) in Chinese before beginning the task. To measure participants’ working memory instead of their language abilities, the experimenter translated the questions to

English on the spot after participants failed to advance to the next round. Therefore, the "English version" of the children from the Mandarin immersion program received was incomplete. Nevertheless, Cronbach's alpha was 0.82 for the Mandarin FI and 0.79 for the English version. The high alpha for the English version could be due to the fact that many of the first action sets (e.g., "touch the yellow cup") were skipped (most children could complete these actions in Chinese) and, therefore, coded as 0. Partially because of the inconsistency across sites in task administration, the analysis in Study 2 excluded the FI measure.

Visuospatial Working Memory: Touch Base. To assess children's spatial working memory, we used a digitized and child-friendly version of the Corsi Block Tapping Test (CBBT; Corsi, 1972); we called the child-friendly version of the game "Touch Base." Digitized CBBT has been shown to work similarly to the traditional version (Brunetti, DelGatto, & Delogu, 2014). In the Touch Base version, on the tablet screen, in place of a regular block was a planet crater. To increase engagement, the experimenter told a story of Alvin the Alien, the secret agent. The children's tasks were to follow the secret agent in the same or reverse order. Children saw the nine craters with a cartoon alien that appeared in one crater and then another. They had to recall the order and the location of the craters in which the alien appeared. In the first part of the task, they had to recall the same order as the appearance of the cartoon alien (the forward span). The top panel on Figure 2.5 displays Touch Base forward version. In the second part, they had to recall the order in reverse (the backwards span). The bottom panel on Figure 2.5 displays Touch Base backward version. The task starts out with two spans; the alien appears in one crater and then in another. There are two questions per set size, and participants proceed onto the next level if they get at least one of the two questions right. The difficulty level increases as children progress through the task in order to gauge their spatial working memory span. The tablet app records participants' progress and responses.

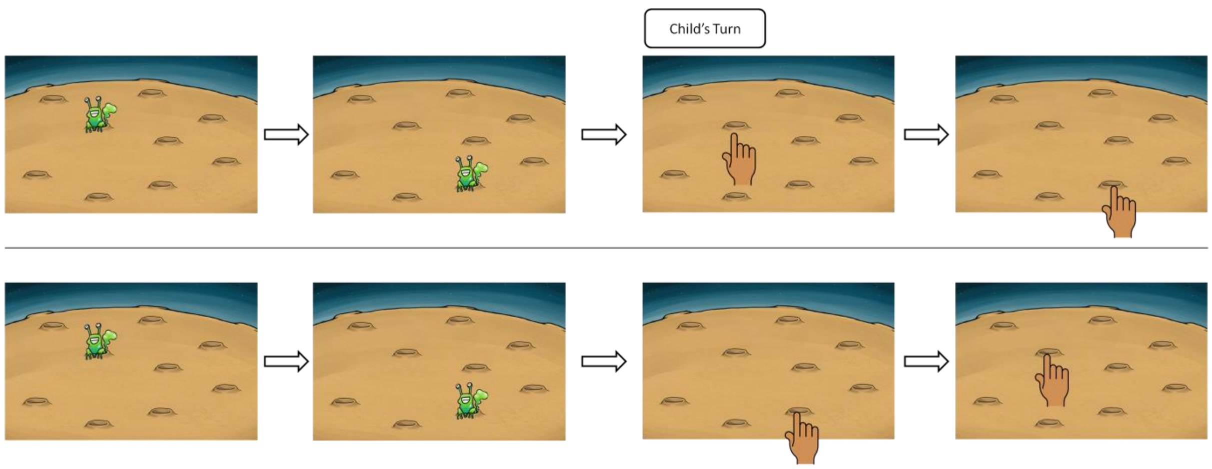


Figure 2.5. A sample visual display of Touch Base forward (top panel) and backward (bottom panel). A trial starts out with Alvin the Alien on one crater and then on another. After the stimuli are presented, the child then gets to touch the crater bases in the same (top panel, forward version) or reverse (bottom panel, backward version) order.

Touch Base has been tested with 187 kindergartners. Cronbach's alpha for the forward version was 0.76. The reliabilities were 0.75 for the 86 children in English instruction classes in California (Study 1 sample and part of Study 2 sample) and 0.62 for the 34 children in the Mandarin immersion classes (includes part of Study 2 sample). Cronbach's alpha for the backward version was 0.70 for all kindergartners tested. For the English instruction classes in California, the reliability was 0.62, and it was 0.60 for the Mandarin immersion sample. As with FI, I calculated the total questions correct for forward and backward versions for the analysis.

Complex Span: Counting Sheep. To assess children's counting span, we modified the original counting span task (Case, Kurland, & Goldberg, 1982) to a child-friendly Counting Sheep task. This task was presented and recorded using E-Prime 2. There were three blocks of trials for this task. Every block contained one trial each of two-, three-, and four-set sizes. For each trial, the child saw a certain set number of "pictures" (ranged from two pictures to four pictures). The pictures displayed different numbers of sheep and wolfs in a green background (see instruction protocol in Appendix VII); the number of sheep ranged from two to five. A sample trial is displayed in Figure 2.6. The wolfs were included as distracters (similar to the role of the red dots or squares in the original counting span). The experimenter instructed the child to count and remember the number of sheep on each picture. After a certain set of pictures, the child was asked to recall the number of sheep they saw in order. We determined the number of sheep per picture and the set size included in the program for this experiment based on pilot testing results. Unlike FI or Touch Base, students received all questions for this measure; therefore, I used the percentage of accurate recall for this measure. Only children in English instruction classes received the Counting Sheep measure, and Cronbach's alpha was 0.87.

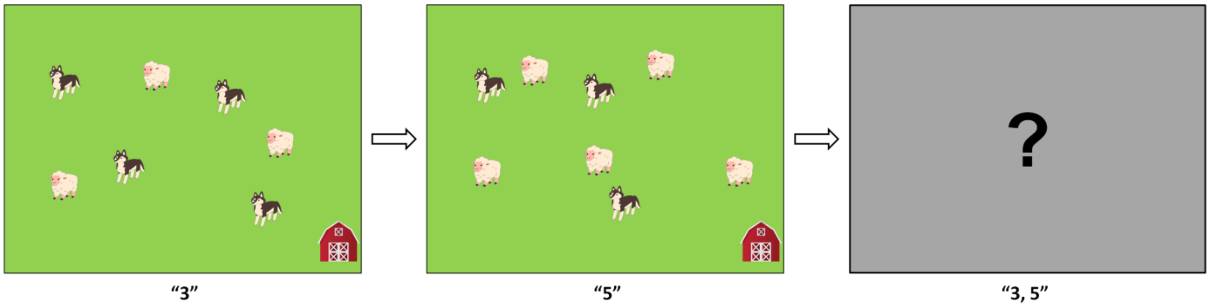


Figure 2.6. A sample Counting Sheep trial. The child counts and gives the number of sheep in the first picture ("3") and repeats for the subsequent display(s) ("5"). When the question mark shows up, the child recalls the number of sheep displayed in the pictures in the same order ("3, 5").

Validity

Concurrent, Convergent, and Discriminant Validity

To examine the concurrent validity of the measures, I explored the correlation among the measures. Additionally, as a first step toward investigating the measures' construct validity, I drew from part of the Multitrait-Multimethod (MTMM) matrix approach (Campbell & Fiske, 1959). MTMM assesses the construct validity by examining the convergent validity and discriminant validity of the measures separately. The mono-trait correlations inform the convergent validity while the hetero-trait correlations inform the discriminant validity. Convergent validity is calculated as the average of all the mono-trait correlations (highlighted in blue in Table 2.1). Discriminant validity is calculated as the mean of the hetero-trait correlations. Typically, MTMM includes multiple methods, each of which would contain at least a task that would assess each trait. Because my 11 measures do not cover multiple "methods" this way, the convergent validity approach used here is based only on the correlations regarding the *traits* of language, mathematics, and executive function.

Table 2.1 displays the correlation matrix for the 11 measures for children attending typical English instruction classes. The mono-trait correlations are highlighted in blue. Because correlations sometimes go in the negative direction (e.g., for number line estimation, lower numbers indicate better performance) and negative correlations may unduly affect the mean, the mean of the absolute values of the correlations were taken for convergent and discriminant validity. The convergent validity from the mono-trait correlations were 0.53. In contrast, the hetero-trait correlations indicate a discriminant validity of 0.32. The higher value for convergent validity indicates that the measures used in this project may indeed meet construct validity criteria. However, this interpretation is subject to criticism common to the MTMM approach, that there is no clear, objective standards for how different the correlation estimates have to be in order to be classified as achieving construct validity (see Kenny & Kashy, 1992).

Closer examination of some individual correlations for children attending the English classes revealed moderately high inter-construct correlations between EVT and measures in math (e.g., number identification) and executive function (e.g., FI). This may be indicative of the way in which the test was conducted. For example, both EVT and number identification required the participant to produce a spoken response to a visual stimuli; the only difference was that for the

EVT, the stimuli were pictures accompanied by a sentence whereas for number identification, the stimuli were Arabic numerals. Likewise, FI required the participants to produce an action after hearing a set of verbal instructions. As the correlations with certain measures such as EVT may be less conclusive, further factor analyses were conducted.

Because of the unplanned procedure/language-switch change at the immersion school, I generated a separate correlation matrix to examine the validity of the measures, particularly the FI measure, at the Mandarin immersion school (see Table 2.2). The construct and discriminant validities were much lower at 0.31 and 0.16, respectively. Most of the measures were not significantly correlated with one another.

Because the correlation matrices provide only rough estimates to the measures' validity, I conducted an exploratory factor analysis using principal component analysis (pca) extraction method with oblique rotation (oblimin) in jamovi 0.8.3.0 (2018) with the measures as further evidence that the tasks can indeed be grouped into the three dimensions of language, math, or executive function. Principal component analysis technique was chosen because, though not aimed at discovering underlying constructs, it takes into account all variances including measurement error (Tabachnick & Fidell, 2007). As Table 2.3 shows, the exploratory analysis with pca resulted in three dimensions (here called factors to be consistent with the idea of factor analysis) that align with the aforementioned theoretical constructs for children attending typical

English instruction classes. (Results from a combined exploratory factor analysis of all participants can be found in Appendix VIII.) However, it is interesting to note that magnitude comparison has high loadings for both constructs related to math and language measures.

Table 2.1

Correlation Matrix for Kindergartners Attending English Instruction Classes

	Language			Math				Executive Function			
	PVT	EVT	PQ	NLE	CCCQ	NI	MC	FI	TB-F	TB-B	CS
Language											
PVT	—										
EVT	0.470 ***	—									
PQ	0.639 ***	0.458 ***	—								
Math											
NLE	-0.111	-0.368 **	-0.286 *	—							
CCCQ	0.319 **	0.464 ***	0.382 **	-0.638 ***	—						
NI	0.254 *	0.526 ***	0.348 **	-0.676 ***	0.699 ***	—					
MC	0.403 ***	0.483 ***	0.366 **	-0.583 ***	0.627 ***	0.757 ***	—				
Executive Function											
FI	0.167	0.503 ***	0.159	-0.212	0.360 ***	0.317 **	0.294 **	—			
TB-F	0.256 *	0.268 *	0.222	-0.300 **	0.349 **	0.422 ***	0.305 **	0.331 **	—		
TB-B	0.400 ***	0.263 *	0.270 *	-0.299 **	0.359 ***	0.328 **	0.419 ***	0.364 ***	0.405 ***	—	
CS	0.220	0.343 **	0.195	-0.300 **	0.353 ***	0.346 **	0.312 **	0.415 ***	0.473 ***	0.456 ***	—

Note. N=86. * p < .05, ** p < .01, *** p < .001. PVT = bilingual Picture Vocabulary Test. EVT = Expressive Vocabulary Test. PQ = Parent Questionnaire. NLE = Number Line Estimation. CCCQ = Common Core Counting Questions. NI = Number Identification. MC = Magnitude Comparison. FI = Following Instructions. TB-F = Touch Base Forward. TB-B = Touch Base Backward. CS = Counting Sheep. Blue boxes indicate monotrait correlations. FI only includes scores from kindergartners attending typical English instruction classes.

Table 2.2

Correlation Matrix for Kindergartners in Mandarin Immersion Classes

	Language			Math				Executive Function		
	PVT	EVT	PQ	NLE	CCCQ	NI	MC	FI	TB-F	TB-B
PVT	—									
EVT	0.122	—								
PQ	0.148	0.073	—							
NLE	-0.120	-0.066	-0.129	—						
CCCQ	-0.030	0.016	0.311	-0.395 *	—					
NI	-0.181	-0.215	0.155	-0.296	0.867 ***	—				
MC	0.113	-0.164	0.164	-0.530 **	0.498 **	0.458 **	—			
FI	0.285	0.299	0.422 *	-0.071	0.257	0.095	0.142	—		
TB-F	0.066	0.173	0.116	-0.260	0.152	0.092	0.203	0.046	—	
TB-B	0.129	0.124	0.238	-0.179	0.218	0.116	-0.139	0.003	0.265	—

Note. N = 34. * p < .05, ** p < .01, *** p < .001. PVT = bilingual Picture Vocabulary Test. EVT = Expressive Vocabulary Test. PQ = Parent Questionnaire. NLE = Number Line Estimation. CCCQ = Common Core Counting Questions. NI = Number Identification. MC = Magnitude Comparison. FI = Following Instructions. TB-F = Touch Base Forward. TB-B = Touch Base Backward. CS = Counting Sheep. Blue boxes indicate monotrait correlations. FI only includes scores from the Mandarin Chinese administration of the task.

Table 2.3

Exploratory factor analysis with all measures for children attending typical English instruction classes: factor loadings (loadings > 0.30 are boldfaced), uniqueness, factor correlations, and explained variance

	Factor 1	Factor 2	Factor 3	Uniqueness
PVT	-0.064	0.098	**0.834**	0.270
EVT	0.122	0.011	**0.714**	0.416
PQ	0.056	-0.067	**0.851**	0.283
NLE	-0.019	**-0.892**	0.201	0.281
CCCQ	0.037	**0.783**	0.071	0.315
NI	0.068	**0.746**	0.201	0.242
MC	0.007	**0.572**	**0.448**	0.282
FI	**0.838**	0.010	0.040	0.268
TB-F	**0.773**	0.100	-0.084	0.370
TB-B	**0.699**	-0.044	0.195	0.419
CS	**0.818**	-0.022	-0.079	0.377
Factor 1	1	.	.	
Factor 2	0.402	1	.	
Factor 3	0.307	0.359	1	
% Variance	23.5	22.5	22.0	

Note. N = 41. PVT = bilingual Picture Vocabulary Test. EVT = Expressive Vocabulary Test. PQ = Parent Questionnaire. NLE = Number Line Estimation. CCCQ = Common Core Counting Questions. NI = Number Identification. MC = Magnitude Comparison. FI = Following Instructions. TB-F = Touch Base Forward. TB-B = Touch Base Backward. CS = Counting Sheep. Extraction method: principal components, rotation: oblimin.

To examine whether the tasks hold the same structure across different groups, I performed EFA with the measures for kindergartners in Mandarin immersion classes as well. However, no clear pattern emerged using various rotation methods of EFA on data from the Mandarin immersion classes. Table 2.4 displays the EFA using principal component analysis with oblique (oblimin) rotation restricted to three factors. As evidenced in this approach, Following Instructions, when administered in Mandarin, as was originally intended, was more aligned with the other language measures of vocabulary tests and parent responses regarding their children's language abilities. Also of interest was how EVT appeared to be split between the language and executive function factors. The three factors also have very low correlation with each other.

Table 2.4

Exploratory factor analysis with all measures in Mandarin immersion classes: factor loadings (loadings > 0.30 are boldfaced), uniqueness, factor correlations, and explained variance

	Factor 1	**Factor 2**	**Factor 3**	**Uniqueness**
PVT	0.106	**0.437**	0.078	0.766
EVT	-0.060	**0.365**	**0.539**	0.545
PQ	-9.13×10^{-4}	**0.742**	0.041	0.442
NLE	**-0.663**	0.190	**-0.308**	0.439
CCCQ	**0.760**	**0.305**	0.071	0.222
NI	**0.818**	0.084	-0.094	0.310
MC	**0.867**	-0.099	-0.087	0.279
FI	0.073	**0.815**	-0.037	0.315
TB-F	0.082	-0.155	**0.774**	0.386
TB-B	-0.114	0.141	**0.688**	0.497
Factor 1	1			
Factor 2	0.183	1		
Factor 3	0.120	0.110	1	
% Variance	25.1	17.7	15.2	

Note. N = 32. PVT = bilingual Picture Vocabulary Test. EVT = Expressive Vocabulary Test. PQ = Parent Questionnaire. NLE = Number Line Estimation. CCCQ = Common Core Counting Questions. NI = Number Identification. MC = Magnitude Comparison. FI = Following Instructions. TB-F = Touch Base Forward. TB-B = Touch Base Backward. The EVT reported here consisted of children's responses in Chinese rather than English. The FI here consisted of children's responses in Chinese. The procedural change meant that children only received the Chinese test with the standardized termination criterion. Extraction method: principal components, rotation: oblimin.

The results show that Following Instructions may not be as valid of a measure for children in the Mandarin immersion program and that the data exhibit a fundamentally different structure pending kindergartners' language of instruction. There could be multiple potential explanations for the different patterns with the Mandarin immersion classes. First, the measures have low to moderate Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, ranging from 0.34 for Touch Base backward to 0.65 for number line estimation and following instructions. With only 32 kindergartners in the immersion program who completed all the measures, the unexpected pattern could possibly be attributed to inadequate sample size. Another potential explanation is that most of the children enrolled in the Mandarin immersion classes were new learners of Chinese. As mentioned earlier, the experimenter had to teach the Chinese names of most objects used in the Following Instructions test. As the factor analysis was conducted with kindergartners' performance on the Chinese FI, it goes to reason that it was their ability to quickly learn the Chinese words that was picked up by the test, rather than their working memory.

Confirmatory Factor Analysis

To further establish the construct validity that the measures fall into the three specified categories of language, mathematics, and executive function, I performed one confirmatory factor analysis (CFA) using only one factor and another CFA using the three factors with the data from the English instruction classes. The one-factor CFA was *not* a good fit for the data, $\chi^2(44) = 99.3$, $p < .001$, CFI = 0.879, RMSEA = 0.102, whereas the three-factor CFA was a good fit, $\chi^2(41) = 46.7$, $p < .250$, CFI = 0.988, RMSEA = 0.034. This indirect comparison provides further indication that the bilingual PVT, expressive vocabulary, and parent questionnaire on their children's language proficiency are valid measures of children's language abilities, that Counting Sheep, FI, and Touch Base Forward and Backward versions are valid as measures of children's executive function, and that number line estimation, Common Core counting questions, number identification, and magnitude comparison are indeed valid measures for assessing mathematics.

Conclusion

Although all 11 measures are adequately reliable for children enrolled in both types of classes, the concurrent and construct validity as demonstrated through correlational and factor analyses tell a different story. All measures appear to be valid for children enrolled in only the English immersion classes. The different patterns could be due to inadequate sampling from the Mandarin immersion classes and/or the unexpected administration change of the Following Instructions task resulting from new learners' lack of proficiency in Mandarin Chinese. Regardless of the reasoning, the emergence of the different EFA pattern calls into question the construct validity of the measures for children enrolled in the Mandarin immersion program. Future studies aimed at assessing children's cognitive abilities should carefully consider both the language of instruction and children's dominant language. As seen with the alignment of the verbal working memory measure with the language measure, perhaps domain-general skills should be assessed with children's dominant language, rather than their language of instruction.

With the discovery of different patterns, adjustments are therefore made to the analyses presented in the following chapters. Analyses for Study 1 (next chapter), with its focus on the latent constructs and their relationship with each other, are restricted to only children attending typical English instruction classes. Analyses in Study 2 that focuses on the math performance of children learning Mandarin Chinese, a language with a transparent base-10 system, exclude Following Instructions as a measure and examine the math tasks separately. The overlap that appeared between children's language ability, as measured by EVT, with their executive function, as measured by Touch Base (see Table 2.4), also justified the inclusion of the two remaining executive function measures as predictors in Study 2 analyses. Without further ado, the next chapter explores the effect of language and executive function on mathematics among kindergartners attending typical English instruction classes.

CHAPTER 3

Study 1. How do linguistic, executive function, and mathematics skills relate to each other in emergent bilingual kindergartners? A SEM analysis

Numbers and mathematics are essential in everyday tasks from telling time to measuring the amount of ingredients needed for cooking. The crucial function of mathematics in modern society is recognized and shared by various coalitions and government entities. Indeed, because of the link between mathematics achievement and economic success, the National Mathematics Advisory Panel was created as an Executive Order to "foster greater knowledge of and improved performance in mathematics among American students" (*Executive Order No. 13398*, 2006).

In addition to the intrinsic value of mathematics itself, there is growing interest in the research community on how different domains of cognition (e.g., language and math) are connected. Though mathematics and language arts have often been thought of as distinct disciplines with separate objectives and class times, recent research and insights into cognitive development show that language, mathematics, and other more "general" domains such as executive function or working memory are more connected than previously thought. The aim of this paper is to investigate the potential concurrent links between the constructs of language, mathematics, and domain-general executive function among kindergartners from diverse backgrounds.

Domain-Specific Skills for Mathematics

Cognitive research has established a specific domain or module meant for numerical processing. In the influential triple-code model (Dehaene, 1992, 2010; Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2003), this math-specific construct is known as the "quantity code." The quantity code is responsible for intuitive number sense, which includes the approximate number system (ANS) and magnitude comparison as well as higher level manipulations such as subtraction. These properties or numerical skills have been grouped under the "quantity code" partly because of overlapping brain regions involved in these tasks. Specifically, neuroimaging studies have implicated the posterior superior parietal lobule (PSPL) and the horizontal segment of the bilateral intraparietal sulcus (hIPS) to be responsible for quantity representation and manipulation, regardless of the form or notation of the numbers (for a review, see Dehaene, Molko, Cohen, & Wilson, 2004).

Indeed, the quantity code, with its inclusion of the ANS, may indicate an evolutionarily older system. Ample research studies have shown that humans and other species have the capacity to do numerical estimation without the use of language (e.g., Rugani, Vallortigara, & Regolin, 2015; Starkey, Spelke, & Gelman, 1990; Xu & Spelke, 2000; for reviews, see Butterworth, 2010; Cantlon, Platt, & Brannon, 2008). Such studies usually take on a non-symbolic format, e.g. with arrays of dots to represent distinct numerosities, and examine participants' or animal subjects' abilities to compare magnitudes or *approximate* the sum or difference of the operands (e.g., Chinello, Cattani, Bonfliglioli, Dehaene, & Piazza, 2013; Knops, Dehaene, Berteletti, & Zorzi, 2014; Knops, Viarouge, & Dehaene, 2009).

Nevertheless, educationally speaking, the development of mathematical skills goes beyond the evolutionarily based ANS. In particular, symbolic numbers are a staple in mathematical tasks. Individuals acquire the meanings of these symbols by mapping them onto the ANS representation of quantity (Piazza, 2010). Because acquiring number words is a basis of math achievement and because symbolic number processing (e.g., magnitude comparison with numerals) is strongly predictive of later math achievement in schools (De Smedt, Noel, Gilmore,

& Ansari, 2013; Holloway & Ansari, 2009), in this paper, I focus on symbolic number tasks when assessing mathematical skills of kindergarten children.

Domain-General Skills for Mathematics

In addition to domain-specific numerical skills, children need domain-general skills in order to successfully acquire mathematics skill (Geary & Hoard, 2005). These "domain-general" components refer to executive function and control skills needed while processing complex tasks.[3] The cognitive control processes regulate and control individuals' attention, serving to select relevant information and suppress irrelevant ones (Engle, 2002, 2010).Sometimes called executive function (EF), this general domain may include skills such as shifting, inhibiting, and updating (Miyake & Friedman, 2012). Shifting (or switching) and inhibition are typically considered executive control processes. Shifting or switching involves being able to allocate and switch attention to the task at hand, whereas inhibiting is the suppression of potentially dominant responses in order to perform the required task. A classic test used to measure executive control such as inhibition is the Simon task (Simon & Wolf, 1963). Participants are shown stimuli on either side of the screen and have to press the corresponding key. Some trials are congruent, where the stimulus appears on the same side as the key, while others are incongruent. Better inhibitory control would mean, for example, faster reaction time in incongruent trials, implying that participants can better able to suppress the response of pressing the key on the same side as the stimulus appears on the screen. On the other hand, the "updating" component of EF is most similar to the idea of working memory (WM), where individuals store and manipulate multiple

[3] The domain-general cognitive control skills have been termed differently by researchers following different traditions and lines of research: executive function (Miyake &Friedman, 2012), central executive (Baddeley, 1992, 2000), working memory (Baddeley, 1992; Engle, 2002), or even executive attention (Engle, 2002, 2010). Because I refer to the most general cognitive control processes, in this paper, I use these terms interchangeably.

concepts in mind. Updating or WM has also been further broken down into smaller systems, such as visuospatial WM (Baddeley, 1992).

Regardless of the specific component or terminology, in numerical processing among children, all types of EF, be it attentional control, inhibition, or working memory, are necessary. However, for the purpose of this study, I focus on the WM component of EF for several reasons. First, take children's math tasks for instance. In counting and adding two groups of star-shaped manipulatives, for instance, a child must suppress or deselect any attention that might be drawn to the irrelevant dimension of "star shapes" ("oh look! It's stars!") and attend to counting the first group, *maintain that number in mind*, and continue counting forward with the second group. Even in simple rote counting, a child must keep the order of the numbers in mind and carefully monitor which number has already been said in order to move on to the next number. In these examples, the type of attention suppression needed does not necessary match the type of inhibitory control measured by the tasks typically used for assessing the executive control aspects of EF (e.g., the Simon task described earlier). The classical tests may be too narrowly focused, whereas in real life situations, multiple factors (e.g., children's motivation) may affect children's attention. Second, research findings have been inconsistent regarding the executive control processes among diverse language groups (e.g., Paap et al., 2014). Some have found superior performance among bilingual groups (see Kroll & Bialystok, 2013), while others have found monolingual advantage on some tasks (e.g., Paap & Greenberg, 2013). Interestingly, at least one research study has suggested that it is actually working memory performance that could predict executive control processes, rather than language group (Namazi & Thordardottir, 2010). Third, as described in the next section, the strongest connection between math and EF appeared in the working memory components of executive function. It is therefore more essential that this study focuses on the WM aspect of EF.

Executive Function and Math. The body of research on executive function (EF) has shown a clear connection between EF, especially the working memory (WM) components, and mathematics (e.g., Bull, Espy, & Wiebe, 2008; Jarvis & Gathercole, 2003; Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013; Krajewski & Schneider, 2009; Simmons, Singleton, & Horne, 2008; Toll, Kroesbergen, & vanLuit, 2016; for reviews, see Bull & Lee, 2014; Cragg & Gilmore, 2014; Jacob & Partkinson, 2015). For example, in a longitudinal study tracking children from kindergarten to the beginning of third grade, researchers found that both verbal and visuospatial working memory affect children's early and later quantity-number competencies, which in turn contributes to their math achievement (Krajewski & Schneider, 2009). Another longitudinal study also finds visuospatial WM, on top of number sense, to be predictive of later math performance (Toll et al., 2016). Similarly, when various WM tasks are grouped under verbal WM or nonverbal WM using confirmatory factor analysis, both types of working memory strongly predict children's academic attainment, including their scores on mathematics tests, equally well (Jarvis & Gathercole, 2003).

Researchers who conducted their studies using a battery of executive function, WM, and mathematic measures also highlighted the importance of EF and WM on math (e.g., Alloway et al., 2005; Alloway & Passolunghi, 2011; Clark, Sheffield, Wiebe, & Espy, 2013; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Jarvis & Gathercole, 2003; Passolunghi, Cargnelutti, & Pastore, 2014). For example, EF as assessed with dual-task tests of listening recall, counting recall, and backward digit recall is key in explaining deficits in children with math learning difficulties (Geary et al., 2007). Structural equation modeling also shows that executive function as measured by three separate working memory tasks and three inhibitory control tasks at preschool is directly and indirectly associated with general math proficiency in kindergarten (Clark et al., 2013). Additionally, a meta-analysis points to verbal updating to be the strongest EF predictor of mathematic performance (Friso-van den Bos, van derVen, Kroesbergen, & vanLuit, 2013). Furthermore, direct training of these executive function skills has also yielded positive effects on children's number line estimation and number identification performance (Ramani, Jaeggi, Daubert, & Buschkuehl, 2017).

Studies focusing on various components of EF have shown differential EF skills to be predictive of different types of numerical skills (e.g., Bull et al., 2008; Kolkman et al., 2013; Van derVen, Kroesbergen, Boom, & Leseman, 2012; Verdine, Irwin, Golinkoff, & Hirsh-Pasek, 2014). For example, in kindergarteners, updating, planning, and inhibition are predictive of early mathematical competence (Kroesbergen, vanLuit, vanLieshout, Loosbroek, & Van deRijt, 2009). In middle school students, inhibition, shifting, and updating are crucial in different parts of the mathematical problem solving process (Kotsopoulos & Lee, 2012). These differential findings with the subcomponents largely depend on the tasks the researchers used. However, all the subcomponents belong neatly to the two main constructs of executive function and mathematics, providing further evidence of the necessity of working memory and executive function for performing and learning mathematical tasks.

Despite the clear connection between executive function and mathematics, it is less clear, nevertheless, how the association between the two constructs will hold up with the addition of a confounding construct—language abilities—that is associated with both executive function and mathematics separately.

Language and Math

In addition to domain general executive function skills, skills in the linguistic domain is also influential and indispensable in numerical processing and development (Dehaene, 2010, 2011; Dehaene, Molko, Cohen, & Wilson, 2004; LeFevre et al., 2010).

Indeed, linguistics skills, such as phonological awareness and combinations of grammar, vocabulary, and listening comprehension, are strongly predictive of elementary school students' math performance (Fuchs et al., 2005; LeFevre et al., 2010; Simmons & Singleton, 2008; Vukovic & Lesaux, 2013). More recently, researchers have also examined even more complicated math problems (e.g. word problems), and the findings similarly indicate a strong linguistic involvement (Wang, Fuchs, & Fuchs, 2016). These findings might not be surprising, as the linguistic complexity of mathematical problems have long been shown to negatively impact language learners (Abedi & Herman, 2010; Abedi & Lord, 2001). Together, these studies clearly establish a link between linguistic and math domains; yet, it is still inconclusive whether the relationship between verbal skills and math performance is direct or indirect or which type of verbal skills and math domains are related.

Though linguistic skills are associated with numerous types of math performance in elementary school students, the association appears to have an even broader influence for preschoolers and kindergartners. Testing close to 200 preschoolers and kindergartners from low and middle socioeconomic home backgrounds, Purpura and Ganley (2014) found that children's linguistic skills significantly predict their performances in counting, cardinality, number comparison, set comparison, number order, number identification, set to numerals, and story problems after controlling for demographic variables, such as age and sex, and other cognitive skills.

Language and Executive Function

It is evident that as a domain-general skillset, EF wields substantial influence over linguistic skills (e.g., Alloway, Gathercole, & Pickering, 2006; Alloway & Ledwon, 2014; Baddeley, Hitch, & Allen, 2009; Baddeley, Gathercole, & Papagno, 1998; Goo, 2010, 2012; Just & Carpenter, 1992). Indeed, the capacity theory of working memory, supported by various experiments that manipulated syntactic ambiguity, sentence complexity, extrinsic load, etc., posits that an individual’s working memory capacity can constrain one’s reading comprehension (Just & Carpenter, 1992). Furthermore, executive function mechanisms can affect something as specific as performance on sentence recall tasks (Alloway & Ledwon, 2014) as well as different forms of language learning, whether it is learning new word forms (Baddeley et al., 1998) or reacting to different methods of learning a second language (Goo, 2010, 2012). It comes as no surprise that EF has such effect on language skills as many linguistic tasks involve EF or working memory. For example, learning a new word—whether in a foreign language or not—may require holding a grammatical structure of a sentence in mind while applying the new word’s form or spelling. Executive cognitive control is clearly implicated in language learning and skills, particularly in adults. However, less research has investigated how linguistic skills can influence domain general skills, especially in children when both skill sets are developing concurrently. At least one study suggests that the connection between cognitive control and language may not be as straightforward and unidirectional (Clark et al., 2013). In particular, instead of EF skills predicting language skills, using structural equation modeling (SEM), researchers have found children’s verbal comprehension to be predictive of their concurrent executive control (inhibitory and working memory) skills. However, more studies are clearly needed.

How are these three cognitive domains related?

While linguistic abilities may contribute to early math development, the connection between executive function and language skills could also affect math.

Though the individual connections between two of the three constructs of language, executive function, and mathematics appear to be clear, few studies have examined all potential links together (see Figure 3.1 for the links). In studies that examine all three constructs, the focus is typically on how language affects specific aspects of math (A1 in Figure 3.1) and how working memory affects specific aspects of math (C1) (e.g., Clark et al., 2013; Purpura & Ganley, 2014; Toll & vanLuit, 2014; Vukovic & Lesaux, 2013b).

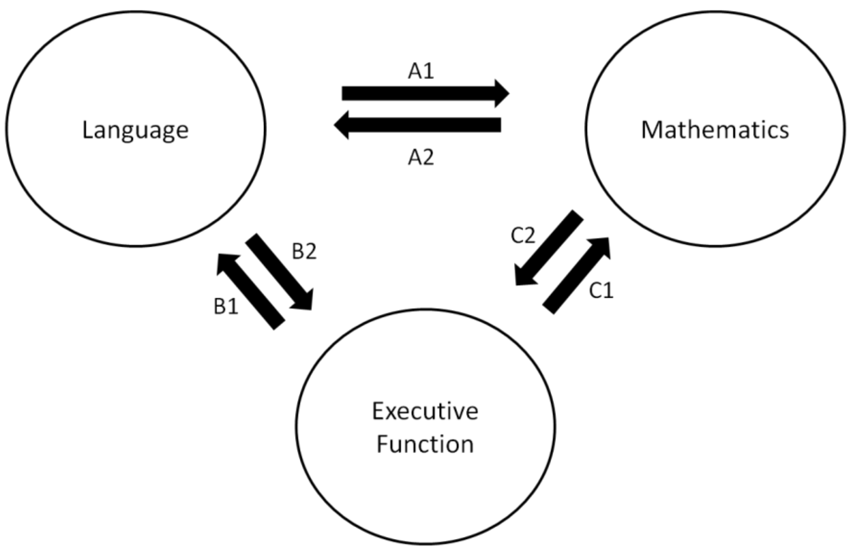


Figure 3.1. Potential links between language, executive function, and mathematics

For example, Vukovic and Lesaux (2013b) examined how language skills in grade 1 or 2 affect performance on algebra, geometry, probability/data analysis, and arithmetic in grade 4; effect was only found for geometry and probability/data analysis. They created a language composite score from picture vocabulary and oral comprehension tests, but only one measure of visuospatial working memory was included as a covariate. Likewise, as the focus was on identifying the relation of language and working memory on different domains of math, Purpura and Ganley (2014) used only one language measure (Expressive One-Word Vocabulary Test) and one word recall verbal memory task. In contrast, in looking for the association between early executive control and later math performance, Clark and colleagues (2013) used the Verbal Comprehension subtest of Woodcock-Johnson III (Woodcock et al., 2001) as a control. Though the test is mainly vocabulary based, it includes four components: picture vocabulary, synonyms, antonyms, and verbal analogies. These studies offer insights into how language or executive function constructs relate to specific types of math, but they tended to use only one or two measures of language or executive function and focus on the individual relation between math and executive function or math and language; how the language-executive function relationship may in turn affect mathematics were often neglected. Therefore, a main goal of the proposed studies is to examine the three constructs together and investigate how they are related and connected.

Furthermore, in studies that do include language effects, most do not examine participants' language or home background or specify whether participants are monolingual, native English speakers or emergent bilingual speakers from a language minority background. The language background of participants may be of particular importance because previous research has established different patterns of language learning and cognitive effects in bilingual individuals (Bialystok, 2007). For example, neuroimaging studies examining white matter structural differences between monolingual individuals and bilingual individuals—even adults who were learning a second language—have suggested that learning a second language is associated with changes in brain regions implicated in language processing (Luk et al., 2010; Kuhl et al., 2016). How the potential changes in neural structures may affect actual behavior needs further exploration. A number of behavioral studies have also suggested a possibility of a bilingual cognitive advantage (e.g., Barac & Bialystok, 2012; Engel de Abreu et al., 2012; Esposito et al., 2013), though this “advantage” is not consistently found (e.g., Paap & Sawi, 2014). The possibility that the language background of participants may affect math findings is further complicated by the consistent findings that bilingual individuals tend to underperform on measures of vocabulary or verbal working memory compared with their monolingual peers (e.g., Bialystok & Luk, 2012; Fernandes et al. 2007). With a potential, albeit controversial, bilingual executive function “advantage” and a disadvantage on verbal measures, and the findings that both executive function and linguistic proficiency separately predict mathematics skills, how the language background can in turn affect math development is, as aptly pointed out by Vukovic and Lesaux (2013), understudied.

Current Study

The present study seeks to understand how language, working memory, and mathematics relate to each other in emergent bilingual children. Specifically, using multiple measures of the three constructs, how do the three constructs relate to each other when considered together? As the focus is on the interrelationship of the latent constructs, these constructs are defined on the individual level by the participants’ exhibited skills. For language, the individual-level focus warrants an examination of the language proficiency of children from linguistically diverse backgrounds. Of particular interest is children’s proficiency in English, the language of instruction, as previous research with emergent bilingual students have demonstrated not just cross-language consistency in numerical tasks such as number identification in both of their languages, but also better performance in counting in the language of instruction, i.e., English (Sarnecka, Negen, & Goldman, 2018).

Our first hypothesis considers the theory that domain-general skills drive skills in both specific cognitive domains (H1). The alternative hypothesis (H2) is more in line with the path of language driving mathematics development regardless of domain-general skills (e.g., LeFevre et al., 2010, Purpura & Ganley, 2014):

H1: domain general skills are associated with both language and math skills, but language and math are minimally related to each other when all three are considered together. In other words, the relationship between language and math is mediated by EF.

H2: language skill is connected to both domain general EF skills and math skills, but EF skills and math skills are minimally related to each other when all three are considered. In other words, EF's effect on math is mediated through its effect on language.

Considering that executive function, language, and mathematics skills can be argued to represent underlying general mental ability (e.g., Schmidt, 2017), yet another alternative hypothesis (H3) is that general mental ability (GMA) would predict each of the three constructs independently. Thus, this alternative hypothesis deviates from the previous two by including one more latent construct, and there are two potential sub-hypotheses related with H3:

H3a: EF, language, and math skills are related to each other (similar to predictions in H1 and H2) even when accounting for GMA.

H3b: No relations remain among EF, language, and math skills after GMA is taken into consideration.

Method

Participants

Eighty-six kindergartners (M_{age} = 73.12 months, *SD* = 7.70) from southern California participated in the study. All children are from schools where the majority of students (over 80%)

are eligible for free and reduced-price lunch. Considering children from lower socioeconomic status (SES) is of particular importance for a couple of reasons. First, children from lower SES backgrounds tend to start kindergarten with lower vocabulary (e.g., Hart &Risley, 1995) and less numerical knowledge than their peers from higher-income families and have a relatively flat growth rate throughout kindergarten (Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan & Levine, 2009). Second, there is a disproportionate number linguistically diverse families, e.g., Hispanic families, living in poverty (e.g., Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Lopez & Velasco, 2011; Flores, Lopez, & Radford, 2017). Yet, the few studies that examined all three constructs together (e.g., Clark et al., 2013; Purpura & Logan, 2014) either excluded children from non-English speaking households or sampled mostly children from typical middle-SES families. It is therefore imperative to see whether previously discovered trends also apply to children from linguistically diverse, lower-SES families.

Of the 86 students, 38 are Caucasian, 31 identify as Hispanic or Latino, 9 as biracial or mixed race, 4 as Asian American or Pacific Islander, and 4 are others or unreported. Furthermore, 31 are monolingual English speakers while 55 are emergent bilinguals with Spanish being the most prevalent household language. Maternal education of these children ranges from some high school courses to postgraduate degree. A breakdown of the children's languages and demographics can be found in Table 3.1.

Not all children completed all of the measures due to absences or fatigue during test sessions. Additionally, 20 parents did not complete the entire parent questionnaire.

Table 3.1

Descriptive Statistics of Demographic Variables

	Total Sample *N* = 86
Age in months, *M*(*SD*)	73.12 (7.70)
Gender (%)	
Female	53
Male	47
Race & Ethnicity of Children (%)	
Caucasian/White	44.19
Hispanic or Latino	36.05
Biracial/Mixed Race	10.47
Asian American or Pacific Islander	4.65
Other or unreported	4.65
Maternal Education (%)	
Some high school	18.6
High school diploma/GED	29.07
Some college/vocational training	25.58
2-year College Degree (Associates)	8.14
4-year College Degree (BA/BS)	8.14
Postgraduate or Professional degree	5.81
Unreported	4.65
Household Income (%)	
Less than $15,000	20
$15,000 - $30,000	23.53
$31,000 - $45,000	11.76
$46,000 - $59,000	10.59
$60,000 - $75,000	10.59
$76,000 - $100,000	7.06
$101,000 or greater	8.24
Unreported	8.24
Household Languages (%)	
Monolingual English speakers	36
Exposed to more than one language	
Spanish	57
Other	7

Note. For Race/Ethnicity, “Other” includes 1 African American, 1 Turkish, 1 American Indian/Alaskan Native, and 1 unreported. Household languages take into account both the languages child speaks at home and languages child hears at home. “Other” includes 2 Tagalog, 1 Turkish, 1 Albanian, 1 Japanese, and 1 unreported

Procedure

Parents filled out and returned informed consent forms as well as a home background questionnaire prior to test administration with their children. Only children whose parents granted consent were tested. Testing was administered in the kindergarten classrooms by trained experimenters. Before testing, the experimenters obtained each child's verbal assent. To make the tests more child-friendly, we referred to them as "games" to the kindergartners. Because of the number of measures, testing was broken down to three days. On the first day, children played the counting game, Following Instructions, and Counting Sheep. Day 2 consisted of the number line task, Touch Base, magnitude comparison, and number identification. On Day 3, we administered the vocabulary measures.

Tasks and Stimuli

Kindergartners completed the 11 measures detailed in Chapter 2. A brief description of the measures is provided below.

Bilingual Picture Vocabulary Test (PVT). To estimate children's language ability, we administered a digitized bilingual Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin, 2013). The established reliability coefficients among 5- and 6-year-olds are 0.97 and 0.94, respectively (N. A.Martin, 2013). The reliability in our sample is 0.96.

Expressive Vocabulary Test (EVT). This expressive vocabulary test was adapted and translated from a previous Mandarin test (Lin & Johnson, 2016) were translated into English. See Appendix I for the stimuli. Cronbach's alpha for this test was 0.85. There are a total of 40 questions on the EVT, and the score reported in this paper is the raw points accumulated for all 40 questions.

Number Line Task. In this digitized task, children indicate the position of 26 numbers between 0 and 100 on a line displayed on a tablet (the 26 numbers can be found in Appendix III). I calculated the average percent absolute error for each individual. Cronbach's alpha for the measure is 0.90.

Common Core Counting Questions. Drawn from the Hawaii Unified School District's sample Common Core assessment question pools ("Standards Toolkit - Common Core Mathematics Assessments," n.d.), there are two sets of counting questions. The first set asks children to count the number of objects in two separate bags. The second set assesses the children's rote counting ability to count from 0-25, 28-40 (or 38-50), 87-100, and 100-120. A total score of 14 is possible (3 points each for counting the items in the two bags correctly, and 2 points for each of the rote counting categories). Students' percentage correct is calculated as total points earned divided by 14. The reliability of this measure is 0.72.

Number Identification and Magnitude Comparison. The experimenter presented 24 single- and double-digit numbers to the participant one at a time and asked the kindergartner to name the number. For magnitude comparison, 24 number pairs were shown to the participant on a ringed binder. The participant is asked to identify the number that is bigger. The percentage correct is calculated by dividing total points by 24. The reliabilities (Cronbach's alpha) are 0.82 and 0.96 for magnitude comparison and number identification, respectively.

Verbal Working Memory: Following Instructions (FI). We assessed kindergartners' verbal working memory by using a hands-on task called "Following Instructions" developed by Gathercole, Durling, Evans, Jeffcock, and Stone (2008). I used the total questions correct for the rest of the analysis. The reliability of our modified version was 0.83 (for the dichotomous scoring method) in our test sample. See Appendix VI for sample questions from Form A of the English version.

Visuospatial Working Memory: Touch Base. To assess children's spatial working memory, we use a digitized and child-friendly version of the Corsi Block Tapping Test (CBBT; Corsi, 1972) called "Touch Base." The reliability in our sample was 0.75 and 0.79 for the forward and backward versions, respectively. As with FI, I calculated the total questions correct for forward and backward versions for the analysis.

Complex Span: Counting Sheep. To assess children's counting span, we modified the original counting span task (Case et al., 1982) to a child-friendly computerized Counting Sheep task. Unlike FI or Touch Base, students received all questions for this measure; therefore, I used the percentage of accurate recall for this measure. The reliability of this measure is 0.87.

Analytical Approach

The descriptive analyses and correlation were performed using Stata15. Because of the number of measures, I first ran exploratory and confirmatory factor analyses using jamovi (jamovi project, 2018) to establish if the tasks were indeed measuring their corresponding constructs (see Chapter 2 for details). I then used structural equation modeling (SEM) to determine potential mediation. In the SEM analyses, latent variables for language, executive function, and mathematics were created with the various tasks as indicators. The SEM analyses were performed using Stata14 and further checked using Ωnyx (von Oertzen, Brandmaier, & Tsang, 2015).

Results

Descriptive Statistics

Language. The kindergartners received a mean raw score of 63.78 (SD = 14.50) on the bilingual PVT exam. For children ages 5 to 6, this average score corresponds to a standard score of 97 (for the highest bracket of ages 6 years and 9 months to 6 years and 11 months) to 111 (ages 5.0 to 5 years and 2 months) (Martin, 2013). To be consistent with the remaining measures, the raw score for subsequent analyses. On the EVT, children received an average of 25.45 (SD = 4.62) points out of 40 questions. The distributions for both vocabulary measures are fairly normal, though the receptive measure is slightly skewed to the left. In contrast, parents tended to score their children higher on the questionnaire regarding their children's English proficiency. The average ratings for speaking English and understanding English are 4.23 (SD = 1.03) and 4.23 (SD = 0.99), respectively. (The highest possible rating is 5.) For reading and writing, they averaged 3.35 (SD = 1.41) and 3.29 (SD = 1.36), respectively. See Table 3.2 for further breakdown and for the overall score.

Executive Function. On the forward version of Touch Base, the children scored a mean of 3.72 points (SD = 2.02). Their backward performance was slightly lower at a mean of 1.81 (SD = 1.49). With nine questions in total with Counting Sheep, the kindergartners answered an average of 14.12% correct (SD = 19.35). However, the range varied widely from 0% to 100%. Consequently, this was the only measure that was heavily skewed. As for the verbal working memory task of Following Instructions, children answered an average of 8.86 (SD = 2.85) correctly. See Table 3.2 for the breakdown.

Mathematics. In contrast to some of the measures for counting sheep that more normal or skewed to the right, the measures for math tended to be skewed to the left (ceiling effect). For example, for magnitude comparison, children scored an average of 79.26% (SD = 21.82) correct, with median being 85.42%. Likewise, children performed well on number identification (M = 71.66%, SD = 31.00, median = 89.58%) and counting (M = 68.94%, SD = 20.17, median = 71.43%). For number line, mean percentage absolute error was around 21.86% (SD = 8.92). See Table 3.2 for a breakdown. To give a context of the PAE performance, if one were to completely guess and move to the middle of the line for every question, the PAE would be around 25%.

Table 3.2

Descriptive Statistics of Outcome Measures

Measures			N	M	SD	min	p25	p50	p75	max	skewness	kurtosis
Language												
	PVT		67	63.78	14.50	19	57	64	72	94	-1.05	5.52
	EVT		74	25.45	4.62	12.5	22	26.5	29	34.5	-0.51	2.93
	Parent Q	Overall (combined)	66	14.47	4.64	2	10	15	20	20	-0.44	2.39
		Speaking	60	4.23	1.03	2	4	5	5	5		
		Understanding	63	4.23	0.99	2	4	5	5	5		
		Reading	65	3.35	1.41	0	2	3	5	5		
		Writing	66	3.29	1.36	1	2	3	5	5		
Executive Function												
	TB Forward		86	3.72	2.02	0	3	4	5	8	-0.21	2.43
	TB Backward		86	1.81	1.49	0	0	2	3	6	0.45	2.50
	Counting Sheep (% correct)		85	14.12	19.35	0	0	11.11	22.22	100	2.04	7.80
	Following Instructions		86	8.86	2.85	0	7	8.5	10	17	0.06	4.89
Mathematics												
	Number Line (PAE)		86	21.86	8.92	4.11	17.1	21.16	25.88	52.11	0.77	4.48
	Magnitude Comparison (%)		84	81.15	18.24	12.50	66.67	87.50	95.83	100	-0.97	3.74
	Number Identification (%)		85	72.50	30.18	4.17	41.67	91.67	100	100	-0.70	1.94
	Counting (%)		86	68.94	20.17	14.29	57.14	71.43	85.71	100	-0.40	2.60

Note. PVT = Picture Vocabulary Test. EVT = Expressive Vocabulary Test. Parent Q = Parents' ratings of their children's English proficiency on the Parent Questionnaire (each subsection is on a scale of 1 to 5, 5 being the most proficient). The overall combined score is the final outcome variable from the Parent Questionnaire. TB = Touch Base, the child-friendly version of the Corsi Block Tapping Task. For Counting Sheep, percentage correct is calculated out of the 9 questions completed. Points correct are reported for PVT, EVT, TB, and Following Instruction tasks.

Correlations

Our variables of interest correlate well with each other within their respective constructs. For example, the language measures of PVT, EVT, and English proficiency based on parent questionnaires are moderately to strongly correlated (*r* ranges from .46 to .64; see Table 2.1). The executive function measures are moderately correlated (r ranges from .33 to .47). The math measures of number line estimation, magnitude comparison, number identification, and counting are strongly correlated with each other (the absolute ranges are from .58 to .76). Interesting, the Expressive Vocabulary measure is significantly correlated with all other measures, but neither PVT nor the English proficiency rating is correlated with Counting Span or Following Instructions. Perhaps the high correlation between EVT and Following Instructions has to do with how both require an active motor planning (one to plan out spoken words and the other a set of actions). All math measures correlate with most other measures. The only exceptions are number line with PVT and Following Instructions. The lack of relation between number line and the vocabulary and verbal working memory measure could be that estimating the position of numbers on a number line entails minimal requirements of language.

Exploratory Factor Analysis (EFA)

The Kaiser-Meyer-Olkin test of sampling adequacy was 0.83, higher than the acceptable index of 0.6. Additionally, the Bartlett's Test of Sphericity was significant ($\chi^2(55) = 178.31$, $p < .001$), verifying the suitability of the data for factor analysis. Principal components method with oblique rotation was then used for the EFA, and the analysis yielded three distinct factors matching the three hypothesized constructs of language, executive function, and mathematics. This structure is further examined with confirmatory factor analysis. As discussed in Chapter 2, the factor analyses confirm the 11 measures with their corresponding constructs. Table 2.3 displays the factor loadings, their correlations, uniqueness, and variance explained.

Structural Equation Modeling (SEM)

After the factor analyses demonstrated the measures fitting into their corresponding categories, I performed the Structural Equation Modeling (SEM) analysis using Stata14 and Ωnyx Version 1.0-972.

I conducted the SEM analysis with the total sample of 86 participants. The total sample of 86 is at the low end of acceptable sample size for the number of parameters in the model. However, if all missing data were excluded, the sample size would reduce to 41. Nevertheless, as a part of a robustness check, the analysis was performed with only the 41 using maximum likelihood estimation method as well as the maximum likelihood with missing values (mlmv) estimation method on the 86 participants using Stata14. Both returned similar results. SEM analyses performed on Ωnyx returned similar path coefficients and model fit results. For the sake of brevity and to keep the analysis with the larger sample size, only the results from the mlmv method from Stata14 are reported here.

SEM is a combination of confirmatory factor analysis (CFA) and multiple regression. As such, the loadings of the observed variables onto the latent constructs of executive function, language, and math differ slightly from that from EFA (see Table 3.3). Most loadings of the indicators to their latent factors were greater than the criterion of 0.60. The lowest loading of 0.57 for Following Instruction points onto the Executive Function latent factor still approached the criterion. To avoid the confusion of potential causal interpretation, I discuss only the direct, indirect, and total effects of the latent constructs for testing Hypotheses 1 and 2 (Schreiber, Nora, Stage, Barlow, & King, 2006).

Table 3.3
Standardized and Unstandardized Coefficients of Measurement Loadings in Structural Equation Modeling

Observed Variable	Latent construct	β(SE)	B
PVT	Language	0.75 (.09)	1
EVT	Language	0.73 (.09)	.27 (.06)
English Proficiency	Language	0.71 (.10)	.27 (.06)
TB Forward	EF	0.65 (.08)	1
TB Backward	EF	0.65 (.08)	.75 (.17)
Following Instructions	EF	0.57 (.09)	1.25 (.31)
Counting Sheep	EF	0.68 (.08)	.10 (.02)
Number Line	Math	0.75 (.05)	1
Magnitude Comparison	Math	0.71 (.06)	2.32 (.36)
Number Identification	Math	0.9 (.04)	4.14 (.50)
Counting	Math	0.82 (.05)	2.46 (.32)

Note. All loadings are significant at p < .001 level. English Proficiency obtained from the Parent Questionnaire. TB = Touch Base (Corsi Block Tapping Task).

Figure 3.2 displays the first model for testing Hypotheses 1 and 2 where language influences both mathematics and executive function, and executive function in turn influences mathematics as well as the standardized coefficients. The standardized coefficients are 0.33 (SE = 0.16, p = .07), 0.59 (SE= 0.12, p < .001), and 0.44 (SE = 0.15, p = .004) for language to math, language to executive function, and executive function to math, respectively for model 1. Even though the directionality is different for model 2, the standardized coefficients remain the same. Therefore, only model 1 is displayed (see Figures 3.2 and Table 3.4). Both models fit the data adequately. The CFI is 1.00, TLI is 1.010, RMSEA is 0.000, and the model does not differ from that of the saturated model, $\chi^2(41) = 38.52$, $p = 0.581$.

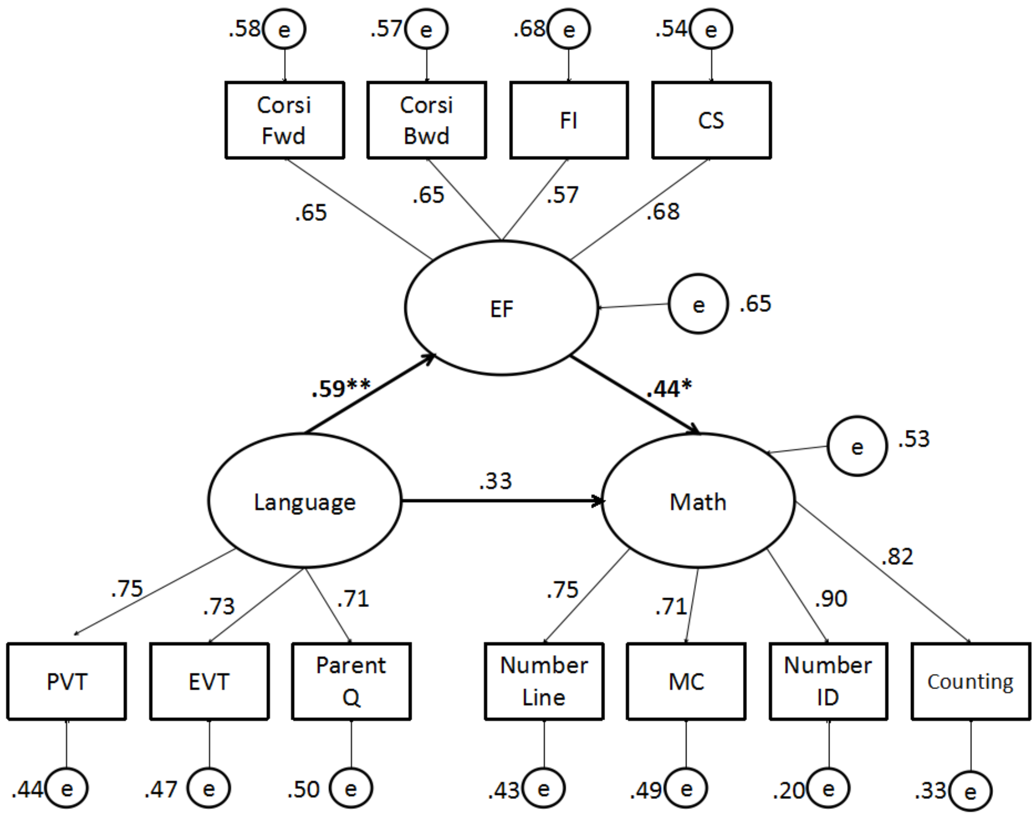


Figure 3.2. SEM of language, executive function, and mathematics with standardized coefficients for model 1. ***p* < .01. * *p* < .05. The structural paths are bolded. All measurement loadings are significant at *p* < .001.

Table 3.4
Direct, Indirect, and Total Effects from SEM of Language, EF, and Math

	Model 1				Model 2			
	B		B		B		β	
	EF (SE)	Math (SE)	EF	Math	Language (SE)	Math (SE)	Language	Math
Direct								
Language	0.06** (.02)	.002 (.001)	0.59** (.12)	0.33		.002 (.001)		0.33
EF		.02* (.008)		.44**	5.58** (1.69)	.02* (.008)	.59**	.44**

Indirect								
Language		.001* (.0006)		.26*				
EF						.01 (.006)		0.19
Total								
Language	.06** (.02)	.003** (.0009)	.59**	.59**		.002 (.001)		0.33
EF		.02* (.009)		.44*	5.58** (1.69)	.03*** (.008)	.59**	.63***

Note. Coefficients are estimated using maximum likelihood with missing values using STATA14. $n = 86$.
$*p < .05$. $**p < .01$. $***p < .001$

Direct Effects. Table 3.4 displays the direct, indirect, and total effects of language and executive function on math. For model 1, language ability was positively associated with executive function skills (standardized coefficient = 0.59, $p = .001$). Additionally, executive function was predictive of math performance (standardized coefficient = .44, $p = .011$). However, language ability was not associated with math performance (standardized coefficient = .33, $p = .07$). Overall, the model explains 0.35 of the variance in executive function and 0.47 of the variance in mathematics.

Although model 1 and model 2 differ in the directionality of the language and executive function path, the standardized coefficients of the links between the three constructs remain the same because SEM uses regression and correlational techniques. Therefore, executive function was positively related with language ability (standardized coefficient = 0.59, $p = .001$, $R^2 = .35$) and predicted math performance (standardized coefficient = .44, $p = .011$). Language is, again, not predictive of math performance (standardized coefficient = .33, $p = .07$). The model also explains .47 of the variance in the latent math construct.

Indirect Effects. In the first hypothesized model, the relationship between language and math appeared be mediated via the indirect effect on executive function (standardized coefficient

= 0.26, $p = .02$). In contrast, the second hypothesized model posits the relationship between executive function and math is mediated through the effect on language. Using the MLMV method, the indirect effect is insignificant (standardized coefficient = .19, $p = .078$).

Total Effects. For model 1, the total effects of all three paths were significant at the .05 level. See Table 3.4 for both the standardized and unstandardized coefficients. In contrast, only executive function has significant total effects on language ability (coefficient = .59, $p = .001$) and math ability (coefficient = .63, $p < .001$) in Model 2.

GMA. The model for testing Hypothesis 3a—that relations among the three constructs of language, EF, and math remains after considering general mental ability—did not converge on Stata14. The overspecified model did, however, converge on Ωnyx. However, because of the discrepancy and the peculiarity resulting from the overspecification, the model for Hypothesis 3a will be discussed in Appendix IX as supplementary material. Hypothesis 3b assumes no relations among the three constructs of interest (language, EF, and math) and therefore no direct, indirect, or total effects. The model generated from Hypothesis 3b (see Figure 3.3) from Ωnyx (it did not converge on Stata14) shows strong association between GMA and language, EF, and math with standardized estimates of .76, .78, and .82 (all *p*s < .001), respectively. The model displays adequate fit with the data (CFI = 0.994, TLI = 0.992, RMSEA = 0.024, $\chi^2(41) = 42.99$, $p = 0.27$).

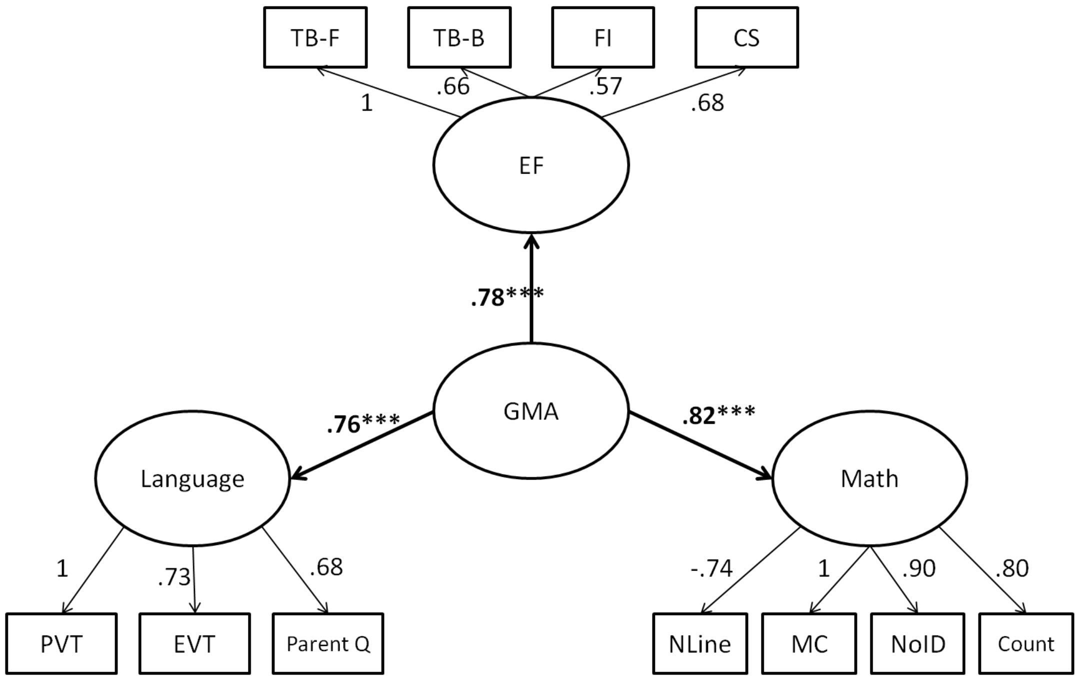


Figure 3.3. SEM of general mental ability, language, executive function, and mathematics with standardized coefficients for Hypothesis 3b. ****p* < .001. * *p* < .05. The structural paths are bolded. The model is based on analysis performed on Ωnyx, as the model would not converge on Stata14.

Discussion

These results suggest that among kindergartners, when all three constructs of language, executive function, and mathematics are considered together, domain-general executive function skills are related to the other two specific domains. Although models constructed from Hypotheses 1 and 2 are equivalent models, the more saturated model (see Figure 3.2) where the third path is freely estimated (instead of constrained to 0 as Hypotheses 1 and 2) resulted in path coefficients that favor Hypothesis 1. Specifically, as depicted in Figure 3.2, language skills predict EF, and EF in turns predicts performance on math tasks, yet there is no significant direct effect of language on mathematics. Therefore, our results with the model containing the three constructs only align more with Hypothesis 1, that the effect of language on mathematics appears to be mediated through its effect on executive function.

The models for testing the hypothesis with general mental ability (GMA) did not converge on Stata14. Based on results from Ωnyx, the model with GMA predicting the three constructs that no longer have direct connections with each other (see Figure 3.3) fits just as well as the model displayed in Figure 3.2. However, as the two models fit equally well and that the one displayed in Figure 3.2 (no GMA) is the more parsimonious one, Occam's razor would demand that the model showing direct connections between the three constructs (rather than connected via a fourth construct of GMA) is preferred. Additionally, the aim of this study is to examine the association of the three constructs without another higher-order construct. Furthermore, the lack of convergence in one of the statistical software suggests that the algorithm employed by Ωnyx may have been more lax in their criteria for convergence. The model that converged in both software programs with estimates that are off by less than a tenth of a standard deviation (see Appendix IX for more details) may in fact be more stable and preferable. The following discussion is, therefore, concentrated on the model displayed in Figure 3.2, with a slant in favor of Hypothesis 1.

Association of EF with Language and Math

The significant associations of executive function skills with both linguistic and mathematical skills align with Geary and Hoard's (2005) framework that domain-general executive function processes are the underlying mechanisms for performances in the mathematical domain. Across multiple measures of math performance, the relatively poor performance of children with math difficulties is mediated by their executive function processes (Geary et al., 2007). It is interesting to see the association between EF and mathematics replicated here to typically developing sample of children from mostly linguistically diverse lower-SES families, suggesting that domain-general skills are required regardless of demographic or individual differences in math skills.

At the same time, the math performance could also be explained by the interplay between linguistic skills and executive processes, as some children with math difficulties also exhibit deficits in both executive processes and language-specific tasks (Geary, Hamson, & Hoard, 2000). Similarly, our findings of the dual association of executive function with language and math also echo Engle's (2002, 2010) theory about working memory capacity where executive processes encompass the regulation of attention. Accordingly, attentional control is necessary in any higher order cognition, be it language processing or mathematical processing. Therefore, performance on domain-general executive function tasks would be associated with performances on language or math tasks.

Language on Math, Mediated via EF

Beyond mere associations, our SEM results suggest that the effect of language on mathematics is mediated through executive function skills. In other words, language has a direct effect on executive function skills, executive function skills have a direct effect on mathematics performance, but there is minimum direct effect of language on math. However, the lack of *significant* direct effect of language on mathematics should be viewed with caution. The loading of 0.33 suggests that a direct effect potentially exists; additionally, the small sample size of 86 may mean a lack of statistical power to detect the effect (indeed, the *p*-value suggests a marginal significance).

Regardless, the minimum direct effect of language on math conflicts with many previous studies on language and cognitive control (e.g., Alloway, Gathercole, & Pickering, 2006; Alloway & Ledwon, 2014; Goo, 2010, 2012; Just & Carpenter, 1992). Although these studies overwhelming suggest that it is the domain-general skills that influence language skills, they typically involve adults. When both domain-general and linguistic skills are developing in children, the story may be slightly different, especially when considering their relation to yet another domain (mathematics). For example, verbal skills have an indirect effect on arithmetic performance in third graders after controlling for working memory and visual-spatial abilities (Vukovic & Lesaux, 2013b).

Moreover, in a time-lag study of children from ages 3 and up, Clark and colleagues (2013) found a similar pattern: children's language comprehension skills at age 3 do not have a significant direct effect on their math performance at age 5; however, they do significantly influence children's executive control skills at the same time point. Our results confirm that when all three constructs are considered together with a more extensive set of tasks, language has only an indirect effect on mathematics. A new contribution of our study is that it demonstrates this pattern without a time lag at kindergarten age and with a sample of mostly emergent bilingual children. That is, without having to look at children's mathematics skills one or two years later, the direct effect of language on *concurrent* math around age 5 or 6 already disappears while accounting for domain-general skills.

Nonetheless, the lack of a significant direct effect on mathematics is inconsistent with a few previous studies that would suggest Hypothesis 2 (Purpura & Ganley, 2014; Purpura & Logan, 2015; Purpura & Reid, 2016). For example, in 4- to 6-year-old children, working memory skills no longer predict performance in counting, number comparison, and number identification after accounting for language skills (Purpura & Ganley, 2014). One potential explanation for this discrepancy is the use of individual tasks versus latent composites. Additionally, even though we also used similar numerical tasks in constructing the latent math composite, we used four working memory measures (Touch Base-Forward, Touch Base-Backward, Following Instructions, and Counting Sheep) to construct a latent executive function variable and three language measures (Receptive One-Word Vocabulary Test, Expressive Vocabulary Test, and parent ratings of English proficiency) to create the language composite. In contrast, because Purpura and Ganley's focus was on disentangling the effects on various mathematical tasks, they only employed one language measure (an expressive vocabulary test) and one word recall working memory measure.

However, in a more recent work where more executive function measures were included, Purpura and Logan (2015) again found executive function skills to be insignificant after controlling for language skills. This discrepancy may be due to the "language" measure used. Purpura and Logan constructed a content-specific "mathematical language" vocabulary list, and the effect of language on mathematics appears to be localized to this content-specific language skill rather than general language skills (Purpura & Reid, 2016).

Language Background and Bilingualism

Yet another potential factor that could explain the discrepancies between our findings and earlier work is the sample. Most of the previous work was conducted with primarily English-speaking children (e.g., Purpura & Ganley, 2014). In contrast, the majority of our participants are emergent bilingual students with a non-English home language. Children from such language backgrounds typically do not fare as well as their native-English speaking counterparts on mathematical assessments (Abedi & Herman, 2010; Kena et al., 2015; Kieffer, Lesaux, Rivera, & Francis, 2009). Furthermore, our sample of students are from low SES backgrounds, and students from low SES backgrounds typically underperform compared with higher SES children (e.g., Jordan & Levine, 2009). Additionally, language skills, *math* language, and different kinds of executive function skills differentially predict early numeracy as a function of the children's math performance (Toll & van Luit, 2014). For example, visual working memory significantly predicts the growth of early numeracy in typically performing children, but not in weaker performers (Toll & van Luit, 2014). These results suggest that the effect of language or working memory on math may be different for different populations.

Nevertheless, a direct comparison of native English speakers and (emergent) bilingual children showed that language abilities are important for both groups of children and that the effect of language on mathematics may be stronger in non-native English speakers after controlling for working memory (Vukovic &Lesaux, 2013a). In the current study, I was unable to directly test the difference between the two language groups, as the low sample size (only n = 33 for children with English as home language). Nonetheless, this present study's finding of minimal direct effect of language on math is with a *majority* of emergent bilinguals, which means that my finding seemingly contradicts with Vukovic and Lesaux's (2013) findings. Though the present study appears to send a different message than the earlier study, there are several key differences. First, only one visual working memory task was used as a control in the previous study; in ours, we created a latent construct from four. It is likely that the importance of executive function only emerges after considering all the multiple types of working memory (both visual and verbal). Second, our participants are younger, consisting of kindergartners. It is possible that the differences from the studies are due to developmental differences between kindergartners and fourth graders. Third, as mentioned previously, we combined both native-English speakers and emergent bilinguals in our study and we did not directly compare children from different linguistic backgrounds. Indeed, one limitation of current study is our relatively small sample size. We do not have enough participants in the two groups to effectively analyze any potential differences between them. Therefore, one direction of future research is to examine the patterns of associations between language, executive function, and mathematics among native English-speaking and emergent bilingual kindergartners.

Directionality

Overall Interpretation. Yet another caveat of current study is its inability to distinguish directionality and address causation among the three constructs. Because SEM analyses are, in their heart, correlational analyses, researchers are cautioned against a strict interpretation of directionality (Thoemmes, 2015). In fact, given the nature of our data, one can interpret our findings two ways. One, EF has significant effect on both language and math but language has no effect on math when all three are considered. Two, language has a direct effect on EF and an indirect effect on math, and EF has a direct effect on mathematics. Correlational analyses cannot distinguish between the two interpretations. However, we lean toward the second interpretation (language with direct effect on EF and indirect effect on math through EF) because of our study design and the use of the parent questionnaire. Parents completed the questionnaire prior to the start of testing, and therefore, it is nonsensical to think that children's performance on various executive function tasks can influence their parents' ratings of their language proficiency.

Between EF and Math. Although our model (and our interpretation) suggests that executive function influences math, the cross-sectional nature of our data does not allow us to test for this supposed directionality. One can easily interpret the significant connection between executive function and mathematics as math having a direct effect on executive function. In fact, even though earlier literature suggests that working memory and executive function skills predict mathematics skills, the reverse also appears to be true (Clements, Sarama, & Germeroth, 2016; Watts et al., 2015). Using a longitudinal dataset, Watts and colleagues (2015) found that first graders' math knowledge actually strongly predicts their third grade EF skills, as measured by Tower of Hanoi, Memory for Sentences, and Continuous Performance Task. Additionally, high quality math program such as *Building Blocks* (Clements & Sarama, 2013) have been shown to improve children's EF as well (Weiland & Yoshikawa, 2013). Unfortunately, our data and the resulting SEM findings cannot tease apart the link and directionality between executive function and mathematical skills in young children. Future longitudinal studies should further investigate the potential bidirectionality of EF and mathematics.

Conclusion

Despite the limitations of the current study, it contributes to the field by being among the first to examine the three interrelated constructs of language, executive function, and mathematics together among a diverse group of kindergartners. By using latent composites for all three constructs, the study is able to speak in more general terms on the indirect effect of language on mathematics and the direct effect of executive function on mathematics rather than focusing on specific measures of any of the constructs. In sum, our findings show that for linguistically diverse kindergartners from mostly lower socioeconomic backgrounds, language skills appear to influence their executive function skills, which in turn affect their mathematical skills. Regardless, future research should address the key caveats of the current study with research designs that could tease apart equivalent and/or alternate, plausible models. For example, future studies could consider increasing the sample size to further examine the difference between native English speakers and emergent bilinguals, implementing a longitudinal design to even better address directionality, and perhaps even conducting randomized-control interventions targeting a specific construct to better establish causation.

CHAPTER 4

Study 2. Learning Mandarin as a second language in an immersion setting vs. receiving mainstream English instructions: A Bayesian comparison of kindergartners' numerical cognition

In Study 1, the linguistic aspect centers on children's English language vocabulary performance and their parents' rating of their English abilities. Here, we shift our focus to a more global, almost external, view of language and investigate how learning a new language, particularly a language with a transparent base-10 number system, may or may not influence kindergartners' numerical cognition.

Impact of Language Structure

Grammatical structure

Although we tend to think of language as something that is internal to a person and as skills a person could acquire, the various features of different languages, which may reflect the macro-level cultural differences, can influence how a child learns.

Cross-linguistic and cross-cultural studies have demonstrated how language features and structures can affect mathematics. The grammatical structure of different languages has, for one, been shown to influence children's understanding of cardinality and plurality. For example, Russian and American three-year-old children, whose native languages include singular/plural markers, outperform age-matched Japanese children on "Give-N" tasks that tested their cardinality (Sarnecka et al., 2007). Similarly, researchers comparing Chinese and American 2- and 3-year-olds also found that English-speaking children more readily acquire understanding of number words beyond "one" as the Chinese language also lacks singular/plural markers (Li et al., 2013).

Number Words

On top of the grammatical features, the number words themselves also exert great influence over the way we conceptualize numbers and math. For example, the Munduruku people indigenous to the Amazon River basin in Brazil lack number words beyond five, and they use words for 3, 4, and 5 as approximate representations (more akin to what English speakers would use "some," "few," or "a handful") instead of exact quantities of exactly three, four, or five objects (Pica, Lemer, Izard, & Dehaene, 2004). Although they are able to add and compare approximate numbers, they are less able to perform exact calculation tasks. When tasked with exact subtraction, their performance decreases drastically if the initial number is above their number naming range. Pica et al.'s finding suggests that exact arithmetic, or more precisely, subtraction, is at least partially dependent on the language of numbers.

Modern societies and languages stand in contrast to that of the Munduruku in that the number systems in most languages can cover a whole number range from negative to positive infinity, or at least very large numbers. Yet the subtle differences in how these number systems are constructed in various languages seem to translate to broader numerical cognition and knowledge.

Transparency of number words. In particular, the base-10 transparency of different languages' number words appears to be crucial in early math learning. For example, East Asian languages (e.g., Korean, Japanese, Chinese) have a succinct syntax for number words such that eleven is "ten-one," twelve "ten-two," twenty "two-ten," and so on. The structure of these number words parallels that of written Arabic numerals and facilitates an earlier mastering of the place-value notation in base-10 system (Geary, Bow-Thomas, Fan, &Siegler, 1993; Laski &Yu, 2014; Miller, Kelly, &Zhou, 2005; Miller, Smith, Zhu, &Zhang, 1995). Conversely, decade-unit inversion of number words[4] that occur in many languages (e.g., Dutch, German) make numbers more difficult to transcribe, and errors related to this inversion is negatively associated with children's math performance (van derVen, Klaiber, & van derMaas, 2014; Xenidou-Dervou, Gilmore, van der Schoot, & van Lieshout, 2015). The higher cognitive load associated with inverted numbers makes it difficult for Dutch-speaking children who were starting to learn arithmetic to perform as well as their English-speaking counterparts in doing symbolic approximation tasks (Xenidou-Dervous et al., 2015). These multiple cross-linguistic and cross-cultural studies demonstrated the effect of the number-language transparency on children's ability to acquire numerical skills such as counting or addition.

However, with the exception of Laski and Yu's (2014) Xenidou-Dervous et al.'s (2015) studies that accounted for school system differences (they assessed younger English-speaking children because formal math instruction started earlier in the UK than in the Netherlands), most were not able to or did not account for the differences between school systems. Also problematic in the interpretation of these findings is the potential cultural difference in perspectives on mathematics learning or cultural tools associated with calculations. In certain cultures, mathematical terms such as count words may appear more frequently in a child's life (Ng & Rao, 2010), and the amount of number-related talks or activities at home is essential in children's numeracy development (Ramani, Rowe, Eason, & Leech, 2015).

Learning in a Different Language

One way to examine the effect of language system in learning math despite cultural background differences is to examine individual learners learning and doing mathematics in more than one language.

[4] In languages such as Dutch, numbers above 20 have an inverted tens and ones unit such that the ones unit is named first. For example, in Dutch, 48 is "achtenveertig" (literally, eight and forty) instead of "forty-eight."

Classic Studies. In fact, the language in which people learn mathematics is so deeply entrenched that even when they become fluent in a second language, individuals still prefer to perform exact number tasks in the language of original instruction (Devlin, 2010). One classic example in support of this view involves bilingual students. A group of English-Russian bilingual participants were taught new addition facts in either English or Russian before they were tested in one of the two languages. When the math tasks involved figuring out the exact number (instead of approximation), the participants took considerably longer to answer when the test was in a language different from the language of instruction (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Spelke & Tsivkin, 2001).

Natural Experiments. A number of studies have since replicated the importance and the use of bilinguals' first language (or the first language in which they learned math) in processing arithmetic regardless of the exact first or second language of the participants (e.g., Salillas & Carreiras, 2014; Wang, Lin, Kuhl, & Hirsch, 2007). Recently, studies conducted in Luxembourg provide additional insights into the effects of mathematical language instruction and number word structures in emerging or balanced bilinguals' arithmetic performance (Van Rinsveld, Brunner, Landerl, Schiltz, & Ugen, 2015). In Luxembourg, children begin primary school instructions in German with French language learning starting in the second grade. However, in the seventh grade, the math instruction switches over to French. Participants of various age groups (grades 7, 8, 10, and young adults) all solved simple one-digit addition questions (e.g., 4+2) faster in German, the language in which they first learned math, than in French. Moreover, the seventh graders were the only ones who were less accurate in French than in German, presumably because they have had less practice in French (Van Rinsveld et al., 2015).

The researchers demonstrated the number word structure effects by including more complex addition questions with addends greater than 70. Number words above 70 in French follows a base-20 format; in contrast two-digit number words in German holds an inversion property but keeps a consistent base-10 format. Comparing within the individuals, the accuracy difference between German and French was much larger with addition of numbers above 70, suggesting that the relative base-10 transparency of German does indeed foster better arithmetic performance (Van Rinsveld et al., 2015).

Language Immersion. It goes to reason, then, that if children were to start learning math in a language having a more base-10 transparent number system, they should be able to benefit from the corresponding numerical knowledge advantages seen in cross-cultural or natural experiment studies. Studies of 6- and 8-year-old children in Wales attending Welsh immersion program suggest that is indeed the case (Dowker, Bala, & Lloyd, 2008; Dowker & Roberts, 2015). The Welsh number system used in school is base-10 transparent where numbers such as 11 (un deg un) and 21 (dau ddeg un) correspond to "one ten one" and "two tens one" respectively. Compared with children enrolled in English-instruction classes, those enrolled in Welsh-medium classes made fewer errors in number line estimation tasks (Dowker & Roberts, 2015) and magnitude comparison (Dowker et al., 2008). Further, children whose home and school languages were both Welsh outperformed those who speak English at home but Welsh at school (Dowker et al., 2008). Together, the findings suggest that not only is base-10 transparent number systems predictive of math performance, children can benefit from such system by learning it—in a different language—in school.

Although one of the best advantages of the Welsh studies is that the children are from similar cultural backgrounds so potential confounds from cultural differences are removed, it also leaves the question of whether the supposed base-10 transparent language advantages can be found where the home language and culture are in stark contrast with that of instruction. Therefore, one major goal of the current study is to examine whether learning in an Asian language that has a base-10 transparent number system can benefit American children whose home language is English only:

RQ1. Is learning in a language with base-10 transparent number system, specifically Mandarin Chinese, predictive of American children's performance on various mathematical measures?

The Role of Working Memory

In addition to investigating the potential difference due to language of instruction, we are also interested in how children's working memory capacity may play a role. It has been established that number word structures in certain languages require higher working memory capacity to process (Helmreich et al., 2011; Pixner et al., 2011; Xenidou-Dervou et al., 2015; Zuber, Pixner, Moeller, & Nuerk, 2009). For example, the inversion property of Dutch and German number words increases the cognitive load for processing numbers. Comparing symbolic addition performance between English-speaking children and Dutch-speaking children, the researchers found that working memory capacity modulate the math performance according to item difficulty level in English-speaking children. That is, English-speaking children with higher working memory capacity, as measured by word recall forward and backward tasks, performed better on difficult addition questions than those with lower working memory. However, working memory capacity appeared unrelated to symbolic addition performance across difficulty levels for Dutch-speaking children, suggesting a generally high demand of working memory for processing even easy numbers in Dutch (Xenidou-Dervou et al., 2015). Not only do number word structures dictate how much working memory is required, the phonological length of number words may also affect performance in arithmetic tasks (Klessinger, Szczerbinski, & Varley, 2012). Examining participants' performance for addition questions with different phonological lengths (e.g., 47 + 16 is considered long because it contains six syllables; 45+18 is considered short because it contains five syllables), researchers found young adults to be significantly quicker but equally accurate at answering questions with shorter phonological lengths compared with longer phonological lengths (Klessinger et al., 2012). Because English and Mandarin Chinese have different base-10 transparency levels and that Chinese numbers are arguably shorter (each unit digit is spoken as a one-syllable word, each number in the teens from 11-19 is spoken as a two-syllable word, e.g., shi-yi for 11, and every other double-digit number is spoken as a three-syllable word, e.g., 37 as san-shi-qi), the amount of working memory required to process math in either language may differ.

Furthermore, language learning by itself requires intense concentration and working memory and may lead to changes in children's working memory capacity (WMC) by way of changes on the neural level. Using neuroimaging techniques (i.e., diffusion tensor imaging), researchers have found bilateral differences in white matter structure between Spanish-English bilingual who have moved to the US in adulthood and native English-speaking monolinguals (Kuhl et al. 2016). These differences cover regions associated with not only language processing and production, but also executive function. Furthermore, the differences appear more pronounced for individuals who have been exposed to their second language longer (Kuhl et al., 2016). Behavioral studies also separately suggest a potential change in WMC. Comparing monolingual Spanish speakers and their peers attending English immersion program, the emergent bilingual children who have been attending the immersion program outperformed their peers in a classic measure of working memory (Hansen et al., 2016). Considering that working memory is implicated in math learning and that immersion program is likely to influence working memory, a second aim of the current study is to examine the role of a transparent base-10 number system in face of working memory differences. Specifically:

RQ2. Are there working memory differences between children attending typical English instruction classes and those attending Mandarin immersion classes? If so, how predictive is learning a base-10 transparent number system of math performance after taking children's working memory into account?

For the first research question, we hypothesize that learning a base-10 transparent language system—Mandarin Chinese—is positively associated with math learning. For the second question, we hypothesize that the base-10 transparency advantage may remain even after taking working memory into account for some of the math measures. In particular, the following are the hypotheses for the current study.

$H_{1.0}$: There is no relationship between receiving instruction in Mandarin Chinese, a base-10 transparent number system, as opposed to receiving instruction in English, and math outcome.

$H_{1.1}$: Relative to receiving instructions in English, learning Mandarin Chinese, a base-10 transparent number system, is positively associated with children's math performance.

$H_{2.0}$: Accounting for working memory differences, there is no relationship between learning Mandarin Chinese, a base-10 transparent number system, as opposed to receiving instruction in English, and math outcome.

$H_{2.1}$: Relative to receiving instructions in English, learning Mandarin Chinese, a base-10 transparent number system, is positively associated with children's math performance even after accounting for working memory differences.

Method

Participants

A total of 120 kindergartners from four schools in three Southern California school districts participated in the study. Eighty-six were enrolled in typical English-instruction programs, and the remaining 34 were enrolled in a Mandarin immersion program. Of the 120 students, 61 children whose home language is indicated as English were included in the final analyses. Table 4.1 displays the demographic breakdown of the final analysis sample. The two groups of children did not differ in terms of age, $t(59) = 0.38$, $p = .71$. However, maternal education level was different between the two groups, χ^2 (5) = 32.07, $p < .001$, as was race/ethnicity, χ^2 (3) = 21.67, $p < .001$.

Table 4.1
Descriptive Statistics of Demographic Variables

	English Instruction N = 33		Mandarin Instruction N = 28	
Age in months, M(SD)	74.65 (3.48)		71.71 (3.92)	
Gender	N	%	N	%
Female	17	51.52	16	57.14
Male	16	48.48	12	42.86
Race & Ethnicity of Children				
Caucasian/White	22	66.67	8	28.57
Hispanic or Latino	25	75.76	0	0
Biracial/Mixed Race	3	9.09	10	35.71
Asian American or Pacific Islander	2	6.06	10	35.71
Maternal Education				
Some high school	1	3.03	0	0
High school diploma/GED	9	27.27	0	0
Some college/vocational training	11	33.33	1	3.57
2-year College Degree (Associates)	5	15.15	1	3.57
4-year College Degree (BA/BS)	3	9.09	15	53.57
Postgraduate or Professional degree	4	12.12	11	39.29

Note. For Race/Ethnicity, the percentage do not add up to 100 as race and ethnicity were separate categories and one can be Hispanic/Latinx while still being of another race (e.g., Pacific Islander and Hispanic or Caucasian/White and Hispanic/Latino).

The Mandarin immersion program structures their classes in an 80:20 fashion. That is, 80% of the instruction is intended to be in the target language (i.e., Mandarin) and the remaining 20% is in English. Both the Mandarin immersion program and the regular English instruction classes implemented curricula that adhere to the Common Core Math standards. (In contrast, in 2016 during the pilot data collection at the Mandarin Immersion program, the program used Singapore math curriculum.)

Measures

Students in both types of programs completed the measures stated in Chapter 2. The following measures were used in this study; refer to Chapter 2 for a full description.

Touch Base Forward and Backward. A gamified tablet version of the Corsi Block Tapping Task. Though studies have shown the forward and backward versions to display similar results (e.g., Donolato, Giofrè, & Mammarella, 2017; Isaacs & Vargha-Khadem, 1989; Kessels, van Zandvoort, Postma, Kappelle, & deHaan, 2000) and that for children ages 4-6, distinction between visuospatial short term memory and working memory may be lacking (Alloway et al., 2006), research also suggests that backward Corsi may be tapping other executive control and spatial processing components (e.g., Cornoldi & Mammarella, 2008; Higo, Minamoto, Ikeda, & Osaka, 2014). Therefore, for analysis purposes, forward and backward performances (number of trials correct) on Touch Base were kept as separate variables.

Magnitude Comparison. 24 questions of symbolic number comparison. Participants were asked to select the larger of the two numbers. The outcome variable from this task is number of questions correct.

Numerical Identification. 24 questions of single or double-digit symbolic numbers presented on index cards. Children were asked to name the numbers. The outcome variable from this task is number of questions correct.

Number Line Estimation. 26 numbers were presented one at a time above a number line on a tablet. Children then indicate where they the number lies between 0 and 100 on the line. The outcome variable from this task is percentage absolute error.

Counting: Children counted set quantities of manipulatives and counted by rote up to 120. A total of 14 points were possible. The outcome variable for this task is the number of points received.

Procedure

The study design and materials were approved by our Institutional Review Board before the researchers reached out to schools. After gaining approval from the school principals and, in one case, the school district as well, the researchers passed out informed consent forms with parent questionnaires home to the parents. Children whose parents signed their approval were then invited to participate in the study. Because of the number of measures involved in the study, testing was broken down into multiple sessions to decrease the possibility of fatigue. Before each testing session, the experimenters asked each participant for his/her verbal assent (e.g., "we will be playing some math games today, will you like to play?"). None of the children decided to withdraw from the study.

Testing Language. As in Dowker and colleagues' (2008; 2015) studies, in the one-on-one testing sessions, the experimenter used the language of instruction (i.e., English in typical English instruction classes and Mandarin in the Mandarin immersion classes) to administer the tasks. However, unlike the Welsh/English studies, the experimenter also used English in the

Mandarin immersion program and asked children to answer in both languages where applicable (e.g., number identification and counting) in order to avoid confounding children's Chinese language abilities with their math performance. For a more direct comparison of children's math performance across language programs, only children's English responses were used for these measures.

Analysis Plan

I conducted a series of multiple regressions and supplemented the OLS regression with Bayesian regression analyses to compare the math performances between children in the Mandarin immersion program and those in typical instruction classes. Bayesian regression is used to supplement the frequentist approach because the theoretical underpinnings and assumptions of frequentist methods (e.g., asymptotic normality) require a large sample; considering the small dataset for this study, Bayesian analysis will provide more accurate estimates (McNeish, 2016; van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015). Moreover, in exploratory analyses, the Bayes factor (BF) for each model allows for a more direct comparison between models in selecting the final regression model for each math outcome variable. The interpretation of the Bayes factor for each remaining predictor variable is also more intuitive (e.g., predictor A is x times more likely to have an effect on the outcome).

All Bayesian analyses were conducted using JASP (2017). I set the priors for the Bayesian regressions as the default for several reasons. First, the default prior on JASP had been carefully considered to meet the theoretically desirable criteria for their resulting BFs: scale and location invariance as well as consistency in information regardless of sample size (see Rouder & Morey, 2012). Second, the default priors set a uniform distribution such that the null and alternative hypotheses are equally likely (Marsman & Wagenmakers, 2017; Wagenmakers et al., 2017). Therefore, the use of the JASP default priors is a *conservative* approach when comparing the children enrolled in the Mandarin immersion program and those in the typical English instruction classes, as pilot data suggest that children in Mandarin immersion program outperforms those in typical instruction classes. However, since Bayesian analysis is suitable for small sample size *if* informative priors are taken into account, I also calculated analyses with various prior values as a robustness check. The results did not vary, so only the results calculated with the default priors are reported below.

To investigate whether any effect may be attributed to the base-10 transparency of the Chinese number system, I also performed item-level analyses on the numbers presented in the number line task. If base-10 transparency is indeed the main, direct reason behind any positive results of the immersion program, we should not expect any differences in the single digits (equally transparent in both languages), but instead see positive results in the 10s and perhaps the 20s decades. (English number system *starts* becoming transparent in the 20s decade, but as the transparency only starts in that decade, it presents a different challenge "switch" for children where the number system acquires a different pattern.)

Results

Descriptive Statistics

Table 4.2 displays the descriptive statistics of the outcome performance by the groups. Comparisons of the receptive and expressive vocabulary performance of children from the two programs confirm that the two groups do not differ in their language skills (ps > .05 for both). Within children from the Mandarin immersion program, more items were completed using English (see Appendix X for a breakdown of the number of items correctly answered in each language for the two vocabulary tests). For all variables of interests (the two working memory measures and the four math measures), children from the Mandarin immersion program outperform those from typical English instruction ones (*p*s < .05; Cohen's d ranges from .656 to 1.565).

Correlation

More surprising is the finding that the associations among the variables display a different trend between the two groups (see Table 4.3). For children from the Mandarin immersion program, only the math measures (number line estimation, magnitude comparison, number identification, and counting) are significantly correlated with each other with the exception of number identification and number line estimation (r = -.32, p = .09). Children's math performances were not correlated with their performances on either of the working memory tasks. In fact, their performances on Touch Base forward and backward performance was also not related (r = .21, p = .29). In contrast, for the children in the typical English instruction classes, performance on the Touch Base forward task was significantly correlated with number identification (r = .43, p = .012) and the backward task (r = .42, p = .015). All four of the math measures were also moderately or strongly correlated with one another (*r*s range from -.38 to -.68, *p*s range from .028 to <.001). The reliabilities of the measures were comparable across the two groups, ranging from Cronbach's alpha of .59 for Corsi forward to .96 for number identification.

Table 4.2

Descriptive Statistics and Statistical Comparisons of Measures Between Children from Different Instruction Types

Measures	English Instruction				Mandarin Immersion				t-test (group comparison)			
	n	M	SD	range [min, max]	n	M	SD	range [min, max]	p-value	effect size	BF	
PVT	24	70.13	14.94	[19, 94]	28	76.30	11.27	[44, 99]	0.100	0.47	BF10	0.88
EVT	26	27.96	3.44	[19.5, 34.5]	28	27.97	1.52	[23.5, 30.5]	0.987	0.00	BF10	0.27
TB Forward	33	3.55	2.12	[0, 7]	28	5.39	1.81	[2, 9]	<.001	0.93	BF10	46.51
TB Backward	33	1.91	1.35	[0, 5]	28	4.14	1.51	[1, 7]	<.001	1.57	BF10	119354.98
Number Line (PAE)	33	20.01	7.87	[4.11, 35.81]	28	15.19	6.48	[5.94, 31.17]	0.012	0.66	BF+0	3.97
Magnitude Comparison	31	20.48	3.84	[11, 24]	28	22.96	2.25	[14, 24]	0.004	0.78	BF-0	9.54
Number Identification	32	18.91	6.67	[2, 24]	28	22.57	4.00	[6, 24]	0.008	0.66	BF-0	3.62
Counting	33	9.94	2.90	[3, 14]	28	12.57	1.81	[6, 14]	<.001	1.07	BF-0	209.54

Note. PVT = Receptive Picture Vocabulary Test. EVT = Expressive Vocabulary Test. TB = Touch Base (Corsi Block Tapping Task). PAE = Percentage Absolute Error. M = mean. SD = Standard Deviation. Effect size reported for t-test is Cohen's d. BF = Bayes Factor. BF10 is the Bayes factor in favor of the alternative hypothesis with the hypothesis test of no difference between the two groups. For the math measures, the BF is calculated with directional hypothesis such that BF+0 indicates that the hypothesis is the first group (English instruction) will have a more positive score for PAE, and BF-0 indicates the hypothesis that the first group (English instruction) will have more negative scores on magnitude comparison, number identification, and counting.

Table 4.3

Pairwise Correlations Between Outcome Variables, Separated by Instruction Types

English Instruction

	PVT	EVT	TB-F	TB-B	NL	MC	NID	Counting
PVT	--							
EVT	**0.54** *	--						
TB-F	**0.48 ***	**0.46** *	--					
TB-B	**0.54** **	0.18	**0.42** *	--				
NL	0.09	**-0.39** *	-0.25	-0.11	-			
MC	0.16	0.24	0.30	0.25	**-0.38** *	--		
NID	0.18	**0.53 ** **	**0.43** *	0.17	**-0.68 *** **	**0.52** **	--	
Counting	**0.47 ***	**0.40** *	0.27	0.34	**-0.56 ** **	**0.43 ***	**0.59** ***	--
Reliability	0.96	0.80	0.79	0.58	0.84	0.83	0.96	0.71

Mandarin Immersion

	PVT	EVT	TB-F	TB-B	NL	MC	NID	Counting
PVT	--							
EVT	-0.03	--						
TB-F	0.17	0.12	--					
TB-B	0.13	0.06	0.21	--				
NL	-0.23	0.02	-0.18	-0.17	--			
MC	0.32	-0.08	0.24	-0.14	**-0.56** **	--		
NID	0.12	0.16	0.07	0.08	-0.32	**0.50** **	--	
Counting	0.20	0.31	0.20	0.18	**-0.46** *	**0.53** **	**0.83** ***	--
Reliability	0.94	0.59	0.59	0.60	0.86	0.88	0.96	0.72

Note. PVT = picture vocabulary test. EVT = expressive vocabulary test. TB = Touch Base (Corsi Block Tapping Task). TB-F = Touch Base Forward. TB-B = Touch Base Backward. NL = Number Line (Percentage Absolute Error). MC = Magnitude Comparison. NID = Number Identification. Significant correlations are bolded. **p* < .05, ** *p* < .01, ****p* < .001. The reliabilities reported are Cronbach's alpha for each measure.

Regression Analyses

As the immersion program and typical instruction students do not differ in terms of their language skills, the vocabulary measures were excluded from the analyses. In contrast, kindergartners' working memory performance as measured by the Touch Base forward and backward versions differ significantly and show different correlation patterns between the groups; they were thus included in subsequent regression models.

Number line estimation. Table 4.4 displays the results of OLS and Bayesian regression analyses of children's number line estimation performance on their instruction type (Mandarin immersion or not) and other predictors. Model 1 shows the simplest regression with immersion program as the only predictor variable. Model 2 adds the mother's education to the model, and Model 3, the full model, includes both working memory measures. Model 4 is the model that best fits the data according to Bayes factors and still contains the immersion program as a predictor variable. The Bayes factors (BF) for each of the models are included. In addition to the BFs for regression models, the BFs for each individual predictors within the last few models are displayed in Table 4.5 along with the predictors' semipartial correlations.

Controlling for maternal education did not yield a better model fit (Model 2 BF = 1.80). However, after accounting for maternal education, children in immersion program outperformed children in the English instruction classes by an average of 6.3% (p = .013), and there is moderate evidence that the immersion program has a unique contribution to children's number line estimation skills (BF = 4.187). Including the working memory measures also did not yield a better fit (Model 3 BF = 1.24), and this model presents anecdotal evidence for the unique contributions of immersion program (BF = 1.24) and Touch Base forward performance (BF = 1.43) on children's number line estimation skill. Interestingly, the best model based on BFs is the one with only Touch Base Forward (Model 5), but this model is no different from Model 4 (BF = 4.41 versus 4.62). In Model 4, there is only anecdotal evidence for unique contributions of either immersion program or performance on Touch Base forward.

Table 4.4
Regression of Number Line Estimation Performance on Immersion Program and Other Predictors

	Model 1	Model 2	Model 3	Model 4	Model 5
Immersion	-.048*	-.063*	-.042	-.033	
	(.02)	(.03)	(.03)	(.02)	
Mother's Ed		.008	.009		
		(0.01)	(.01)		
TB Forward			-.007	-.008†	-.011*
			(.01)	(.01)	(.01)
TB Backward			-.005		
			(.01)		
Constant	.200***	.174***	.204***	.229***	.228***
	(.01)	(.03)	(.03)	(.02)	(.02)
Adj-R2	.086	.084	.110	.116	.092
BF	3.97	1.80	1.24	4.41	4.62
N	61	61	61	61	61

Note. *** $p < .001$, *$p < .05$, † $p < .10$. Standard errors are in parentheses. Immersion is an indicator variable of whether the student attends the Mandarin immersion program. Mother's Ed (education) is an ordinal variable. Results with indicator variables of the categories of maternal education display the same trend. For simplicity's sake, only the ordinal variable is used and displayed on the table. TB Forward and Backward are the points received in the forward and backward versions of the Touch Base game.

Table 4.5
Unique Contributions of Individual Predictors on Number Line Estimation

	Model 2			Model 3			Model 4		
	Sr	Sr^2	BF	Sr	Sr^2	BF	Sr	Sr^2	BF
Immersion	-.32	.102	4.19	-.19	.035	1.24	-.20	.044	1.49
MEd	.11	.012	0.44	.14	.019	0.46	--		--
TB-F				-.18	.033	1.43	-.21	.050	1.82
TB-B				-.08	.007	0.70	--		--

Note. Sr = semipartial correlation. Sr^2 = semipartial correlation squared. BF = Bayes factor. MEd = Mother's Education. TB-F = Touch Base Forward. TB-B = Touch Base Backward.

Magnitude Comparison. Table 4.6 displays the regression results for children's performance on the magnitude comparison task. Model 1 is the simple regression with only the immersion program as the predictor variable. Model 2 adds maternal education as a predictor while Model 3 includes children's working memory performance. Model 4 is the best model that fits the data according to BF. Table 4.7 displays the semipartial correlations and individual BFs for predictors in Models 2-4.

Controlling for maternal education, children in the immersion program on average scored 2.5 points higher on the magnitude comparison task (*p* = .035), and the immersion program explains over 7% unique variance, BF = 5.456. However, the best model according to Bayes factors is actually one that includes only immersion program and performance on the Touch Base Forward task. This model is over six times more likely as the full model (Model 3), but barely 1.5 times as likely as the simplest model. Using the best model, children in the immersion program on average performed 1.8 points (*p* = .051) higher than children in the typical English instruction class, controlling for the forward version of visuospatial working memory. Though the difference is not significant, there is anecdotal evidence (BF = 2.42) that the immersion program explains about 5.8% unique variance after accounting for working memory.

Table 4.6.
Regression of Magnitude Comparison on Immersion Program and Other Predictors

	Model 1	Model 2	Model 3	Model 4
Immersion	2.48**	2.50*	1.72	1.78
	(0.83)	(1.15)	(1.25)	(0.89)
Mother's Ed		-0.01	-0.16	
		(.39)	(.39)	
TB Forward			0.38	0.40
			(0.22)	(.21)
TB Backward			0.19	
			(0.31)	
Constant	20.48***	20.51***	19.27***	19.01***
	(0.57)	(1.40)	(1.51)	(0.95)
Adj-R2	0.12	0.10	0.14	0.16
BF	9.54	2.90	2.26	13.48
N	59	59	59	59

Note. *** $p < .001$, **$p < .01$, *$p < .05$. Standard errors are in parentheses. Immersion is an indicator variable of whether the student attends the Mandarin immersion program. Mother's Ed (education) is an ordinal variable. Results with indicator variables of the categories of maternal education display the same trend. For simplicity's sake, only the ordinal variable is used and reported on the table. TB Forward and Backward are the points received in the forward and backward versions of the Touch Base game.

Table 4.7
Unique Contributions of Individual Predictors on Magnitude Comparison

	Model 2			Model 3			Model 4		
	Sr	Sr^2	BF	Sr	Sr^2	BF	Sr	Sr^2	BF
Immersion	.27	.072	5.46	.17	.029	1.47	.24	.058	2.43
MEd	-.003	.000	0.40	-.05	.003	0.39	--		--
TB-F				.21	.042	1.72	-.23	.054	2.09
TB-B				.08	.006	0.66	--		--

Note. Sr = semipartial correlation. Sr^2 = semipartial correlation squared. BF = Bayes factor. MEd = Mother's Education. TB-F = Touch Base Forward. TB-B = Touch Base Backward.

Number Identification. Table 4.8 displays the regression results for performance on the number identification task, and Table 4.9 displays the corresponding semipartial correlations and BFs for individual predictors. The best model (Model 5, BF = 7.414) is the model that includes only the Touch Base forward performance. Even if other predictors are included, performance on the forward version of the Touch Base game still marginally predicts performance on number identification (Model 3, b=.73, p = .068) and explains 5.3% unique variance, BF = 2.36, in children's number identification performance. However, the simple model with only the Touch Base forward performance is over seven times as likely as the full model. So the effect of Touch Base performance is much higher, such that a one-point increase in Touch Base forward is associated with a 0.97-point increase in number identification. Attending immersion program, however, does not predict performance on the number identification task, as the Bayes factors suggest only weak evidence for and against the null hypothesis (BFs range from 0.88 to 1.79 controlling for various other predictors).

Table 4.8
Regression of Number Identification on Immersion Program and Other Predictors

	Model 1	Model 2	Model 3	Model 4	Model 5
Immersion	3.67^{*}	2.83	2.06	2.38	
	(1.45)	(1.99)	(2.16)	(1.55)	
Mother's Ed		0.42	0.23		
		(0.68)	(0.69)		
TB Forward			0.73	0.74^{*}	0.97^{**}
			(.39)	(0.37)	(0.34)

TB Backward			-0.05		
			(0.55)		
Constant	18.91***	17.55***	15.59***	16.20***	16.27***
	(0.99)	(2.44)	(2.64)	(1.65)	(0.97)
Adj-R2	0.084	0.074	0.100	0.130	0.110
BF	3.62	1.36	0.95	6.14	7.41
N	60	60	60	60	60

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. Standard errors are in parentheses. Immersion is an indicator variable of whether the student attends the Mandarin immersion program. Mother's Ed (education) is an ordinal variable. Results with indicator variables of the categories of maternal education display the same trend. For simplicity's sake, only the ordinal variable is used and reported on the table. TB Forward and Backward are the points received in the forward and backward versions of the Touch Base game.

Table 4.9
Unique Contributions of Individual Predictors on Number Identification

	Model 2			Model 3			Model 4		
	Sr	Sr^2	BF	Sr	Sr^2	BF	Sr	Sr^2	BF
Immersion	.18	.032	4.19	.12	.014	0.88	.19	.035	1.16
MEd	.08	.006	0.44	.04	.002	0.52	--		--
TB-F				.23	.053	2.36	.25	.060	2.94
TB-B				-.01	.000	0.42	--		--

Note. Sr = semipartial correlation. Sr^2 = semipartial correlation squared. BF = Bayes factor. MEd = Mother's Education. TB-F = Touch Base Forward. TB-B = Touch Base Backward.

Counting. Simple regression with the immersion program as the only predictor shows decisive evidence (BF = 209.79) that children in the immersion program outperformed those in typical English instruction classes by an average of 2.6 points ($p < .001$). The effect remains after controlling for maternal education (see Tables 4.10 and 4.11, Model 2), though the model is three times weaker than the simple model. (In other words, the data is three times more likely to occur under Model 1 than Model 2.) No effect remains with the full model, though there were anecdotal evidence for the unique contributions of the immersion program and for the backward version of the Touch Base game (see Table 4.11, Model 3). Nevertheless, in the best model according to Bayes factor (BF = 392.43), both immersion program and children's performance on the backward version of the Touch Base game remain as significant predictors. Controlling for performance on the backward Touch Base, children in the immersion program on average received 1.6 more points on counting (p = .049). However, the remaining BF for either predictor can only be considered as anecdotal evidence (Wagenmakers et al. 2017).

Table 4.10
Regression of Counting on Immersion Program and Other Predictors

	Model 1	Model 2	Model 3	Model 4
Immersion	2.63^{***}	2.22^{*}	1.18	1.58^{*}
	(0.63)	(0.84)	(0.92)	(0.79)
Mother's Ed		0.21	0.13	
		(0.29)	(0.28)	
TB Forward			0.22	
			(0.17)	
TB Backward			0.36	0.47^{*}
			(0.23)	(0.22)
Constant	9.94^{***}	9.22^{***}	8.05^{***}	9.04^{***}
	(0.43)	(1.05)	(1.14)	(0.59)
Adj-R2	0.213	0.207	0.259	0.259
BF	209.79	68.70	90.76	392.43
N	61	61	61	61

Note. *** $p < .001$, *$p < .05$. Standard errors are in parentheses. Immersion is an indicator variable of whether the student attends the Mandarin immersion program. Mother's Ed (education) is an ordinal variable. Results with indicator variables of the categories of maternal education display the same trend. For simplicity's sake, only the ordinal variable is used and reported on the table. Touch Base Forward and Backward are the points received in the forward and backward versions of the Touch Base game.

Table 4.11
Unique Contributions of Individual Predictors on Counting

	Model 2			Model 3			Model 4		
	Sr	Sr^2	BF	Sr	Sr^2	BF	Sr	Sr^2	BF
Immersion	.30	.092	19.81	.14	.020	1.71	.22	.050	2.22
MEd	.09	.008	0.39	.05	.003	0.46	--		
TB-F				.15	.022	0.93	--		
TB-B				.17	.029	2.22	.24	.057	3.14

Note. Sr = semipartial correlation. Sr^2 = semipartial correlation squared. BF = Bayes factor. MEd = Mother's Education. TB-F = Touch Base Forward. TB-B = Touch Base Backward.

Item-Level Analyses

Mean group comparisons of PAE (percentage absolute error) on each of the 26 numbers presented on the number line estimation task revealed significant differences with moderate to strong evidence between children attending English instruction classes and those attending

Mandarin immersion classes on only a few numbers: 8, 21, 24, 25, and 29. Table 4.12 displays the comparisons, effect sizes (Cohen's d), and also Bayes factors (BFs) associated with the comparisons. For numbers 18, 33, and 48, children in the Mandarin immersion classes outperformed those in the English instruction classes significantly (*p*s < .05), but Bayes factors revealed only anecdotal evidence for this group difference.

Table 4.12
Group Comparison for Number Line PAE at the Item Level

Number	EE			EC			t-test	p-value	Cohen's d	BF
	n	M	SD	n	M	SD				
3	32	0.13	0.15	28	0.08	0.16	1.29	0.20	0.33	0.52
4	33	0.15	0.13	28	0.11	0.15	1.21	0.23	0.31	0.48
6	32	0.22	0.16	28	0.20	0.25	0.41	0.69	0.11	0.28
8	32	0.23	0.14	28	0.13	0.12	2.91	0.01	0.75	8.09
12	32	0.27	0.15	28	0.20	0.18	1.53	0.13	0.40	0.70
14	32	0.27	0.17	28	0.19	0.14	1.96	0.06	0.51	1.29
17	32	0.31	0.17	28	0.24	0.20	1.54	0.13	0.40	0.70
18	32	0.30	0.16	28	0.21	0.17	2.13	0.04	0.55	1.70
21	32	0.26	0.19	28	0.15	0.11	2.75	0.01	0.71	5.61
24	32	0.33	0.19	28	0.15	0.13	4.35	< .001	1.13	366.08
25	32	0.27	0.17	28	0.16	0.13	2.93	0.01	0.76	8.33
29	33	0.29	0.17	28	0.13	0.12	3.97	< .001	1.02	124.12
33	32	0.22	0.17	28	0.14	0.11	2.09	0.04	0.54	1.60
39	33	0.22	0.14	28	0.17	0.13	1.69	0.10	0.43	0.85
42	32	0.20	0.16	28	0.15	0.12	1.47	0.15	0.38	0.65
48	32	0.16	0.10	28	0.11	0.09	2.35	0.02	0.61	2.54
52	33	0.12	0.11	28	0.12	0.10	0.09	0.93	0.02	0.26
57	32	0.12	0.08	28	0.11	0.08	0.90	0.37	0.23	0.37
61	32	0.12	0.10	28	0.12	0.09	0.27	0.79	0.07	0.27
64	32	0.11	0.09	28	0.14	0.11	-1.13	0.26	-0.29	0.45
72	32	0.11	0.11	28	0.17	0.16	-1.61	0.11	-0.42	0.77
79	32	0.13	0.11	28	0.14	0.11	-0.32	0.75	-0.08	0.27
81	32	0.19	0.14	28	0.16	0.10	0.96	0.34	0.25	0.39
84	32	0.19	0.12	28	0.17	0.13	0.51	0.61	0.13	0.29
90	32	0.15	0.16	28	0.14	0.11	0.31	0.76	0.08	0.27
96	33	0.24	0.21	28	0.19	0.17	0.96	0.34	0.25	0.39

Note. PAE = Percentage Absolute Error. EE = English-speaking children in English instruction classes. EC = English-speaking children in Mandarin Chinese instruction classes.

Discussion

The aim of this study is to examine whether attending a Mandarin immersion program and presumably learning the base-10 transparent number system in Mandarin Chinese are associated with higher performance on various math skills. Across the four mathematical tasks, evidence for the Mandarin immersion program effect is found in number line estimation, magnitude comparison, and counting. The evidence is inconclusive whether Mandarin immersion program has any effect for number identification. In this section, I discuss the findings from the task with the weakest support to the one with the strongest.

Number Identification

Although enrolling in the Mandarin immersion program and presumably learning Mandarin Chinese, a language containing base-10 transparent number system, is associated with higher score on the number identification task, the evidence is only moderate (BF = 3.62) even when excluding other covariates. In fact, when other predictors are included, the null hypothesis that immersion program is not predictive of number identification performance is 1.14 times more likely given our data. Even under the best model that includes the immersion program as a predictor (BF = 6.14), it is only 1.16 times more likely that the immersion program is associated with children's performance on number identification. Given the weak results from Bayes analysis, it is inconclusive whether learning a base-10 transparent language is associated with number identification. This finding aligns with previous conflicting studies on children in Wales attending Welsh-medium instruction classes (Dowker & Roberts, 2015; Dowker et al. 2008). The number system in Welsh is also base-10 transparent. Yet, the two studies also showed contrasting results when it comes to number identification, or reading numbers. Combining children from both English-speaking and Welsh-speaking families, children in the Welsh-medium schools did not do better in reading single-, double-, and triple-digit numbers than the children in English-medium schools after controlling for age (Dowker & Roberts, 2015). In contrast, English-speaking children attending Welsh-medium schools outperformed English-medium school children in reading double-digit numbers (Dowker et al., 2008). It is likely that the conflicting results could be due to children's overall high performance in number identification. In the current study, both children attending Mandarin immersion program and those English instruction classes, performance on number identification were skewed to the left, revealing a near ceiling effect. The finding from the present study should, however, be taken with a grain of salt, as only children's number identification in *English* was used in the analyses to facilitate the comparison to the English instruction group and to avoid confounding their nondominant Chinese language skills with their math ability.

Number Line Estimation

Kindergartners' number line estimation skill was associated with their instruction type even after controlling for maternal education. The result suggests that an education system that teaches base-10 transparent number system may help facilitate numerical knowledge development, as performance on the number line estimation task is mutually predictive with math achievement (Friso-van den Bos et al., 2015). Previous studies examining language system and education system have shown more transparent number systems to be associated with more accurate estimation on the number line (Helmreich et al., 2011; Laski & Yu, 2014). For example, Laski and Yu (2014) have focused on comparing Chinese children with Chinese American children. Despite the similar linguistic and cultural background, the Chinese American children received regular English instruction, and their numerical knowledge as measured by the number line estimation task was comparable to that of Chinese children two years younger (Laski & Yu, 2014). Our current study supplements the previous study by looking on the flip side: how numerical knowledge development of children from English-speaking background may be affected by attending a program where they learn in Mandarin Chinese. The advantage of the transparent counting system appears to hold even for new learners who speak English at home.

The inclusion of children's working memory performance tells a different story. Being able to track item placement in a forward fashion is predictive of children's number line estimation. The finding is in line with previous studies that demonstrate executive function to explain unique variance in number line estimation skills (Friso-van den Bos, Kolkman, Kroesbergen, & Leseman, 2014; Friso-van den Bos et al., 2013; Kroesbergen et al., 2009). Specifically, visuospatial working memory skill (Odd One Out) significantly predicts the shape of the mental number line fit (Friso-van den Bos et al., 2014). In our study, with visuospatial working memory taken into account, being in the Mandarin immersion program and learning Chinese is no longer significantly associated with performance on the number line task. However, as learning a second language can have some influence over children's executive function (Hansen et al., 2016; Poarch & van Hell, 2012), the effect of one (executive function) versus the other (immersion program) may not be as clean cut. For example, investigating potential differences in executive function (as measured by the Simon task) among four groups of German children—monolinguals, second language learners, bilinguals, and trilinguals, with the three non-monolingual groups all attending German-English immersion classes—researchers found that bilinguals and trilingual students who have been immersed in the dual language program for longer are better able to resolve conflicts (i.e., the Simon effect of the difference between congruent and incongruent trials was smaller) than their monolingual peers. The second language learners who have spent shorter periods of time in the language immersion program did not show any significant difference with the other groups (Poarch & van Hell, 2012). It is therefore possible that being enrolled in the Mandarin immersion program for almost a year has had some executive function effect on the children in our study. However, as the current study is cross-sectional and cannot adequately examine mediation of language learning through executive function, it is premature to pinpoint the mechanism of any potential effect (if any) of being enrolled in Mandarin immersion classes.

Magnitude Comparison

Just like the Welsh-speaking children in Dowker et al.'s (2008) study, children learning the base-10 transparent number system in Chinese are better able to compare two-digit numbers, and this effect holds even after taking maternal education into account. Unlike the previous two tasks, magnitude comparison showcases the importance of both immersion program and visuospatial working memory in the forward condition. Though evidence for the unique effect of immersion program is only anecdotal after controlling for visuospatial working memory, the model containing the two predictors was six times more likely than the full model including mother's education level and backward visuospatial working memory. This finding aligns with previous research demonstrating that performance on both linguistic and spatial working memory skills are associated with symbolic quantity measures including number comparison (Cirino, 2011). The importance of executive function on magnitude comparison is hardly surprising, as previous studies have highlighted the critical role of verbal/visuospatial updating and attention skills in not just magnitude comparison, but later math abilities (Cirino, 2011; Kolkman et al., 2013; Martin, Cirino, Sharp, & Barnes, 2014). Our current finding, however, is in contrast with LeFevre and colleagues' (2010) study, where spatial attention as measured by the Corsi block tapping task does not uniquely explain variance in number comparison, but language skill does. This discrepancy could be related to our use of the immersion program as the language predictor, as our two groups of students do not differ in their vocabulary skills.

In short, despite the expected effect of visuospatial working memory (or spatial attention), evidence, albeit weak or anecdotal, still remains that learning a base-10 transparent number system in Mandarin immersion classes is associated with better performance on magnitude comparison. Whether the effect of the immersion classes is due to the base-10 transparency of the Chinese language or potential classroom level differences (e.g., curriculum, activities) is to be further explored.

Counting

The strongest evidence for the positive effect of attending a Mandarin immersion program and learning a base-10 transparent number system appears in counting. The best model shows decisive evidence for the combined role of immersion program and performance on the backward version of the visuospatial working memory task. However, the Bayes factors indicate only anecdotal or moderate evidence for each of the individual predictors. As individual predictors' BFs are akin to semipartial correlations, this finding could indicate substantial overlap between kindergartners' visuospatial working memory skill (backwards) and the type of instruction they receive. Consider the Venn diagram in Figure 4.1. The two predictors overlap substantially with the outcome variable (counting), illustrating the decisive evidence for the model. Yet, each predictor only has a smidge of unique overlap with the outcome variable, illustrating their resulting anecdotal or moderate evidence (based on BF).

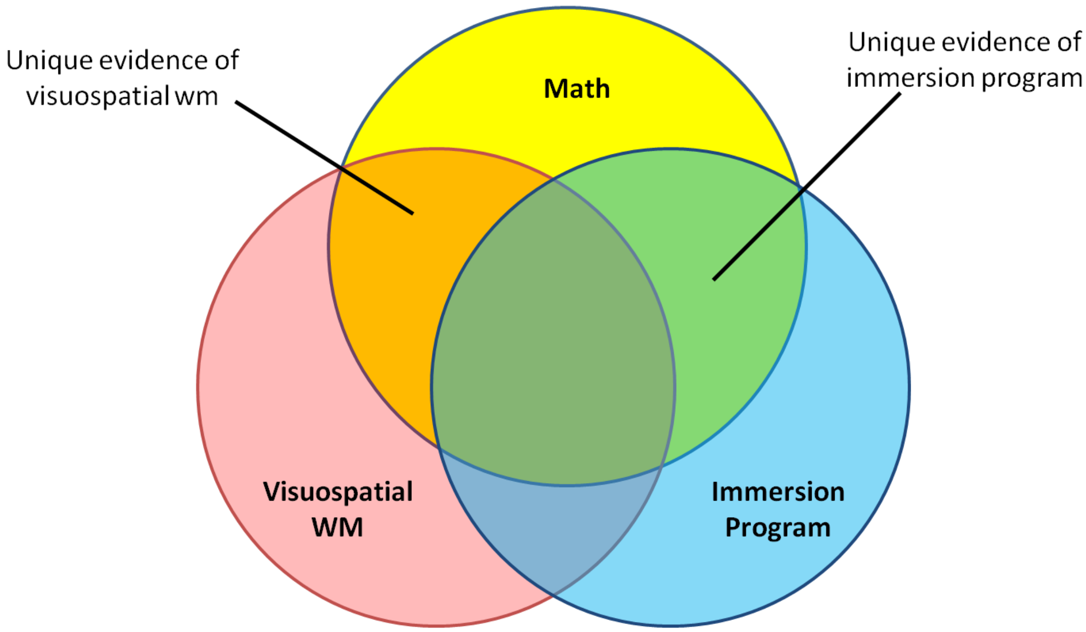


Figure 4.1. Explaining the regression Bayes factors
Combined (the darkest portion where the three circles overlap as well as the orange portion—unique evidence of visuospatial working memory—and the green section—the unique evidence of immersion program), there is decisive evidence that the Mandarin immersion program and visuospatial working memory are positively predictive of counting abilities. However, there is only moderate or anecdotal evidence for the individual predictor's *unique* roles (orange portion for visuospatial working memory, and green portion for immersion program).

Because children from the immersion program are new learners of Chinese, one explanation for the potential overlap is that in learning another language and taking the first steps toward bilingualism, these children's executive functions are also growing (Hansen et al., 2016; Poarch et al., 2012). However, as this study was cross-sectional and the EF ability was measured at the same time as their counting ability, it is difficult to tease apart whether it was indeed the attendance of the language immersion program that contributed to the growth of executive function or if the children just came to the program with higher working memory abilities.

The alternative explanation that the immersion program students came to kindergarten with higher working memory abilities is not implausible, as these children came from families with higher SES as measured by maternal education and the school is located in a more affluent neighborhood. Studies examining the effect of socioeconomic background on executive function have shown children from higher SES families to demonstrate higher executive function skills, and these findings have been shown with different measures of SES (e.g., maternal education, income) and executive function (e.g., inhibitory control, working memory, cognitive flexibility) (e.g., Blair et al., 2011; Noble, McCandliss, & Farah, 2007; Ursache & Noble, 2016). For example, parent education and income independently predict children's performance on measures list sorting test (a test of working memory), dimensional change card sort task (measure for cognitive flexibility), and flanker task (a test of inhibitory control) after controlling for demographic variables such as age and ethnicity (Urasache & Noble, 2016). It is, therefore, premature to rule out the possibility that any EF difference may be attributed to SES differences. It should nonetheless be noted that the effect of the immersion program on *all four math skills* in the current study withstood the inclusion of the maternal education variable, and that maternal education was not a significant predictor on any models.

Effect of Base-10 Transparency?

The item-level analyses on the number line task revealed stark group contrasts in the two groups' performances on the number 8 and numbers in the 20s. It should also be noted that there were significant group differences (though the evidence was only anecdotal) in the two numbers surrounding the 20s decade (18 and 33). The better performance of children in the Mandarin immersion program on this range of numbers indicates that they may have a more linear representation of the number line rather than a logarithmic one.

Despite a potential shift in the shape of the mental number line representation, the lack of differences in the 10s decade (with the exception of 18) and the significant difference in the unit digit 8 suggest that the group difference might *not* be due to base-10 transparency of Mandarin Chinese. However, the lack of findings in the 10s could be due to instructional approaches. The standards for kindergarten math curriculum include counting up to 20, and all teachers—based on researcher observations—have incorporated practices in their classrooms to ensure this standard has been met. Moreover, the significant and moderate to strong evidence for differences in the 20s decade—beyond the required standards of kindergarten curriculum—could possibly indicate that children in the Mandarin immersion program were able to better extrapolate the meaning of the numbers beyond 20 due to the regularity of the numbers. (Having this understanding of the 20s decade perhaps also shifted these children to a more linear representation of the number line.) An alternate, additive explanation is that the teachers in the Mandarin immersion program incorporated numbers beyond 20 in their classroom activities. This practice was indeed observed inside their classrooms. It is, therefore, inconclusive whether the differences reflect classroom teaching or the extrapolation of number meanings due to the regularity of the base-10 transparent number system.

Regardless, attending a Mandarin immersion program and presumably learning a base-10 transparent number system does seem to have a positive influence on children's ability to estimate number placements on a number line and count to 120, even in English. The heightened counting performance could be related to the regular number system in Chinese that makes it easier for the teachers to incorporate them in practice. In my observations of the immersion program classrooms, the teachers regularly incorporated counting up to 100 in Chinese, despite the kindergarten curriculum of just counting up to 20. For example, both classrooms had point systems in place to encourage good behavior. At least once a month, the teachers would count the marbles earned by each table with the children to see who the winners were. The kindergartners chimed in enthusiastically in these counting exercises. When I spoke with one of the teachers about how she managed to get the kids to count that high, she responded, "I think the regularity helps" before proceeding on to demonstrate the pattern and rhythmicity of counting in Chinese. It is therefore plausible that the effects on any mathematics skills (even magnitude comparison) could be attributed to classroom practice differences rather than the base-10 transparency of Chinese. Yet, if we were to take the teachers' explanation into account, it was the base-10 transparency that allowed them to incorporate more number words in their practice. Thus, even if it was the classroom teaching that has the direct effect, the base-10 transparency of the number words has an indirect influence by affecting teachers' decisions.

While these classroom counting exercises were conducted in Chinese, the counting effect was demonstrated in English. This peculiar finding could be due to 1) spillover effects from Chinese to English, or 2) home environment where the kindergartners counted more at home with their parents. The latter explanation is plausible, considering that the Mandarin immersion program was located in a more affluent neighborhood. However, survey responses from the parents indicated no difference in the amount of time the parents from the two groups spent counting with their children. In fact, parents from regular instruction schools reported slightly higher frequency of counting with their children at home. The remaining explanation is then children managed to learn the numbers in Chinese and became curious about their corresponding words in English, and they could have then learned these English number words from teachers at school (20% of the instruction is meant to be in English) or parents at home.

General Discussion

To recap, we found strongest evidence for the effect of learning math in a Mandarin immersion setting in counting, followed by magnitude comparison and number line estimation. The strength of the evidence corresponds with the developmental trajectory of numerical knowledge. Past research have shown children's counting abilities to be predictive of their ability to place numbers on a number line; children who are able to count to higher numbers display a more linear pattern—as opposed to logarithmic pattern—for the number line task (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). This linearity of mental representation is in turn found to be predictive of number comparison and math achievement (Laski & Siegler, 2007), though the relationship is not found in all studies (e.g., LeFevre et al., 2013). Our results thus suggest that the base-10 transparent number system in Chinese might have indirectly helped the children in the Mandarin immersion program acquire the ability to count to a higher number sooner, and the effect becomes smaller as it trickles down the path of the numerical knowledge development.

Limitations and Future Directions

This interpretation is, of course, with some reservations. First, the current study is not longitudinal. It is therefore difficult to directly examine the developmental trajectory of numerical knowledge in this sample, though it is likely that any effect of learning another language on later math skills would come through the ability to count. It is also difficult to disentangle the roles of executive function, second language learning, and base-10 transparent number system. Future studies should examine children's early math and working memory capacity at the start of enrolling in a Mandarin immersion program to better investigate potential moderation and mediation of executive function on mathematics later in the year. Furthermore, to better tease apart the contributions of language learning from that of learning a language that has a base-10 transparent number system, more items that could examine base-10 effects (e.g., performance on estimating numbers on a number line between 100 and 200) should be included and expansion to language immersion programs with languages with less transparent number system would be ideal.

Second, even though special care was taken to limit our study sample to children who only hear English at home and control for socioeconomic status (SES), we were only able to use maternal education as a proxy for SES and were unable to control for potential selection bias. We were therefore unable to determine whether any effect could be due to potential income-level differences between the two groups or other motivational or unobserved differences in the parents who would want their kids to be enrolled in a Mandarin immersion program. As discussed previously, multiple studies have found family SES to be predictive of children's executive function (e.g., Noble et al., 2007), math (e.g., Lawson & Farah, 2017; Sarnecka et al., 2017), or language skills (e.g., Hart & Risley, 1995; Noble et al., 2007). Some have suggested that the effect of SES could have manifested through the activities and the resources available in the home (e.g., Hackman, Farah, & Meaney, 2010; Ramani et al., 2015). However, the alternative explanation that the Mandarin immersion program effect seen in the current study may be due to SES differences or selection bias is not very likely for several reasons. First of all, we did not find any significant vocabulary-level differences between the two groups. Granted, this lack of language differences could be the result of slightly different measurement or task administration methods. For the receptive picture vocabulary test, children from the English instruction classes received the digitized version that played the English audio first with the option of playing the Spanish audio, whereas children from the Mandarin immersion classes received the version that played Mandarin first with the option of playing the English audio. On the expressive test, children in the Mandarin immersion classes were asked the questions in Mandarin first and only received the questions in English if they were unable to answer in Mandarin. However, beyond these differences, the tests were extremely similar such that the lack of significance in vocabulary-level differences should still hold. Second, self-report of counting activities at home between the two groups did not differ significantly. Third, maternal education contributed little to no unique variance in math performance across the measures. Finally, parents selecting a Mandarin immersion program likely selected it for language learning reasons (e.g., learning Mandarin Chinese so their children can speak with their grandparents) rather than math learning reasons. Nevertheless, future studies should expand the current research to include more diverse student populations and more demographic measures. Qualitative interviews with the parents may also help decipher selection reasons that may or may not affect children's math achievement.

Despite these limitations, the current study is the first to demonstrate a potential advantage in mathematics of a base-10 transparent language immersion program in children from a very different linguistic (and, in some cases, cultural) background.

CHAPTER 5

Summary and Conclusion

There is growing recognition that language and executive function may independently affect numerical cognition. This dissertation aims to contribute to this scientific conversation by expanding upon what is meant by "language" and the approaches of exploring its relation to mathematics, namely, by considering executive function simultaneously. In two studies, I examined the relationship between language and math while considering executive function by operationalizing language first as a function of kindergartners' language proficiency and, second, as a function of the lexical property of the number system in the target language of instruction. Below, I summarize the findings from the key studies and the major takeaways that emerged from this work. I conclude with new directions of research stemming from this dissertation.

Summary of Findings

Chapter 2 introduced the 11 measures—three for language, four for executive function, and four for mathematics—used throughout the studies in this dissertation and examined the psychometric properties of the measures. Cronbach's alpha indicated that all measures were adequately reliable. Correlational and factor analyses ascertained that the measures achieved concurrent validity for children attending typical English instruction classes. However, Following Instructions, a verbal working memory task, proved to be better aligned with the language construct for children attending Mandarin immersion classes. Moreover, the measures showed a distinct structural pattern between children in different language instruction types, such that the language and executive functions barely relate intra- or inter-constructs for children in the Mandarin immersion classes. The distinct structural pattern and lack of expected correlations could have resulted from lack of sampling adequacy. The alignment of Following Instructions with the linguistic construct might be due to the kindergartners' lack of proficiency in Chinese. Considering that many learned the Chinese terms of the objects (e.g., cup, box) for the first time at the start of testing, it goes to reason that the test picked up their ability to learn new vocabulary terms rather than their working memory. As such, Study 1 included only children from typical English instruction classes and Study 2 examined each of the math measures separately.

Study 1

In Study 1, I investigated the relationship between language, executive function, and mathematics using structural equation modeling and latent variables for the three constructs. The language construct in this study is defined as kindergartners' proficiency as measured by objective vocabulary tests and parent reports. With this approach, I showed that the influence of language on mathematics may not be as straightforward as previous research of linguistic influence on math suggests (see Dowker & Nuerk, 2017). I showed that language has a direct effect on executive function, and that executive function, in turn, has a direct effect on overall mathematics such that the effect of language proficiency on mathematics appeared to be at least partially mediated through its effect on executive function. This finding suggests that when considering the relation of two specific cognitive domains, the general cognitive domain (i.e., executive function) cannot be neglected. There have been previous attempts at examining components of all three constructs together. For example, some researchers considered verbal working memory as a type of language-related construct and investigated its role on multiplication performance in relation to the role spatial working memory plays (Soltanlou, Pixner, & Nuerk, 2015). Yet in such studies, language and executive function are intertwined. Study 1 takes this line of investigation one step further by teasing apart all three constructs and revealing the complexity when all three are examined side by side. That is, language proficiency as a whole has effects on mathematics performance, and these effects could have been mediated via its effect on executive function skills. Nevertheless, alternate, plausible models should also be considered. For example, the three constructs may only be interrelated due to a fourth, higher-level construct of general mental ability. This alternative model may be plausible, but does not seem to fit our data very well, as the model only converged using one program possibly because of overspecification with an additional construct. Despite its convergence issues and despite that this alternate model with yet another construct may be less satisfying in terms of the possibility of generating educational interventions, future studies should consider study designs that could test the differences between these potential alternatives with larger, more diverse samples. Additionally, longitudinal designs should be incorporated to tease apart directionality and causation.

Study 2

In Study 2, the perspective taken for "language" shifts to a broader lexical context: how learning a language with a base-10 transparent number system may or may not influence kindergartners' different mathematics performance. Using multiple regressions supplemented by Bayesian analyses, I demonstrated that being exposed to the base-10 transparent Chinese number system in Mandarin immersion classes is associated with higher performance on all four math tasks of counting, identifying numbers, comparing magnitude, and estimating the position of a number on a number line. The strength of the effect on various tasks differs once children's executive function skills are taken into account, with being in the immersion program having the strongest effect with counting followed by magnitude comparison and number line estimation. Socioeconomic status as measured by maternal education does not have an effect on children's math performance. The current dataset offers only inconclusive evidence on whether the differences between children attending English instruction classes with those attending Mandarin immersion classes might be due to learning a different language or learning a different language *that is base-10 transparent*. Nevertheless, the positive finding may be attributed to the way the Mandarin immersion teachers conduct their classroom activities, which might have been affected by the regularity and base-10 transparency of the Chinese number system.

Key Lessons or Themes

(Emergent) Bilingualism and Math

In addition to cognitive domains such as executive function and language, research in bilingualism has recently made the foray into how bilingualism may affect numerical cognition. As such, there is yet no definitive conclusion regarding the role bilingualism plays in numerical cognition. For example, studies with preschoolers attending Spanish-English dual-language immersion program demonstrated that young dual-language learners may be trailing behind their monolingual peers in mathematics due to SES differences, and that among high-SES groups, there was no difference in number knowledge between monolingual and bilingual groups (Sarnecka, Negen, & Goldman, 2018). In contrast, studies with large datasets (e.g., ECLS-K: 2011) suggest that bilingualism positively influences math achievement after controlling for age, sex, SES, and language proficiency (Hartanto, Yang, & Yang, 2018).

The studies in this dissertation contribute to this conversation by connecting language—whether it is proficiency or learning—to executive function and math among emergent bilingual students. Study 1 affirmed that proficiency in the language of instruction is an important predictor of both executive function and mathematics among monolingual and emergent bilingual kindergartners. However, due to sample size limitations, a direct model comparison of the emergent bilingual kindergartners with their monolingual peers was not possible. It is nonetheless important to note that this study was conducted with a sample majority of emergent bilingual students. It is reasonable, then, that the overall "big picture" of how the three constructs interrelate—that language proficiency, as a whole, has direct effects on executive function (as constructed by multiple measures of working memory), which has direct effects on overall math performance, and that the linguistic effect on math may be mediated via executive function skills—can be applied to emergent bilingual kindergartners.

Study 2 echoed the importance of both linguistic and executive function influences on mathematics among a different population of emergent bilingual kindergartners. Furthermore, Study 2 was able to provide a direct contrast by comparing only children from monolingual English-speaking households. The findings suggest shared variance between learning a new language in the immersion program and children's executive function skills, but also unique positive contributions of learning Mandarin Chinese in the immersion setting. In other words, Study 2 suggests that *becoming* an emergent bilingual by being exposed to the Chinese language in a language immersion setting may have effects on children's numerical cognition.

Concerning Language of Assessments

Another key lesson from this dissertation work is that assessment tasks do not always work the same way for different populations. As shown in Chapter 2, the 11 measures composed distinct patterns and structures for children in mainstream typical English instruction classes and those enrolled in Mandarin immersion classes. Previous studies have shown that tests sometimes have differential functions for English learners (e.g., Abedi & Lord, 2001). However, the reverse appears to be true in this dissertation. The majority of children attending typical English instruction classes in the study sample are English learners and from lower SES families, whereas the ones attending Mandarin immersion classes are mostly native English speakers and from higher SES families. Yet, we found the expected relationship between the measures for children attending the English instruction classes, but not for the ones attending Mandarin immersion classes. The different patterns could possibly be attributed to power issue associated with smaller sample size, as the low values from the KMO measure of sampling adequacy indicated that the data from the Mandarin immersion classes might not be suited for factor analysis. Additionally, as noted earlier, the Following Instructions test might have picked up children's ability to quickly learn new words.

This key discrepancy in the verbal working memory measure calls into question the role the language of assessment plays in estimating children's skills in different cognitive domains. Having tested early numeracy skills in emergent bilingual children in both their languages (i.e., Spanish and English), researchers have found no difference between children's performances when tested in the two languages of assessment (Sarnecka et al., 2008). The finding led to the recommendation that due to limited researching resources, it should suffice to test children in only their language of instruction. While the recommendation is sensible for content areas such as math where children are actively acquiring new concepts in classes (and hence in the language of instruction), the same cannot be said of domain-general skills that might not be specifically taught in school. It may be reasonable, then, for researchers to consider emergent bilingual children's dominant language when attempting to assess their domain-general executive function skills.

Future Directions

Language or Classroom Environment?

The work presented in this dissertation suffers from the caveat of limited samples. For example, Study 2 compared new learners of Mandarin Chinese with their monolingual peers. As it stands, there is no information on whether the emergent bilingual children's higher performance should be attributed to the fact that they were actively learning a second language or that the target language of instruction was one with a base-10 transparent number system. Future study should therefore expand to include children learning non-base-10 transparent languages to tease apart the contributions of language learning from that of distinct lexical features. Expanding the research to more language immersion programs would also be beneficial in teasing apart the contribution of languages (whether it is language learning or the lexical features of the target language) from that of the classroom environment. Research presented in Chapter 4 (Study 2) was conducted at two Mandarin immersion classes and four mainstream typical English instruction classes. With such a small sample of classrooms, it was difficult to disentangle (base-10 transparent) language learning effects from that of the classroom learning context.

To be more specific, although this dissertation work provided one of the first looks into new learners in Mandarin immersion classes on non-linguistic and executive function domains, the scope of this dissertation, with its focus on language, math, and executive function skills, also meant a somewhat limited perspective at the learning that took place in such dual-language immersion programs. Future work on student cognition should therefore consider mixed methods design by incorporating qualitative methods such as observations and interviews to examine the learning context, as classroom environment may differ drastically from one room to another within the same school. These work would then be able to provide more detailed accounts to *how* the learning took place. For example, what strategies did the teachers incorporate? Which languages did the teachers use in their classrooms to convey different points? Did language switching occur? If so, under what circumstances? How often did number-related activities occur beyond the math segments of class? Asking for and providing a qualitative overview of the classrooms may shed light into who these learners are and how they might have been able to acquire mathematical concepts in a new language.

Furthermore, to combat potential selection bias that the parents who choose to place their children into immersion programs may be very different from other parents, baseline performance of the participants should be established at the time of kindergarten entry or even earlier. The incorporation of a longitudinal design, along with multiple time points for classroom observations, can serve to make research studies investigating these special, niche student populations much stronger. The use of qualitative methods, such as parent surveys or interviews, may also supplement this line of research by identifying parents' rationale for selecting into these language immersion programs.

Better Integration of the Three Constructs

This dissertation has demonstrated the interrelationship between language, executive function, and mathematics among emergent bilingual kindergartners and highlighted the importance of examining not only domain-general skills in math performance, but also that of different types of language skills. The field of cognition, especially numerical cognition, has started to investigate the interplay of components within the three constructs. For instance, using multiple measures of executive function (e.g., listening recall for verbal working memory, dimensional card sorting for cognitive flexibility) and language (e.g., phonological awareness, print knowledge, and even a researcher-developed "mathematical language" measure) and employing regression tree analyses, Purpura, Day, Napoli, and Hart (2017) have been able to suggest the skill or component most predictive of math performance. While the movement toward interdisciplinary exploration by taking into account multiple skill sets from multiple domains has been able to generate more insights into human cognition, it can still greatly benefit from a broader, potentially unifying theory or model connecting all three constructs.

One approach toward connecting the three constructs conceptually may involve integrating more models and theories from non-numerical, and perhaps linguistic framework. Dehaene's triple-code model (Dehaene, 2011; Dehaene & Cohen, 1997) connects language (verbal code and Arabic code) with mathematics, though the roles allocated to "language" was somewhat restricted (e.g., phonological processing, reading Arabic numerals). There is more to be gained by integrating insights from linguistic research to the field (Brysbaert, 2018). One such example at a more specific, analytical level is to consider the semantic distance between numbers (Brysbaert, 2018). Such incorporation of insights from semantic network analyses can potentially offer explanations for why certain numbers may be more difficult in addition to even more intimately connect language research with mathematics research.

Yet, on an even broader scale, combinations of psycholinguistic *theories*, particularly those pertaining to bilingual population, may be able to push the field to further delineate whether the relations among language, executive function, and mathematics may vary according to individuals' linguistic experiences. As presented in Chapter 1 Figure 1.1, Paivio's bilingual dual coding theory (2014) could be integrated with Dehaene's triple-code model (Dehaene, 2011; Dehaene & Cohen, 1997). The languages of bi- or multilingual individuals could mean different ways of storing or accessing semantic information (e.g., imagens). How, for example, would the potentially different linguistic processing affect math processing (the link between the new "verbal" code with that of quantity code) in emergent bilinguals? Is it possible to draw only upon the "nonverbal" semantic imagens for bilingual individuals without activating either L1 or L2 logogens when processing mathematics? Although this dissertation touched upon emergent bilinguals' performance in language and mathematics, it has not been able to test out any links in the integrated model (see Figure 1.1). Future research should attempt to test out theories specifically considering (emergent) bilingual processing of both language and mathematics.

Furthermore, with the exception of LeFevre's Pathways to Mathematics (2010) model, few theoretical models have tried to integrate any part of executive function with language and mathematics with the three constructs being placed at the *same level*. There are admittedly conceptual models that connect executive function and mathematics while including some part of language. For example, Geary and Hoard's (2005) conceptual framework for approaching the study of mathematical disabilities incorporates Baddeley's (1992) model of working memory and a large part of the triple-code model of numerical processing. However, as such, "language" becomes subsumed under executive function and is considered as a support system the same way the "visuospatial system" is. Under this framework, language is narrowly responsible for presenting and manipulating verbal information, such as in retrieving arithmetic facts (similar to triple-code model's verbal code). A model that would place the language construct at the same level of relevance as executive function or mathematics construct may be beneficial in helping researchers, scholars, and educators realize the significant role it plays and further consider how the three interrelate. As many recent studies (e.g., see special issue edited by Dowker & Nuerk, 2016) and this dissertation have demonstrated, the role of language in mathematics goes beyond that of fact retrieval and can be considered as distinct from the role of executive function. A more integrative synthesis of theories and models concerning all three constructs should therefore be further explored in future research to uncover the underlying mechanisms through which the constructs influence each other.

REFERENCES

Abedi, J., & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teacher College Record, 112*(3), 723–746.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234. http://doi.org/10.1207/S15324818AME1403_2

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology*, *23*, 417–426. http://doi.org/10.1348/026151005X26804

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716. http://doi.org/10.1111/j.1467-8624.2006.00968.x

Alloway, T. P., & Ledwon, F. (2014). Working memory and sentence recall in children. *International Journal of Educational Research*, *65*, 1–8. http://doi.org/10.1016/j.ijer.2013.12.001

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137. http://doi.org/10.1016/j.lindif.2010.09.013

Baddeley, A. (1992). Working Memory. *Science*, *255*, 556–559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*. http://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173. http://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61*(3), 438–456. http://doi.org/10.1016/j.jml.2009.05.004

Baldo, J., & Dronkers, N. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia*, *45*(2), 229-235. doi: 10.1016/j.neuropsychologia.2006.07.014

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Developement*, *83*(2), 413–422. http://doi.org/10.1111/j.1467-8624.2011.01707.x.Bilingual

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism, 10*(3), 201–223. http://doi.org/10.2167/beb441.0

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*(2), 397–401. http://doi.org/10.1017/S136672891100040X

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading, 9*(1), 43–61. http://doi.org/10.1207/s1532799xssr0901_4

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24*, 27–44. http://doi.org/10.1017.S014271640300002X

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition, 112*(3), 494–500. http://doi.org/10.1016/j.cognition.2009.06.014

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., Kivlighan, K. T., Fortunato, C. K., & the FLP Investigators. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development, 82*(6), 1970-1984. doi: 10.1111/j.1467-8624.2011.01643.x

Booth, J., & Siegler, R. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/42/1/189/

Brunetti, R., DelGatto, C., & Delogu, F. (2014). eCorsi: Implementation and testing of the Corsi block-tapping task for digital tablets. *Frontiers in Psychology, 5*(AUG), 1–8. http://doi.org/10.3389/fpsyg.2014.00939

Brysbaert, M. (2018). Numbers and language: What's new in the past 25 years? In A. Henik &W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition*. Amsterdam: Elsevier.

Bull, R., Espy, K. A., & Wiebe, S. a. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228. http://doi.org/10.1080/87565640801982312

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36–41. http://doi.org/10.1111/cdep.12059

Bunch, G. C. (2013). Pedagogical language knowledge: Preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education, 37*, 298–341. http://doi.org/10.3102/0091732X12461772

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences, 14*(12), 534-541. doi: 10.1016/j.tics.2010.09.007

California Proposition 58, Non-English Languages Allowed in Public Education (2016). Retrieved December1, 2016, from https://ballotpedia.org/California_Proposition_58,_Non-English_Languages_Allowed_in_Public_Education_(2016)

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. http://doi.org/10.1037/h0046016

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2008). Beyond the number domain. *Trends in Cognitive Sciences, 13*(2), 83-91. doi: 10.1016/j.tics.2008.11.007

Case, R., Kurland, D. M., &Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*(3), 386–404. http://doi.org/10.1016/0022-0965(82)90054-6

Chinello, A., Cattani, V., Bonfiglioli, C., Dehaene, S., & Piazza, M. (2013). Objects, numbers, fingers, space: clustering of ventral and dorsal functions in young children and adults. *Developmental Science, 16*(3), 377-393. doi: 10.1111/desc.12028

Cirino, P. T. (2011). The interrelationships of mathematical precursors in kindergarten. *Journal of Experimental Child Psychology, 108*(4), 713–33. http://doi.org/10.1016/j.jecp.2010.11.004

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662–677. http://doi.org/10.1111/j.1467-8624.2012.01854.x

Clements, D. H., & Sarama, J. (2013). Rethinking early mathematics: What is research-based curriculum for young children? In L. D.English &J. T.Mulligan (Eds.), *Reconceptualizing Early Mathematics Learning* (pp. 121–147). Springer Netherlands. http://doi.org/10.1007/978-94-007-6440-8_7

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36*, 79–90. http://doi.org/10.1016/j.ecresq.2015.12.009

Cornoldi, C., & Mammarella, I. C. (2008). A comparison of backward and forward spatial spans. *Quarterly Journal of Experimental Psychology, 61*(5), 674–682. http://doi.org/10.1080/17470210701774200

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(2), 891B.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*(2), 63–68. http://doi.org/10.1016/j.tine.2013.12.001

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. http://dx.doi.org/10.1037/h0040957

De Smedt, B. Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48-55. doi: 10.1016/j.tine.2013.06.001

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1–2), 1–42. http://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2010). The calculating brain. In D. A.Sousa (Ed.), *Mind, brain, & education: Neuroscience implications for the classroom* (pp. 179–200). Bloomington, IN: Solution Tree Press.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics* (Revised Ed). New York: Oxford University Press.

Dehaene, S. & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*(1), 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*(2), 219–250. http://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14*, 218-224. doi: 10.1016/j.conb.2004.03.008

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3-6), 487-506. doi: 10.1080/02643290244000239

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science, 284*(5416), 970-974. doi: 10.1126/science.284.5416.970

Devlin, K. (2010). The mathematical brain. In D. Sousa (Ed.), *Mind, brain, & education: Neuroscience implications for the classroom* (pp. 163-178). Bloomington, IN: Solution Tree Press.

Doabler, C. T., Clarke, B., Kosty, D. B., Baker, S. K., Smolkowski, K., & Fien, H. (2016). Effects of a core kindergarten mathematics curriculum on the mathematics achievement of Spanish-speaking English learners. School *Psychology Review, 45*(3), 343–361.

Donolato, E., Giofrè, D., & Mammarella, I. C. (2017). Differences in verbal and visuospatial forward and backward order recall: A review of the literature. *Frontiers in Psychology, 8*, 663. http://doi.org/10.3389/fpsyg.2017.00663

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. http://doi.org/10.1080/09515080802285511

Dowker, A., & Nuerk, H.-C. (2016). Editorial: Linguistic Influences on Mathematics. *Frontiers in Psychology, 7*, 1035. http://doi.org/10.3389/fpsyg.2016.01035

Dowker, A., & Roberts, M. (2015). Does the transparency of the counting system affect children's numerical abilities? *Frontiers in Psychology, 6*(July), 1–7. http://doi.org/10.3389/fpsyg.2015.00945

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99*(1), 1–17. http://doi.org/10.1016/j.jecp.2007.08.006

Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science, 23*(11), 1364–1371. http://doi.org/10.1177/0956797612443836

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. http://doi.org/10.1111/1467-8721.00160

Engle, R. W. (2010). Role of Working-Memory Capacity in Cognitive Control. *Current Anthropology, 51*(1). http://doi.org/10.1086/650572

Esposito, A. G., Baker-Ward, L., & Mueller, S. (2013). Interference suppression vs. response inhibition: An explanation for the absence of a bilingual advantage in preschoolers' Stroop task performance. *Cognitive Development, 28*(4), 354–363. http://doi.org/10.1016/j.cogdev.2013.09.002

Executive Order No. 13398, 71 Federal Register 20519 (April 21, 2006).

Fernandes, M. A., Craik, F., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian Journal of Experimental Psychology, 61*(2), 128–141. http://doi.org/10.1037/cjep2007014

Flores, A., Lopez, G., & Radford, J. (2017, September 18). Facts on U.S. Latinos, 2015: Statistical portrait of Hispanics in the United States. Retrieved from < http://www.pewhispanic.org/2017/09/18/facts-on-u-s-latinos-trend-data/>

Friso-van den Bos, I., Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2014). Explaining variability: Numerical representations in 4- to 8-year-old children. *Journal of Cognition and Development, 15*(2), 325–344. http://doi.org/10.1080/15248372.2012.742900

Friso-van den Bos, I., Kroesbergen, E. H., VanLuit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van derSchoot, M., & VanLieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 13*4, 12–29. http://doi.org/10.1016/J.JECP.2015.02.002

Friso-van den Bos, I., van derVen, S. H. G., Kroesbergen, E. H., & vanLuit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44. http://doi.org/10.1016/j.edurev.2013.05.003

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493–513. http://doi.org/10.1037/0022-0663.97.3.493

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives, 11*(3). Retrieved from: < https://epaa.asu.edu/ojs/article/view/264>

García, O. (2009). Emergent bilinguals and TESOL: What’s in a name? *TESOL Quarterly, 43*(2), 322–326. http://doi.org/http://www.jstor.org/stable/27785009

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children’s performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, *22*, 1019–1037. http://doi.org/10.1002/acp.1407

Geary, D. C., Bow-Thomas, C. C., Fan, L., & Siegler, R. S. (1993). Even before formal instruction, Chinese children outperform American children in mental addition. *Cognitive Development, 8*(4), 517–529. http://doi.org/10.1016/S0885-2014(05)80007-3

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and Arithmetical Cognition: A Longitudinal Study of Process and Concept Deficits in Children with Learning Disability. *Journal of Experimental Child Psychology, 77*(3), 236–263. http://doi.org/10.1006/jecp.2000.2561

Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D.Campbell (Ed.), *Handbook of Mathematical cognition* (pp. 253–267). New York: Psychology Press.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359.

Goo, J. (2010). Working memory and reactivity. *Language Learning, 60*(4), 712–752. http://doi.org/10.1111/j.1467-9922.2010.00573.x

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*(3), 445–474. http://doi.org/10.1017/S0272263112000149

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11*(9), 651–9. http://doi.org/10.1038/nrn2897

Hansen, L. B., Macizo, P., Duñabeitia, J. A., Saldaña, D., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2016). Emergent bilingualism and working memory development in school aged children. *Language Learning*, *66*(S2), 51–75. http://doi.org/10.1111/lang.12170

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.

Hartanto, A., Yang, H., & Yang, S. (2018). Bilingualism positively predicts mathematical competence: Evidence from two large-scale studies. *Learning and Individual Differences, 61*, 216–227. http://doi.org/10.1016/j.lindif.2017.12.007

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language Effects on Children's Nonverbal Number Line Estimations. *Journal of Cross-Cultural Psychology, 42*(4), 598–613. http://doi.org/10.1177/0022022111406026

Higo, K., Minamoto, T., Ikeda, T., & Osaka, M. (2014). Robust order representation is required for backward recall in the Corsi blocks task. *Frontiers in Psychology*, *5*, 1285. http://doi.org/10.3389/fpsyg.2014.01285

Hoff, E., Core, C., Place, S., & Rumiche, R. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*(1), 1–27. http://doi.org/10.1017/S0305000910000759

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17-29. doi: 10.1016/j.jecp.2008.04.001

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. British Journal of *Developmental Psychology, 7*(4), 377–380. http://doi.org/10.1111/j.2044-835X.1989.tb00814.x

Jacob, R., & Partkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 41*(4), 41. http://doi.org/10.3102/0034654314561338

jamovi project (2018). jamovi (Version 0.8.3) [Computer Software]. Retrieved from https://www.jamovi.org

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and nonverbal working memory and achievements on national curriculum tests at 7 and 14 years of age. *Educational and Child Psychology, 20*, 123–140.

JASP Team (2017). JASP (Version 0.8.1.2) [Computer software]. Retrieved from https://jasp-stats.org/

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*(1), 36-46. https://doi.org/10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850-867. https://doi.org/10.1037/a0014939

Jordan, N.C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*(1), 60-68. https://doi.org/10.1002/ddrr.46

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1). Retrieved from http://repository.cmu.edu/psychology

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., & Velez, E. D. (2015). The Condition of Education 2015. NCES 2015-144. National Center for Education Statistics. National Center for Education Statistics. Available from: ED Pubs. P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827; Web site: http://nces.ed.gov/.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*(1), 165-172.

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology, 7*(4), 252–258. http://doi.org/10.1207/S15324826AN0704_8

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168–1201. http://doi.org/10.3102/0034654309332490

Klessinger, N., Szczerbinski, M., & Varley, R. (2012). The role of number words: the phonological length effect in multidigit addition. *Memory & Cognition*, 1289–1302. http://doi.org/10.3758/s13421-012-0228-y

Knops, A., Dehaene, S., Berteletti, I., & Zorzi, M. (2014). Can approximate mental calculation account for operational momentum in addition and subtraction? *The Quarterly Journal of Experimental Psychology, 67*(8), 1541-56. doi: 10.1080/17470218.2014.890234.

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from the operational momentum effect. *Attention, Perception, & Psychophysics, 71*(4), 803-821. doi: 10.3758/APP.71.4.803

Kolkman, M. E., Hoijtink, H. J. A., Kroesbergen, E. H., & Leseman, P. P. M. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences, 24*, 145–151. http://doi.org/10.1016/j.lindif.2013.01.004

Kotsopoulos, D., & Lee, J. (2012). A naturalistic study of executive function and mathematical problem-solving. *Journal of Mathematical Behavior, 31*(2), 196–208. http://doi.org/10.1016/j.jmathb.2011.12.005

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103(*4), 516–531. http://doi.org/10.1016/j.jecp.2009.03.009

Kroesbergen, E. H., vanLuit, J. E. H., vanLieshout, E. C. D. M., Loosbroek, E.Van, & Van deRijt, B. A. M. (2009). Individual differences in early numeracy. *Journal of Psychoeducational Assessment, 27*(3), 226–236. http://doi.org/10.1177/0734282908330586

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science, 23*(3), 159–163. http://doi.org/10.1177/0963721414528511

Kuhl, P. K., Stevenson, J., Corrigan, N. M., van den Bosch, J. J. F., Can D. D., & Richards, T. (2016). Neurogimaging of the bilingual brain: Structural brain correlates of listening and speaking in a second language. *Brain & Language, 162*, 1-9. doi: 10.1016/j. bandl.2016.07.004

Laski, E.V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*(6), 1723–1743. http://doi.org/10.1111/j.1467-8624.2007.01087.x

Laski, E.V., & Yu, Q. (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology, 117*(1), 29–44. http://doi.org/10.1016/j.jecp.2013.08.007

LeFevre, J. A., Lira, C. J., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*(SEP). http://doi.org/10.3389/fpsyg.2013.00641

LeFevre, J., Fast, L., Smith-Chant, B. L., Skwarchuk, S.-L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics : Longitudinal predictors of performance. *Child Development, 81*(6), 1753–1767.

Li, Y., Hu, Y., Zhao, M., Wang, Y., Huang, J., & Chen, F. (2013). The neural pathway underlying a numerical working memory task in abacus-trained children and associated functional connectivity in the resting brain. *Brain Research, 1539*, 24–33. http://doi.org/10.1016/j.brainres.2013.09.030

Lin, L.-C., & Johnson, C. J. (2016). Mandarin-English bilingual vocabulary development in an English-immersion preschool: How does it compare with monolingual development? *International Journal of Bilingualism, 20*(2), 173–189. http://doi.org/10.1177/1367006914547662

Lopez, M. H., & Velasco, G. (2011). Childhood poverty among Hispanics sets record, leads nation. Washington, DC: *Pew Hispanic Center*. Retrieved from <http://www.pewhispanic.org/2011/09/28/childhood-poverty-among-hispanics-sets-record-leads-nation/>

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition, 74*(3), 347–357. http://doi.org/10.1016/j.bandc.2010.09.004

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research, 50*(4), 940. http://doi.org/10.1044/1092-4388(2007/067)

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*(5), 545–555. http://doi.org/10.1080/17405629.2016.1259614

Martin, N. A. (2013). Receptive One-Word Picture Vocabulary Test - 4: Spanish-Bilingual Edition. Novato, CA: ATP Assessments.

Martin, R. B., Cirino, P. T., Sharp, C., & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learning and Individual Differences*, *34*, 12–23. http://doi.org/10.1016/j.lindif.2014.05.006

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. http://doi.org/10.1080/10705511.2016.1186549

Messick, S. (1979). Test validity and the ethics of assessment. *ETS Research Report Series, 1979*(1), i-43. doi: 10.1002/j.2333-8504.1979.tb01178.x

Miller, K. F., Kelly, M. K., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of mathematical development. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163-178). New York: 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*(1), 56-60. doi: 10.1111/j.1467-9280.1995.tb00305.x

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions : Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. http://doi.org/10.1177/0963721411429458

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology, 114*(2), 187–202. http://doi.org/10.1016/j.jecp.2012.09.002

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism, 13*(5), 597-616. doi: 10.1080/13670050.2010.488288

National Council on Measurement in Education. (2018). Glossary of Important Assessment and Measurement Terms. Retrieved April22, 2018, from http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorV

National Governors Association Center for Best Practices, &Council of Chief State School Officers. (2010). Common Core State Standards - Mathematics. Retrieved February3, 2017, from http://www.corestandards.org/Math/

Ng, S. S. N., & Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research, 80*(2), 180-206. http://doi.org/10.3102/0034654310364764

Nicolay, A. C., & Poncelet, M. (2013). Cognitive abilities underlying second-language vocabulary acquisition in an early second-language immersion education context: A longitudinal study. *Journal of Experimental Child Psychology*, *115*(4), 655–671. http://doi.org/10.1016/j.jecp.2013.04.002

Nicolay, A. C., & Poncelet, M. (2015). Cognitive benefits in children enrolled in an early bilingual immersion school: A follow up study. *Bilingualism: Language and Cognition, 18*(4), 789–795. http://doi.org/https://doi.org/10.1017/S1366728914000868

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science, 10*(4), 464-480. doi: 10.1111/j.1467-7687.2007.00600.x

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232-258. doi: 10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6), 615-639. doi: 10.1080/20445911.2014.944914

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*, 962. doi: 10.3389/fpsyg.2014.00962

Paivio, A. (2014). Bilingual dual coding theory and memory. In Heredia R., Altarriba J. (Eds.), *Foundations of Bilingual Memory* (pp. 41-62). New York, NY: Springer. doi: 10.1007/978-1-4614-9218-4_3

Passolunghi, M. C., Cargnelutti, E., & Pastore, M. (2014). The contribution of general cognitive abilities and approximate number system to early mathematics. *British Journal of Educational Psychology*, *84*(Pt 4), 631–49. http://doi.org/10.1111/bjep.12054

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542-551. doi: 10.1016/j.tics.2010.09.008

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499–503. http://doi.org/10.1126/science.1102085

Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities, 32,* 2683–2689. http://doi.org/10.1016/j.ridd.2011.06.004

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113*(4), 535–551. http://doi.org/10.1016/J.JECP.2012.06.013

Poulin-dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, *108*, 567–579. http://doi.org/10.1016/j.jecp.2010.10.009

Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, *122*(1), 104–121. http://doi.org/10.1016/j.jecp.2013.12.009

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–1724. http://doi.org/10.1037/dev0000055

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*, 259–268. http://doi.org/10.1016/j.ecresq.2015.12.020

Purpura, D. J., Day, E., Napoli, A. R., & Hart, S. A. (2017). *Journal of Numerical Cognition*, *3*(2), 365–399. doi:10.5964/jnc.v3i2.53

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition*, *3*(2), 468–495. http://doi.org/10.5964/jnc.v3i2.31

Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Cognitive development math talk during informal learning activities in Head Start families. *Cognitive Development*, *35*, 15–33. http://doi.org/10.1016/j.cogdev.2014.11.002

Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, *39*(8), 582–590. http://doi.org/10.3102/0013189X10389811

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research, 47*(6), 877–903. http://doi.org/10.1080/00273171.2012.734737

Rugani, R., Vallortigara, G., & Regolin, L. (2015). At the root of the left-right asymmetries in spatial-numerical processing: From domestic chicks to human subjects. *Journal of Cognitive Psychology, 27*(4), 388-399. doi: 10.1080/20445911.2014.941846

Salillas, E., & Carreiras, M. (2014). Core number representations are shaped by language. *Cortex, 52*, 1–11. http://doi.org/10.1016/j.cortex.2013.12.009

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: early meanings of “one,” “two,” and “three” in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136–168. http://doi.org/10.1016/j.cogpsych.2006.09.001.

Sarnecka, B.W., Negen, J. & Goldman, M.C. (2018). Early Number Knowledge In Dual-Language Learners From Low-SES Households. In D. Berch, D. Geary and K. Mann-Koepke (Eds.), *Language and Culture in Mathematical Cognition*. (pp. 197-228). New York: Elsevier.

Schmidt, F. L. (2017). Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology, 5,* 32-41. doi: http://dx.doi.org/10.1037/arc0000033

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–337. http://doi.org/10.3200/JOER.99.6.323-338

Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–243.

Sijtsma, K. (2009). Reliability beyond theory and into practice. *Psychometrika, 74*(1), 169-173. doi: 10.1007/s11336-008-9103-y

Simmons, F. R., & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia, 14*, 77–94. http://doi.org/10.1002/dys

Simmons, F., Singleton, C., & Horne, J. (2008). Brief report—Phonological awareness and visual-spatial sketchpad functioning predict early arithmetic attainment: Evidence from a longitudinal study. *European Journal of Cognitive Psychology*, *20*(4), 711–722. http://doi.org/10.1080/09541440701614922

Simon, J.R. and Wolf, J.D. (1963). Choice reaction times as a function of angular stimulus-response correspondence and age. *Ergonomics, 6,* 99-105.

Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation. *Frontiers in Psychology*, *6*, 1062. http://doi.org/10.3389/fpsyg.2015.01062

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*(1), 45-88. doi: 10.1016/S0010-0277(00)00108-6

Standards Toolkit - Common Core Mathematics Assessments. (n.d.). Retrieved October7, 2015, from http://standardstoolkit.k12.hi.us/common-core/mathematics/mathematics-assessments/

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*(2), 97-127. doi: 10.1016/0010-0277(90)90001-Z

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.

Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology, 37*(4), 226–234. http://doi.org/10.1080/01973533.2015.1049351

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15*(4), 426–445. http://doi.org/10.1177/1367006911403202

Toll, S. W. M., Kroesbergen, E. H., & Van Luit, J. E. H. (2016). Visual working memory and number sense: Testing the double deficit hypothesis in mathematics. *British Journal of Educational Psychology, 86,* 429–445. http://doi.org/10.1111/bjep.12116

Toll, S. W. M., & Van Luit, J. E. H. (2014). Explaining numeracy development in weak performing kindergartners. *Journal of Experimental Child Psychology, 124,* 97–111. http://doi.org/10.1016/j.jecp.2014.02.001

Ursache, A., & Noble, K. G. (2016). Socioeconomic status, white matter, and executive function in children. *Brain & Behavior, 6*(10), e00531. doi: 10.1002/brb3.531

van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology, 6*(1), 25216. http://doi.org/10.3402/ejpt.v6.25216 van der Ven, S. H. G., Klaiber, J., & van der Maas, H. L. J. (2014). Four-and-twenty blackbirds : The effect of number word structure at different ages. Retrieved from http://dspace.library.uu.nl/handle/1874/305837

van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*(1), 100–119. http://doi.org/10.1111/j.2044-8279.2011.02035.x

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., & Ugen, S. (2015). The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition. *Frontiers in Psychology, 6*, 265. http://doi.org/10.3389/fpsyg.2015.00265

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology, 126*, 37–51. http://doi.org/10.1016/j.jecp.2014.02.012

von Oertzen, T., Brandmaier, A.M., & Tsang, S. (2015). Structural Equation Modeling with Ωnyx. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 148-161. doi: 10.1080/10705511.2014.935842

Vukovic, R. K., & Lesaux, N. K. (2013a). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology, 115*(2). http://doi.org/10.1016/j.jecp.2013.02.002

Vukovic, R. K., & Lesaux, N. K. (2013b). The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences, 23*, 87–91. http://doi.org/10.1016/j.lindif.2012.10.007

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Smíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., & Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76. http://doi.org/10.3758/s13423-017-1323-7

Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learning and Individual Differences, 52*, 79–87. http://doi.org/10.1016/j.lindif.2016.10.015

Wang, Y., Lin, L., Kuhl, P., & Hirsch, J. (2007). Mathematical and linguistic processing differs between native and second languages: An fMRI study. *Brain Imaging and Behavior, 1*, 68–82. http://doi.org/10.1007/s11682-007-9007-y

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., Engel, M., Siegler, R., & Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*(6), 1892–1907. http://doi.org/10.1111/cdev.12416

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112–2130. http://doi.org/10.1111/cdev.12099

Xenidou-Dervou, I., Gilmore, C., van der Schoot, M., & van Lieshout, E. C. D. M. (2015). The developmental onset of symbolic approximation: beyond nonsymbolic representations, the language of numbers matters. *Frontiers in Psychology, 6*(APR), 487. http://doi.org/10.3389/fpsyg.2015.00487

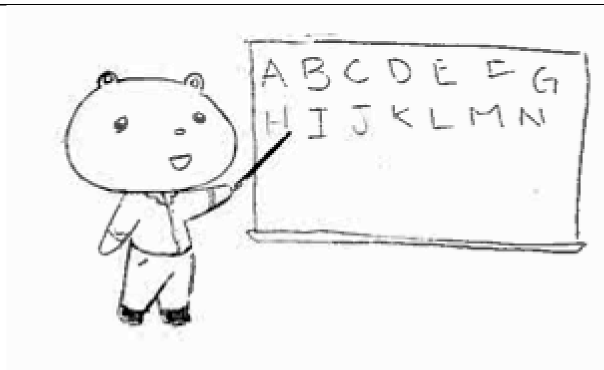
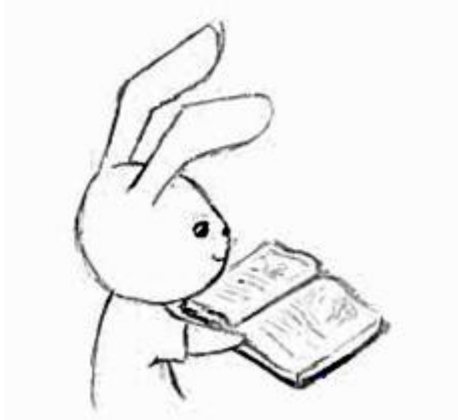
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11. doi: 10.1016/S0010-0277(99)00066-9

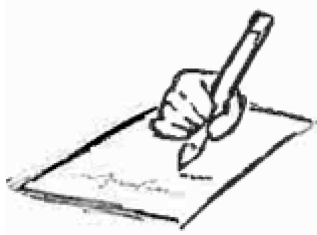
Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102*(1), 60–77. http://doi.org/10.1016/j.jecp.2008.04.003

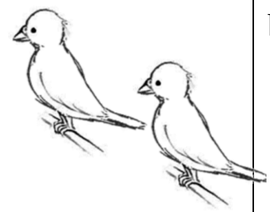
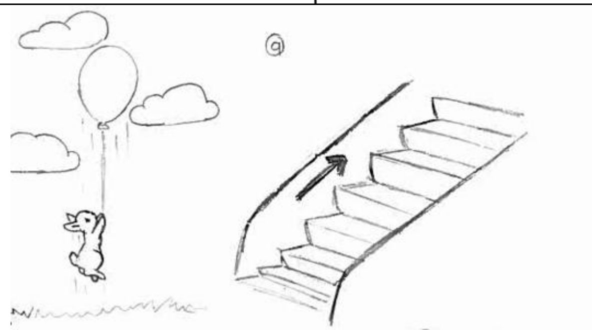
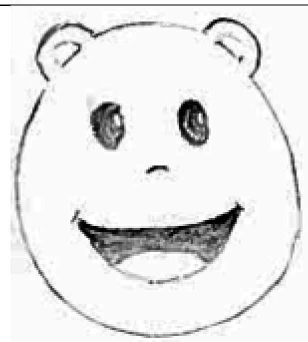
APPENDIX I

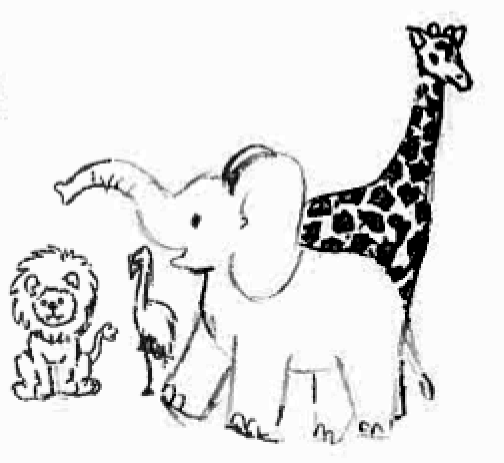
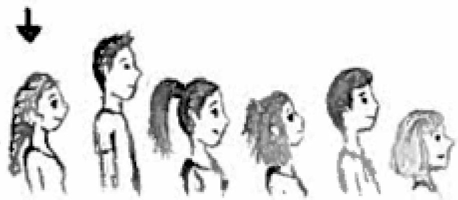
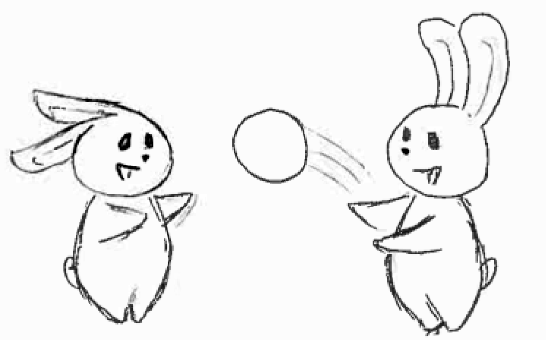
Modified EVT Questions and Stimuli

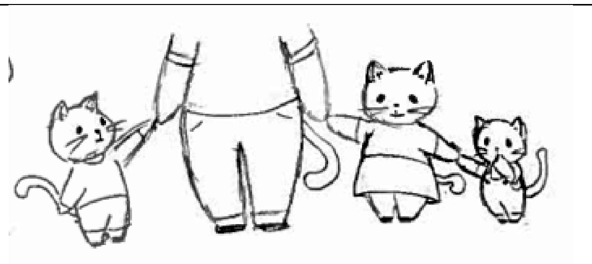
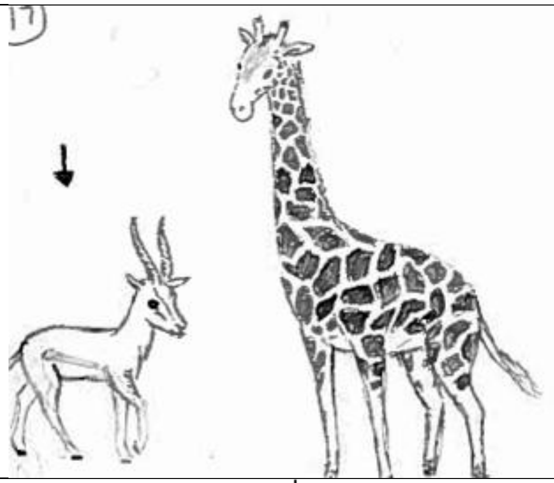
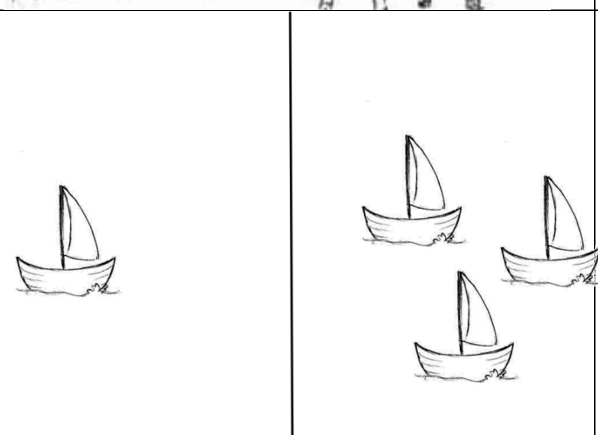
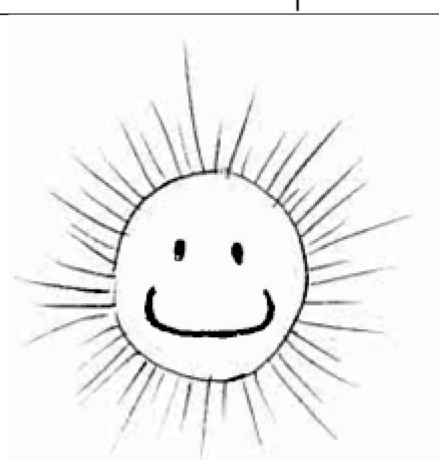
English Version

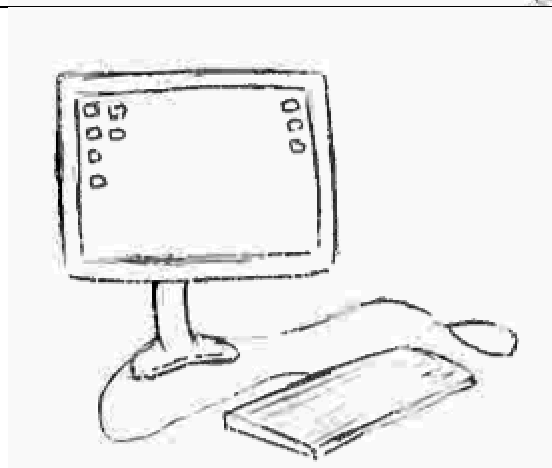
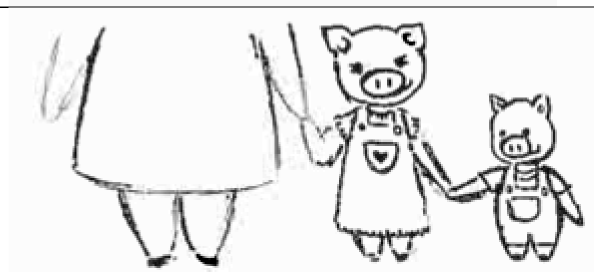
1	Who is this (point to the character)? (If they don't get it, prompt, Ms. __ is your ---?) Prompt if bear: this bear is a _____.		Teacher 老師
2	What is the bunny doing?		Read/Reading 唸書 看書 讀書
3	What is this place		School 學校

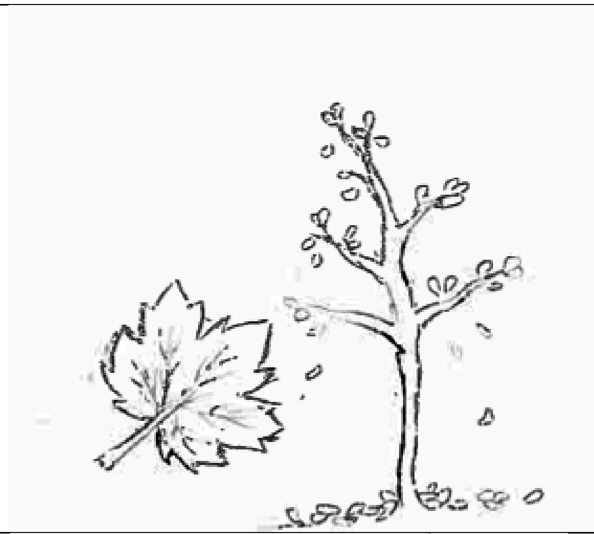
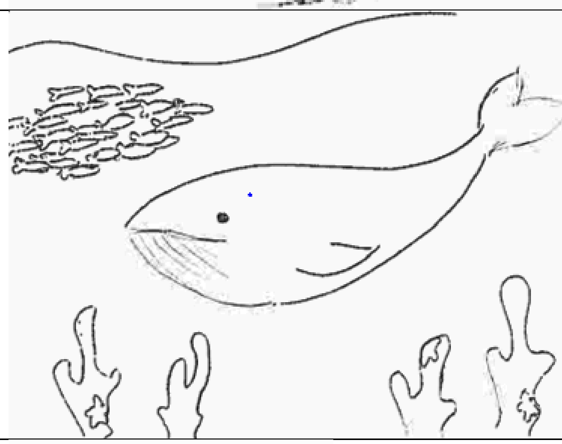
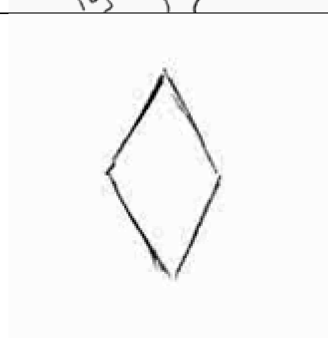
4	Who is holding the baby?		Mom, mommy, mama 媽媽
5	What is this person doing?		Walk/Walking 走路
6	What is the hand doing? (point to had on picture)		Write/writing 寫字 Draw/drawing (p)
7	What is this? (point to the side with 1 book) P with “how many?” if they only say “book”	 	One book 一本書 Book: 0.5

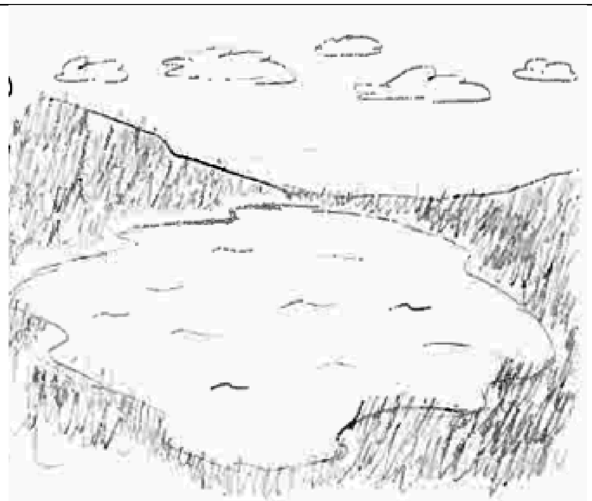
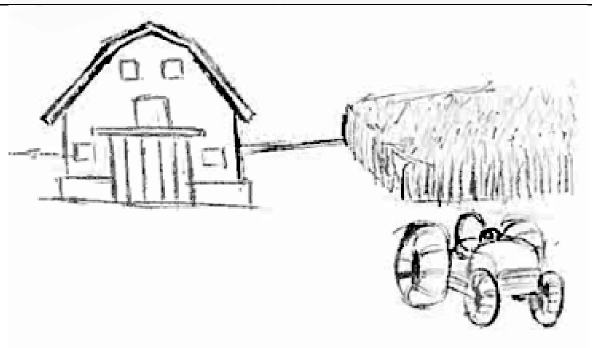
8	What is this? (point to the side with 1 bird) P with “how many?” if they only say “bird”			One bird 一隻鳥 bird: 0.5
9	I’m down here, I want to go __. (point to picture) OR Where is the bunny going? It’s going ___.			Up 上去 Upstairs(p)
10	He’s not sad, so he’s very ___.			Happy Glad Joyful 快樂 開心 高興
11	What is this? (If they only say “hand” then prompt “how many” and “this is one ___of hands?” If they said “holding hands” or similar, then point to hands on picture or our own hands and ask “what is this?”)			One pair of hands 一雙手 0.5: Two hands Hand(s) Holding hands

12	What are these? (prompt "all of them are called ___?" if child says individual animal names)		Animals 動物 P if: Elephant, giraffe, lion, crane, etc.
13	If this girl (point) is first in line, then she is (point to last in line)?		Last 最後
14	What are the two bunnies doing? (prompt with "another way to say this?" if they say throwing or catching or similar verbs. If they keep saying specific verbs, prompt with "what do you do with your friends at recess?")		**Playing** (ball/catch) 玩(球)
15	This is a little kid (point). This is the kid's ___. P if "dad" or "papa"		Father 父親 0.5: Dad, daddy, papa, pop 爸爸

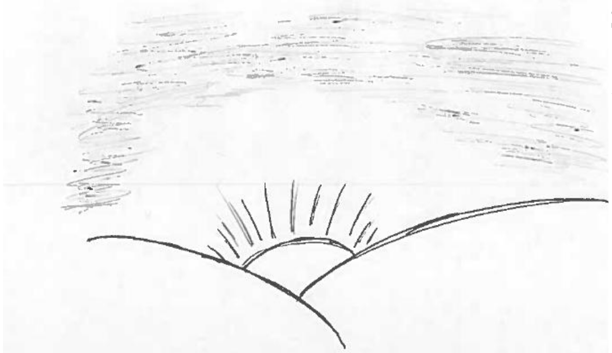
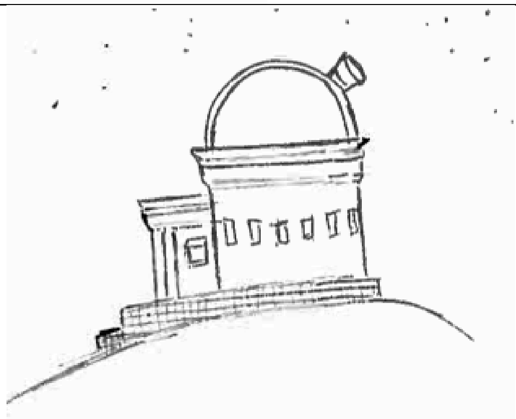
16	This is an adult/grown up, so these are (point to the kids)		Children Kids 小孩 小朋友
17	The giraffe is very tall, and the gazelle is very ___.		Short 矮(的)
18	What is this? (point to the side with 1 book. Prompt with “how many?” if they only say “boat”)		One boat 一艘船 一條船 0.5 if boat, sailboat
19	What is this?		Sun 太陽

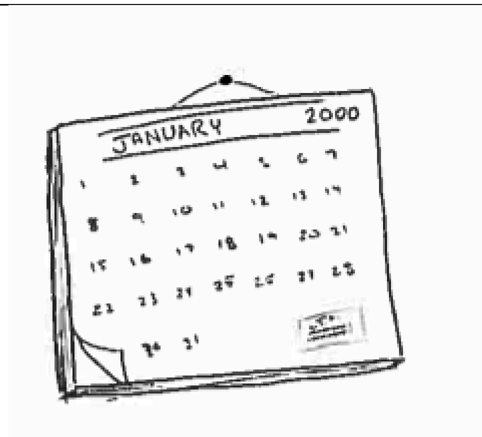
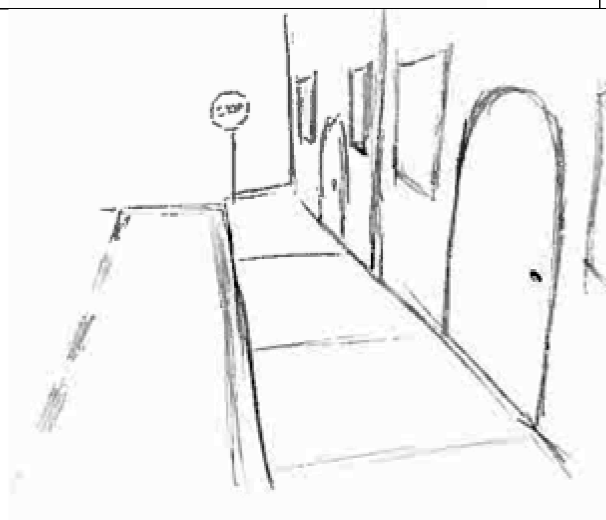
20	When does the moon come out?		Night(time) Evening 晚上
21	What is this?		Computer 電腦
22	If this is the older sister (point), this little boy is her ___.		Younger/little/ baby brother 弟弟
23	What did you do with the music? You used your ear and ___.		Hear/heard 聽到/聽見

24	Which season is this? (can also prompt it's after summer.)		Autumn / Fall 秋天
25	Where do the whales and fish swim and live?		Ocean/sea 海洋 Under the sea In the ocean
26	What is this shape?		Diamond/ Rhombus 菱形
27	What is this?		Tornado 龍捲風

28	What is this? (picture might not be obvious, so may have to say "there's a lot of water" as you point to the lake area. If they said pond or sea, might have to prompt with "bigger" or "smaller")		Lake 湖泊 0.5 Pond
29	Which industry is this?		Agriculture 農業 0.5 farm
30	Who are these people? (if they start saying animal names, ask them "what do they do?" If they say "playing music" or any of the playing variety, then you prompt with "what do we call people who play music/instruments?")		Musicians 音樂家 0.5 pianist

31	Dogs are mammals; what are frogs?		Amphibians 兩棲(類)動物
32	What is this smelly thing?		Trash Garbage 垃圾
33	What is this?		Statue (of liberty) 雕像 自由女神(p)

34	What happens before it gets dark? (point to the sun) If they say "the sun goes down," prompt with "what's another way to say that?"		Sunset 日落 夕陽
35	She is very ____ (or she is ___ citizen), but she still has a lot of energy.		Senior Elderly 年老
36	Where do you go to see the stars?		Observatory 天文台
37	What are the people who saw boards or work with wood called?		Carpenters 木匠

38	What is this?		Calendar 日曆 月曆
39	The cars go on the road (point to it), what do we call the place that people go on? (point to sidewalk)		Sidewalk 人行道
40	The farmer plants and sows the seeds in the spring, what does he do in the fall?		Harvest 收割

P = prompt

Please contact author for the Chinese prompts.

APPENDIX II

Parent Questionnaire

Please tell us a little about you and your child. The information will be kept confidential and will not influence you and your child's eligibility to participate in the study.

1. What is your child's age (years) and date of birth (month, day, year) ______________________

2. What is your child's gender (circle one): MALE FEMALE

3. What is your child's ethnicity? (Check the one that best describes your child)
   - ❑ Hispanic or Latino
   - ❑ Not Hispanic or Latino
   - ❑ Other (please list all groups that apply) _____________________________

4. What is your child's race? (Check the one that best describes your child)
   - ❑ African American or Black
   - ❑ Caucasian/White
   - ❑ Asian or Pacific Islander
   - ❑ American Indian or Alaska Native
   - ❑ Biracial/Mixed Race (Please list all groups that apply) _____________________________

5. Please indicate your relationship to your child (e.g. father, mother): __________________

6. Please indicate the highest level of education completed by your child's mother:
   - ❑ Some High School Coursework
   - ❑ High School Diploma/GED
   - ❑ Some College Coursework/Vocational Training
   - ❑ 2-year College Degree (Associates)
   - ❑ 4-year College Degree (BA/BS)
   - ❑ Postgraduate or Professional degree (MA, PhD, MD, JD)

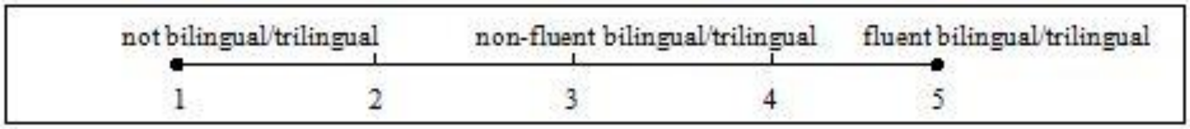
7. Please indicate the highest level of education completed by your child's other parent:
   - ❑ Some High School Coursework
   - ❑ High School Diploma/GED
   - ❑ Some College Coursework/Vocational Training
   - ❑ 2-year College Degree (Associates)
   - ❑ 4-year College Degree (BA/BS)
   - ❑ Postgraduate or Professional degree (MA, PhD, MD, JD)

8. How many people typically reside in your household? ______________

9. Annual household income:
   - ❑ Less than $15,000
   - ❑ $15,000 - $30,000
   - ❑ $31,000 - $45,000
   - ❑ $46,000 - $59,000
   - ❑ $60,000 - $75,000
   - ❑ $76,000 - $100,000
   - ❑ $101,000 - $150,000
   - ❑ $151,000 or more

Language Background

1. Overall, how would you describe ***your child's*** level of bilingualism or trilingualism? Choose one.



1 – speak predominantly one language
- only know a few words in the other language.

2 – weak bilingual / trilingual
- know enough to carry out some conversation to a very limited extent (use key words with not much grammar)
- need to listen to sentences more than once before understanding.

3 – unbalanced bilingual / trilingual
- able to carry out basic conversation with minor grammatical errors without the other speake repeating
  the sentence
- have difficulty producing a fluent conversation.

4 – practical bilingual / trilingual
- can carry on fluent conversation
- do not use the second language every day

5 – fluent bilingual / trilingual
- able to converse fluently and actively; use two or three languages every day
- lived abroad in a community that has English as the dominant language

2. List all the languages and dialects your child can speak including English, *in order of fluency:*
1. ____________________ 2. ______________________ 3. ________________________

3. What language does your child speak most at home? _______________________________

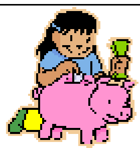
4. What language does your child hear most at home? _______________________________

5. Please rate your child's current abilities in each language on a scale from 0-5, compared to other children his/her age (1 = low, 2 = slightly lower than average, 3 = fair, 4 = slightly higher than average, 5 = high)

	Language #1	**Language #2**	**Language #3**
Speaking			
Understanding spoken language			
Reading			
Writing			

Play Background

In the past month, how often have you and your child engaged in the following activities? ***Circle 0*** *if the activity did not occur,* ***1*** *if it occurred less than once per week, but a few times a month,* ***2*** *if it occurred about once a week,* ***3*** *if it occurred a few times a week,* ***4*** *if it occurred almost daily, 5 if it occurred every day* ***and N/A*** *if the activity is not applicable to your child.*

	Did not occur	1-3 times per month	Once per week	2-4 times per week	Almost daily	Daily	Activity is not relevant to my child
Reading together	0	1	2	3	4	5	N/A
Saying/singing the ABC's	0	1	2	3	4	5	N/A
Counting out loud	0	1	2	3	4	5	N/A
Counting by a number other than 1 (by 2's, by 5's, by 10's)	0	1	2	3	4	5	N/A
Noticing letters and words	0	1	2	3	4	5	N/A
Counting objects	0	1	2	3	4	5	N/A
Labeling letters or words	0	1	2	3	4	5	N/A
Talking about how many objects are in a set (e.g. there are 5 toys in the basket)	0	1	2	3	4	5	N/A
Memorizing letters/sounds or sight words	0	1	2	3	4	5	N/A
Memorizing math facts	0	1	2	3	4	5	N/A
Writing numbers	0	1	2	3	4	5	N/A
Point to letters/words while reading	0	1	2	3	4	5	N/A
Comparing numbers (e.g. "2" is bigger than "1")	0	1	2	3	4	5	N/A
Counting down (10, 9, 8, 7…)	0	1	2	3	4	5	N/A
Talking about meanings of words	0	1	2	3	4	5	N/A
Talking about what letters words start with	0	1	2	3	4	5	N/A
Introducing new words and definitions	0	1	2	3	4	5	N/A

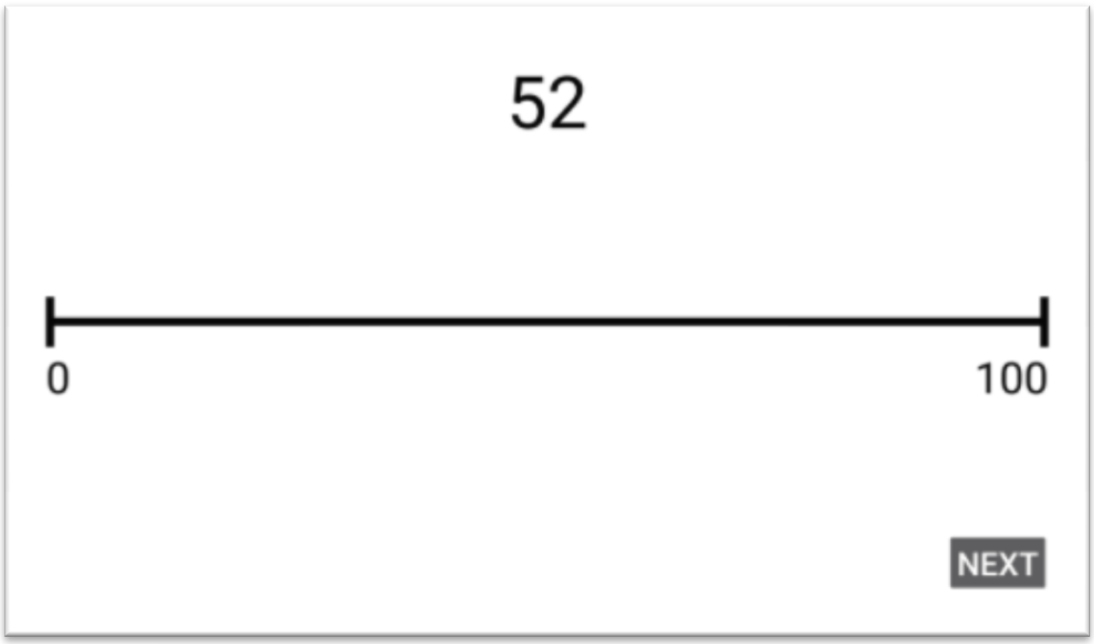
Counting out money	0	1	2	3	4	5	N/A
Asking questions when reading together	0	1	2	3	4	5	N/A
Comparing amounts (e.g. 3 cookies is more than 1 cookie)	0	1	2	3	4	5	N/A
Talking about letter sounds	0	1	2	3	4	5	N/A
Using fingers to indicate how many	0	1	2	3	4	5	N/A
Sounding out words	0	1	2	3	4	5	N/A
Learning simple sums (e.g., 2 + 2)	0	1	2	3	4	5	N/A

Thank you for taking the time to complete this questionnaire!

APPENDIX III

Number Line Estimation Task

Screenshot of the task on the tablet:



Children get to touch the number line to indicate where between 0 and 100 the number (e.g., 52) would be.

Children perform the task 26 times. The numbers below are displayed in randomized order on the digital task.

3	29	81
4	33	84
6	39	90
8	42	96
12	48	
14	52	
17	57	
18	61	
21	64	
24	72	
25	79	

APPENDIX IV

Common Core Base-10 Questions

Highlighted Questions are given to all students. All others are only given to children in the Mandarin immersion classes.

1) Counting

- Mystery Bag - 3 questions
  - How many stars are in bag A?
  - How many stars are in bag B?
  - How many stars are there all together?
- Rote Counting from 1 to 120

2) Number Pattern

- 4 questions; ask participants to complete the pattern and how they figured it out

5, 10, 15, 20, ______, ______, ______, ______

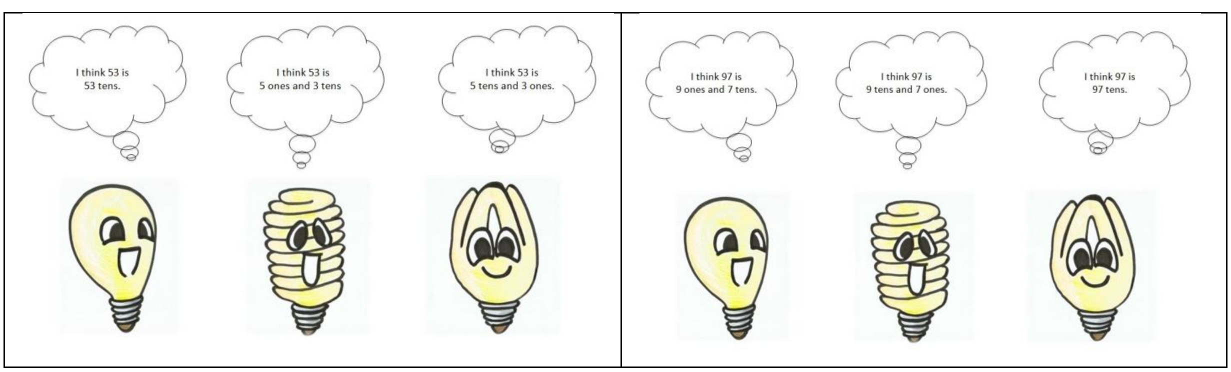
30, 40, 50, 60, _____, _____, _____, _____

4, 6, 8, 10, ______, ______, ______, ______

23, 33, 43, 53, _____, _____, _____, _____

3) Farmer Brown and Who is Right

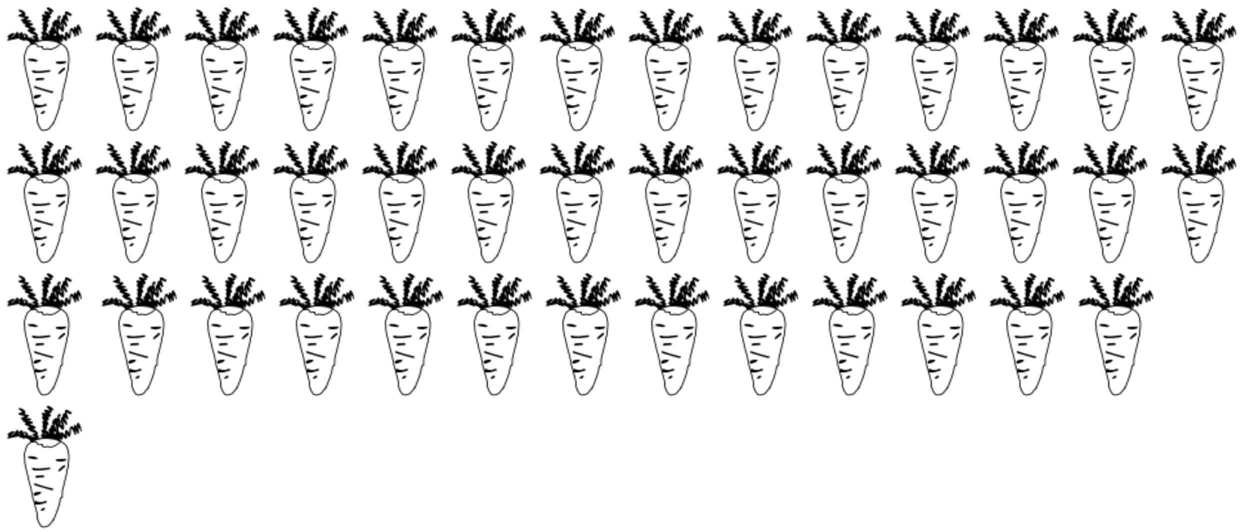
- Farmer Brown
  - one worksheet requiring participants to count and group items into groups of 10 (see next page)
- Who is Right
  - Two questions, each asking the child to choose which "light bulb" explained a number correctly



Secret Code:

Farmer Brown

Help Farmer Brown count the vegetables he has picked from his field. Count the carrots. Circle in groups of 10.



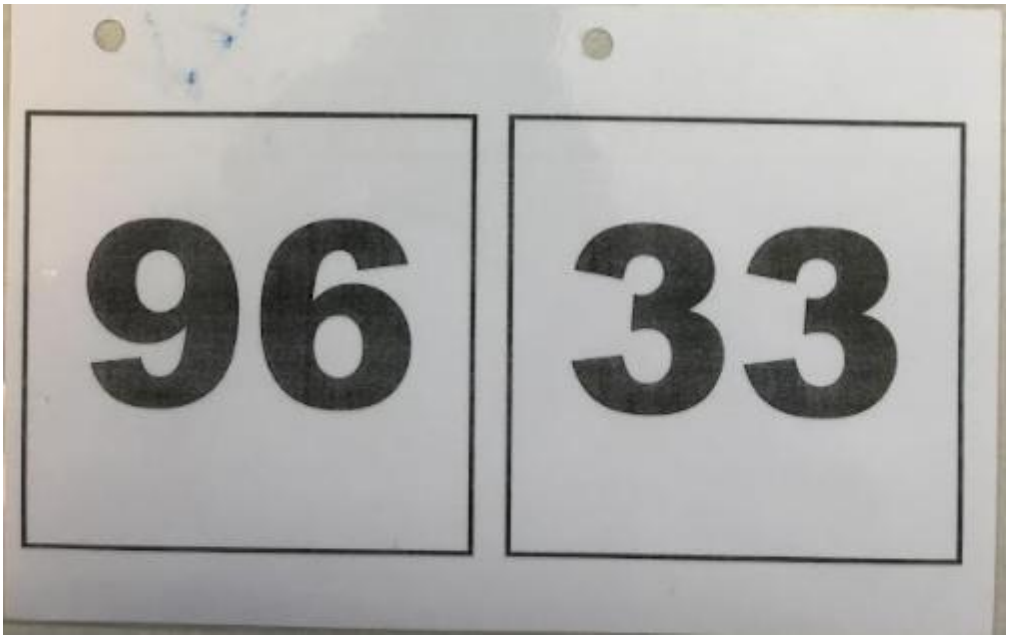
How many carrots did Farmer Brown pick?   ______ carrots

How many tens and ones is that? _____tens   _____ones

APPENDIX V

Magnitude Comparison Number Pairs

Sample stimuli



Children view a number pair as the experimenter reads the numbers out loud and asks them to select the bigger of the two numbers ("which one is bigger/more?")

The Number Pairs and their Ratios

Ratios	Number 1	Number 2	Ratios	Number 1	Number 2
0.32	79	25	0.58	57	33
0.32	8	25	0.59	57	96
0.36	33	12	0.60	42	25
0.38	8	3	0.66	64	42
0.39	25	64	0.67	8	12
0.42	79	33	0.67	64	96
0.44	25	57	0.68	84	57
0.44	96	44	0.72	79	57
0.48	12	25	0.74	42	57
0.50	84	42	0.76	25	33
0.52	33	64	0.76	84	64
0.53	42	79	0.79	33	42

APPENDIX VI

Following Instructions Sample Questions Form A

1 Action

1) Touch the yellow cup
2) Touch the blue plane
3) Touch the red car
4) Touch the blue fish

2 Action

1) Pick up the yellow car and put it in the blue cup
2) Pick up the red fish and put it in the yellow box
3) Pick up the blue car and put it in the red cup
4) Pick up the yellow plane and put it in the red box

3 Action

1) Touch the blue fish THEN pick up the yellow car and put it in the red box
2) Touch the red car THEN pick up the blue plane and put it in the yellow cup
3) Touch the yellow fish THEN pick up the blue car and put it in the red box
4) Touch the red plane THEN pick up the red car and put it in the yellow box

4 Action

1) Pick up the yellow plane and put it in the red box THEN pick up the blue fish and put it in the yellow box

2) Pick up the blue car and put it in the red box THEN pick up the yellow plane and put it in the blue cup

3) Pick up the blue fish and put it in the yellow cup THEN pick up the yellow plane and put it in the red box

4) Pick up the blue car and put it in the red cup THEN pick up the yellow fish and put it in the blue box

5 Action

1) Touch the red plane THEN pick up the yellow car and put it in the blue cup THEN pick up the red plane and put it in the blue box

2) Touch the blue box THEN pick up the red plane and put it in the yellow cup THEN pick up the red fish and put it in the yellow cup

3) Touch the red plane THEN pick up the blue car and put it in the yellow cup THEN pick up the blue fish and put it in the red box

4) Touch the yellow fish THEN pick up the red plane and put it in the blue box THEN pick up the yellow car and put it in the red cup

6 Action

1) Pick up the blue car and put it in the yellow box THEN pick up the red fish and put it in the blue cup THEN pick up the red plane and put it in the yellow cup

2) Pick up the red fish and put it in the yellow cup THEN pick up the blue fish and put it in the blue cup THEN pick up the red plane and put it in the yellow cup

3) Pick up the yellow plane and put it in the blue cup THEN pick up the blue car and put it in the red cup THEN pick up the red fish and put it in the yellow box

4) Pick up the blue fish and put it in the red box THEN pick up the yellow plane and put it in the yellow box THEN pick up the blue car and put it in the red cup

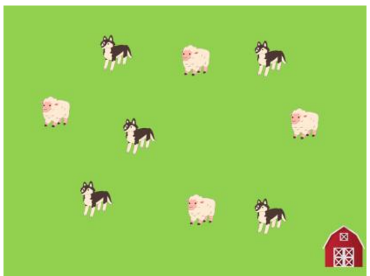
APPENDIX VII

Counting Sheep Administration Protocol

Counting Sheep

Instructions

Introduction to the game: **“In this game, you will see sheep and wolves on the screen.”**

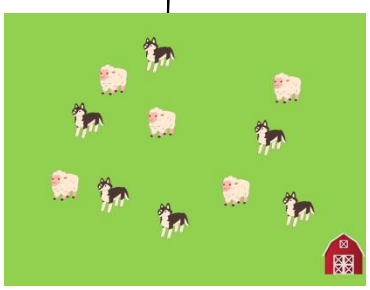


1) “I am going to ask you to count all the *sheep* out loud as fast as you can! After you’re done counting, I want you to tell me the number of sheep that you counted.”

“Let’s try one together! There are: 1,2,3,4.” Point to sheep as you count. **“So how many did you count?”** Wait for child’s answer.

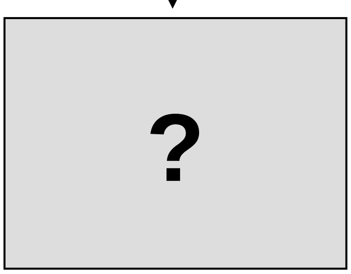
If child counts wolves also say, **“Remember – just count the *sheep,* not the wolves.”**

“Now, I want you to *remember* the number you just told me.”



2) “Now there’s another picture of sheep and wolves. For this picture, I want you to do the same thing you just did. You’re going to count the *sheep* and then tell me the number.” Let child count. If they need prompting say, “**How many sheep do you see?”**

“Be sure to remember that number too!”



3) **“Now a big question mark will show up. That means that it is time to remember the number of sheep that you counted for each picture. Think hard, and tell me the numbers in the order that you told me.**”

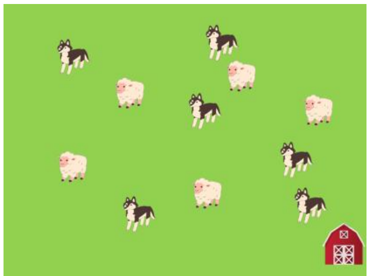
If the child is silent or says they can’t remember say, “**So, what was the first number you told me?”** Wait for answer. **“What was the second number you told me?”** Wait for answer.

If the child says the correct numbers, but in the wrong order say, “**Remember to tell me your numbers *in the right order*, *the same order you saw them*.”**

“Also, don’t worry about doing anything with the computer – I will do that for you. I just want you to focus on counting the sheep and remembering the number of sheep in each picture in the same order you see them.”

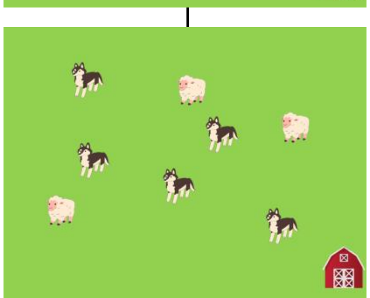
"Do you have any questions?" Wait for questions. **"All right. Let's practice!"**

[All entries are done by the experimenter! The practice consists of 2 trials with set size 2.]



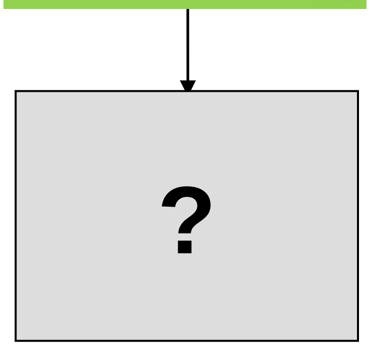
Alright, can you count the sheep for me? How many sheep do you see?

Enter the response in the computer. The entered number will be displayed in the lower right corner on the screen. After you hit ENTER, the next array will appear.



How about in this picture? How many sheep do you see here?

Enter the response in the computer. After you hit ENTER, the next array appears.



Now, I want to see if you can remember how many sheep you counted in the two pictures you just saw. How many sheep did you see in the first picture? And, *then*, how many did you see in the second picture?

Enter the responses in the computer. Also here, the number you entered will show up on the screen. After you hit ENTER, the next series will appear.

Important notes:

If the participant recalls the results in the wrong order, please remind her to tell you the results in the correct order!

If the task is not clear after the practice items, please repeat the practice by clicking on key Y (otherwise start the task by clicking on key N).

If there was a mistake in entering the numbers (if either experimenter hits the wrong digit or participant corrects), the entry can be corrected. After practice, there will be set sizes 2-3-4-5 in random order. These sequences will be repeated 3 times. The participant is allowed to take a short break after each round.

Test trials

"I like how you worked really hard on that. Let's try some more!"

If child does not tell you their count say, **"How many sheep do you see?"**

If child does not tell you sequence of counts say, **Insert prompt for question mark screen here.**

APPENDIX VIII

Combined Exploratory Factor Analysis

Table 1. Exploratory factor analysis with all measures for all kindergartners: factor loadings (loadings > 0.30 are boldfaced), uniqueness, and factor correlations

	Factor 1	**Factor 2**	**Factor 3**	**Uniqueness**
PVT	0.067	-0.060	**0.821**	0.305
EVT	0.014	0.149	**0.554**	0.600
PQ	-0.014	0.074	**0.748**	0.405
NLE	**-0.732**	-0.024	0.154	0.533
CCCQ	**0.762**	0.025	0.027	0.377
NI	**0.822**	0.020	0.095	0.221
MC	**0.622**	-0.049	0.118	0.563
FI	-0.028	**0.845**	0.003	0.308
TB-F	0.243	**0.595**	-0.092	0.487
TB-B	0.018	**0.598**	0.173	0.522
CS	-0.034	**0.683**	0.008	0.553
Factor 1	1	.	.	
Factor 2	0.611	1	.	
Factor 3	0.567	0.447	1	
% Variance	21.4	18.6	15.7	

Note. N = 59. PVT = bilingual Picture Vocabulary Test. EVT = Expressive Vocabulary Test. PQ = Parent Questionnaire. NLE = Number Line Estimation. CCCQ = Common Core Counting Questions. NI = Number Identification. MC = Magnitude Comparison. FI = Following Instructions. TB-F = Touch Base Forward. TB-B = Touch Base Backward. CS = Counting Sheep. Extraction method: minimal residual, rotation: oblimin.

APPENDIX IX

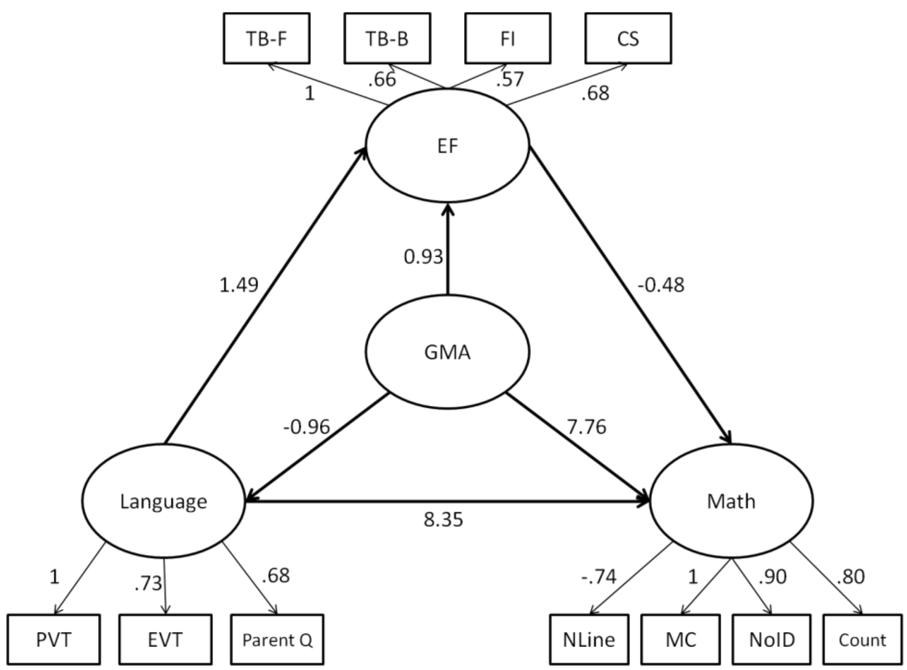
Supplementary Materials for Chapter 3 SEM Analyses

Model Fit Indices for Chapter 3 Models from Ωnyx.

Model	CFI	TLI	RMSEA	Chi-Sq	BIC (sample-size adjusted)
Figure 3.2	0.994	0.992	0.024	$\chi^2(41) = 42.99$ $p = .39$	1994.769
Hypothesis 3a	0.986	0.979	0.039	$\chi^2(38) = 42.99$ $p = .27$	2008.132
Hypothesis 3b	0.994	0.992	0.024	$\chi^2(41) = 42.99$ $p = .39$	1994.769

As can be seen from the χ^2 tests, the model (Figure 3.2) for testing Hypotheses 1 and 2 are statistically equivalent with the model for Hypothesis 3b. The estimates displayed in Figure 3.2 were obtained using Stata 14. The Ωnyx estimates were similar: 0.60 for language to executive function, 0.42 for executive function to math, and 0.38 for language to math.

The difference of around 13.4 between the adjusted BIC for Hypotheses 3a and 3b (and Figure 3.2) suggests strong evidence that 3b (and Figure 3.2) are better models. This is unsurprising as this model for Hypothesis 3a is overspecified, resulting in standardized path coefficients greater than 1 in many cases. (Standardized path coefficients should be between 0 and 1.) Because of the poorer model fit and the nonsensical path coefficients resulting from overspecification, this model was not presented in the results or discussion in Chapter 3.



The model displayed here matches that of Hypothesis 3A from Chapter 3: A higher-order latent construct of general mental ability (GMA) would predict the three individual constructs of language, executive function (EF), and math, and that the relations among the three constructs remain to be freely estimated. This model did not converge using Stata14 but converged in Ωnyx.

APPENDIX X

Mandarin and English Vocabulary

The information here pertains only to children enrolled in the Mandarin immersion classes.

Table X.

Mean (standard deviation) of the Number of Correctly Answered Items in Each Language

	Mandarin	**English**	**Overall (Conceptual Score)**
PVT	16.89 (11.69)	38.59 (12.61)	76.30 (11.27)
EVT	9.12 (6.47)	18.86 (6.57)	27.97 (1.52)

For the receptive Picture Vocabulary Test, conceptual score is not a sum of the correct items answered after listening to Mandarin or English stimuli as items below basal are assumed to be correct (a score of 1 point each). Out of all attempted questions that were correctly answered, kindergartners from the Mandarin immersion program answered a mean of 70% of questions (SD = 18.3%) in English.

For the Expressive Vocabulary Test, the items displayed are the questions that were answered correctly in the given languages. Children attempted to answer in Mandarin first, and they answered in English if they were unable to provide a correct response in Mandarin. Therefore, children’s English vocabulary should be higher than the average displayed on the table, as they were not asked to answer in English the questions they answered correctly in Mandarin. The overall score for EVT *is* a sum of their scores in Mandarin and English as the kindergartners attempted all 40 questions.