Running head: COMPARING…EARLY ARITHMETIC INSTRUCTION

Comparing the Efficacy of Early Arithmetic Instruction Based on a Learning Trajectory and Teaching-to-a-Target

Douglas H. Clements[1], Julie Sarama[1], Arthur J. Baroody[2], Traci S. Kutaka[1], Pavel Chernyavskiy[3], Candace Joswick[4], Menglong Cong[1], and Ellen Joseph[1]

[1] University of Denver
[2] University of Illinois at Urbana-Champaign
[3] University of Wyoming
[4] University of Texas at Arlington

Author Note

Douglas H. Clements https://orcid.org/0000-0003-1800-5099

Julie Sarama https://orcid.org/0000-0003-1275-6916

Arthur J. Baroody. https://orcid.org/0000-0003-4296-8434

This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R305A150243. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. Researchers from an independent institution oversaw the research design, data collection, and analysis and confirmed findings and procedures. The authors wish to express appreciation to the school districts, teachers, and children who participated in this research. We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Douglas H. Clements University of Denver Katherine A. Ruffatto Hall Rm. 152/154 Denver CO 80208-1700. Email: Douglas.Clements@du.edu

Abstract

Although basing instruction on a learning trajectory (LT) is often recommended, there is little evidence regarding the premise of a LT approach—that to be maximally meaningful, engaging, and effective, instruction is best presented one LT level beyond a child's present level of thinking. We evaluated this hypothesis using an empirically-validated LT for early arithmetic with 291 kindergartners from four schools in a Mountain West state. Students randomly assigned to the LT condition received one-on-one instruction one level above their present level of thinking. Students in the counterfactual condition received one-on-one instruction that involved solving story problems three levels above their initial level of thinking (a teach-to-target approach). At posttest, children in the LT condition exhibited significantly greater learning, including target knowledge, than children in the teach-to-target condition, particularly those with low entry knowledge of arithmetic. Child gender and dosage were not significant moderators of the effects.

KEYWORDS: Achievement, curriculum, early childhood, instructional design/development, learning trajectories, learning environments, mathematics education

Educational Impact and Implications Statement

The results of this study underscore the benefits of teaching early arithmetic following learning trajectories, that is, providing instruction that is just beyond a child's present level of thinking. Children who experiences this approach learned significantly more than those who were taught the target skills for the same time period. Therefore, instruction following learning trajectories may promote more learning, including learning target competencies, than an equivalent amount of instruction on these target competencies with developmentally unready children.

The use of learning trajectories (LTs) in early mathematics instruction has received increasing attention from educators, curriculum developers, and researchers {Baroody, 2019 #8346;Clements, 2014 #5679;Maloney, 2014 #4653;Sarama, 2009 #3380}. For example, LTs were a core construct in the NRC {National Research Council, 2009 #3857} report on early mathematics education (note the subtitle: "Paths toward excellence and equity") and the notion of levels of thinking was a key first step in the writing of the Common Core State Standards — Mathematics {NGA/CCSSO, 2010 #4143}. Despite these recommendations, little research has directly tested the specific contributions of LTs to teaching compared to instruction provided without LTs {Frye, 2013 #4610}. The goal of the present study was to compare the learning of kindergarteners who received arithmetic instruction grounded in an empirically-validated LT to those who received an equal amount of time dedicated to solving story problems at the target level – three levels beyond the child's initial level.

Background and Theoretical Framework

Learning Trajectories are not only under-researched, they are often defined differently {Frye, 2013 #4610}. For example, some have confused LTs with a logical task analysis, hierarchies or sequences based solely on the structure of mathematics content {Resnick, 1981 #1971}, or the on accretion of facts and skills {Carnine, 1997 #2558}. Others have valid, but distinct, definitions of related constructs, such as learning progressions, sequences of assessment tasks, or cognitive patterns of thinking {e.g.`, \National Research Council, 2007 #3247;Steedle, 2009 #7725}. In contrast, to be optimally useful to educators, learning trajectories must include and integrate educational standards, children's learning, and teaching strategies. Therefore, we define a LT as having three components: a goal, a developmental progression of levels of thinking, and instructional activities (including curricular tasks and pedagogical strategies) designed explicitly to promote the development of each level {Clements, 2004 #2125;Maloney, 2014 #4653;National Research Council, 2009 #3857;Sarama, 2009 #3380}. *Goals* are based on the structure of mathematics, societal needs, and research on children's thinking about and learning of mathematics, and require input from those with expertise in mathematics, policy, and psychology {Clements, 2004 #1717;Fuson, 2004 #1720;Sarama, 2009 #3380;Wu, 2011 #3385}.

Descriptions of the other two components of learning trajectories requires more detailed consideration of the theory in which they are embedded, *hierarchic interactionalism* {Sarama, 2009 #3380}. The term indicates the influence and interaction of global and local (domain specific) cognitive levels and the interactions of innate competencies, internal resources, and experience (e.g., cultural tools and teaching). Consistent with Vygotsky's construction of the zone of proximal development {Vygotsky, 1935/1978 #2610}, the theory posits that most content knowledge is acquired along developmental progressions of levels of thinking within a specific topic, consistent with children's informal knowledge and patterns of thinking and learning. Each level is more sophisticated than the last and is characterized by specific concepts (e.g., mental objects) and processes (mental "actions-on-objects") that underlie mathematical thinking at level *n* and serve as a foundation to support successful learning of subsequent levels. However, levels are not stages but probabilistic patterns of thinking through which most children develop {e.g.`, an individual may learn multiple levels simultaneously or in a slightly different order`, \Sarama, 2009 #3380}. Developmental progressions are the second component of a LT.

The theory also posits that teaching based on those developmental progressions is more effective, efficient, and generative for most children than learning that does not follow these paths. Thus, each LT includes a third component, recommended *instructional activities* corresponding to each level of thinking. That is, based on the hypothesized, specific, mental constructions (mental actions-on-objects) and patterns of thinking that constitute children's thinking, curriculum developers design instructional tasks that include external objects and actions that mirror the hypothesized mathematical activity of children as closely as possible. These tasks are sequenced, with each corresponding to a level of the developmental progressions, to complete the hypothesized learning trajectory. Such tasks will theoretically constitute a particularly efficacious educational program; however, there is no implication that the task sequence is the only path for learning and teaching; only that it is hypothesized to be one fecund route. In sum, LTs are “descriptions of children’s thinking and learning in a specific mathematical domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking” {Clements, 2004 #2125`, p. 83;Sarama, 2009 #3380`, provides a complete description of hierarchic interactionalism’s 12 tenets}.

Turning to the evidentiary base, the goals and developmental progressions for many topics have been supported and validated by theoretical and empirical work describing consistent sequences of thinking levels, although the amount of empirical support differs for different topics and ages {Confrey, 2019 #9684;Daro, 2011 #4343;Gravemeijer, 1994 #1449;Maloney, 2014 #4653;National Research Council, 2009 #3857}, especially in domains such as the approximate number system and subitizing {e.g.`, \Clements, 2019 #4384;vanMarle, 2018 #8597;Wang, 2016 #8184}, counting {e.g.`, \Fuson, 1988 #948;Purpura, 2013 #10112;Spaepen, 2018 #9315}, and arithmetic {e.g.`, \Hickendorff, 2010 #8638`, see the following section for early arithmetic}. Further, the application of developmental progressions as curricular guides {e.g.`, \Clarke, 2001 #2057} and complete learning trajectories {i.e.`, \Clements, 2008 #2785;Clements, 2011 #4177} have been successfully applied in early mathematics intervention projects, with significant effects on teachers’ professional development {Clarke, 2008 #4294;Kutaka, 2016 #8188;Wilson, 2013 #5964} and children’s achievement {Clarke, 2001 #2057;Clements, 2008 #2785;Clements, 2011 #4177;Kutaka, 2017 #8189;Murata, 2004 #2571;Wright, 2006 #2868}.

Despite this research foundation, there is little research that directly tests the theoretical assumptions and specific educational contributions of LTs. That is, most studies showing positive results of LTs confound the use of LTs with other factors {Baroody, 2017 #5605;Frye, 2013 #4610}, thus suggesting the efficacy of the use of LTs without identifying their unique contribution, particularly beyond that of other instructional approaches {Clarke, 2001 #2057;Clements, 2007 #2091;Clements, 2011 #4177;Fantuzzo, 2011 #4529;Gravemeijer, 1999 #1412;Jordan, 2012 #5144}. For example, preschoolers who experienced a curriculum specifically designed on LTs increased significantly more in mathematics competencies than those in a business-as-usual control group score (effect size, 1.07) and more than those who experienced an intervention using a research-based curriculum that followed a sequence of mathematically-rational topical units {effect size`, .47`, \Clements, 2008 #2785}. Given that the contents of the two curricula were closely matched, the latter difference may be due to the use of LTs (e.g., the developmental progressions of the LTs provided benchmarks for formative assessments, especially useful for children who enter with less knowledge). However, the two curricula also differed in organization (e.g., interwoven counting, arithmetic, geometry and patterning LTs vs. separate units on these topics) and in specific activities. Therefore, again, several factors were confounded and the specific effects of LTs could not be distinguished {Clements, 2008 #2785}.

The Present Study

To address these gaps in the research corpus, we designed a series of experiments to examine the unique contributions of LTs to mathematics teaching and learning covering different ages and topics {e.g.`, \Clements, 2019 #9686`, reports on shape composition with preschoolers}. For the present study, we choose a central topic for kindergarten mathematics: solving arithmetical story problems. This domain has been extensively researched and, thus, has a solid empirical foundation for a detailed LT and may hold implications for the use of LTs across multiple domains {e.g.`, \Alonzo, 2012 #5442;National Research Council, 2007 #3247}. Further, informal arithmetic competence is one of the best predictors of mathematical disabilities/difficulties and later achievement in not just mathematics but also in reading {Geary, 2011 #5419;Gersten, 2005 #2731}.

The Arithmetic Learning Trajectory

The following describes the three components of our LT for arithmetic and the research that underlies them, focusing on the levels most relevant to kindergarteners{all levels are available in \Clements, 2014 #5679;, 2020 #8608;Sarama, 2009 #3380}.

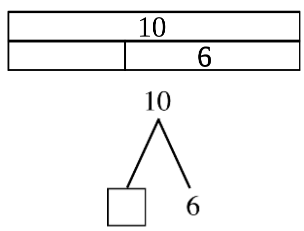
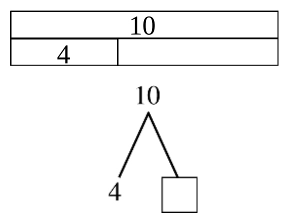
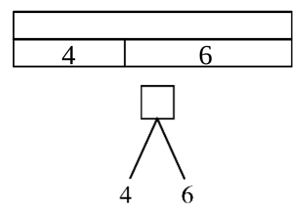
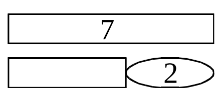
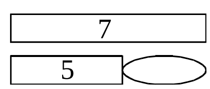
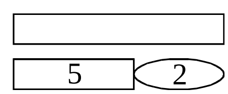
1. The goal. An overarching aim of early arithmetic goal is enabling children to understand and solve simple addition (word) problems. Children initially and informally do both in terms of counting {Ginsburg, 1977 #1154;National Research Council, 2009 #3857}. Ideally, instruction would foster children's use of a relatively efficient informal strategy. One main goal of the early arithmetic LT, then, is the verbal (abstract) counting-on strategy. For example, solving 4 + 7 by starting the count at "four" and continuing the count for 7 more numbers: 4; 5, 6, 7, 8, 9, 10, 11.

Also important is children's ability to solve different types of problems. The *type*, or *structure* of the word problem depends on the *situation* and the *unknown* determines its difficulty {Carpenter, 1992 #1921}. There are four different real-world situations (shown in the four rows of Figure 1). For each situation, the unknown can be any of the three quantities – differences in the location of this unknown quantity in part explains how difficult it is for children to model and solve these problems. Consider "Change add to (Join)" problems (row 1) in which items are added to a set. Result-unknown problems are relatively easy because they conform to children's informal change add-to view of addition (as adding more items to an existing collection to make it larger) and, thus, can be readily understood and modeled. Change unknown are more difficult than result unknown, because children need to create an initial set, then understand that they do not then create another set but instead add on to the set to create the total named. Even if they can do that, they may not have anticipated needing to keep the additional objects separate from the initial set. Thus, modeling change unknown involves more working memory demands. Start unknown are the most difficult, as there is no initial quantity stated, so “getting started” in the modeling process is especially challenging. Change take away (Separate) involving taking items away from a set and are similar in the relative difficult across the columns. Part-part-whole problems embody a more formal meaning of addition but are often assimilated to children’s informal change-add-to view of addition. Here, there is not difference in difficulty between the first and second unknowns. Finally, compare situations, regardless of the unknowns, are equally difficult {Artut, 2015 #10212;Carpenter, 1992 #1921;Fuson, 2018 #9540}. A main *goal* of the addition and subtraction LT is that children learn to solve all 12 types of arithmetic problems.

2. The developmental progression. The second component of the learning trajectory, the developmental progression, is based on many empirical studies {Baroody, 1987 #2467;Carpenter, 1992 #1921;Carr, 2011 #3473;Fuson, 1992 #2147;Fuson, 2014 #6311;Steffe, 1988 #610;Sarama, 2009 #3380;Steffe, 1988 #610;Tzur, 2019 #9541} and has been supported by many others {Clements, 2014 #5679}, including international research {Artut, 2015 #8686;Dowker, 2007 #4463;Gervasoni, 2018 #10190}.

The levels for the arithmetic learning trajectory are shown in the first column in Figure S-1 (see the online Supplemental Material). In addition to the type of problem involved (Fig. 1), the difficulty of a level is determined in part by the size of the numbers involved, which

Figure 1: Addition and Subtraction Problem Types {Carpenter, 1992 #1921;adapted from \Clements, 2014 #5679}.

Situation	First Unknown	Second Unknown	Third Unknown
Change add to (Join) A physical act of joining, or adding more items to a set, increases the number in a set.	*start unknown* □ + 6 = 11 Al had some balls. Then he got 6 more. Now he has 11. How many did he start with?	*change unknown* 5 + □ = 11 Al had 5 balls. He bought some more. Now he has 11. How many did he buy?	*result unknown* 5 + 6 = □ Al had 5 balls and gets 6 more. How many does he have in all?
Change take away (Separate) An action of separating decreases the number in a set.	*start unknown* □ - 5 = 4 Al had some balls. He gave 5 to Barb. Now he has 4. How many did he have to start with?	*change unknown* 9 - □ = 4 Al had 9 balls. He gave some to Barb. Now he has 4. How many did he give to Barb?	*result unknown* 9 - 5 = □ Al had 9 balls and gave 5 to Barb. How many does he have left?
Part-Part-Whole Two parts make a whole, but there is no physical action—the situation is static.	*first part unknown*  Al has 10 balls. Some are blue, 6 are red. How many are blue?	*second part unknown*  Al has 10 balls; 4 are blue, the rest are red. How many are red?	*whole unknown*  Al has 4 red balls and 6 blue balls. How many balls does he have in all?
Compare The numbers of objects in two sets are compared.	*smaller unknown*  Al has 7 balls. Barb has 2 fewer balls than Al. How many balls does Barb have?	*difference unknown*  Al has 7 balls. Barb has 5. How many more balls? does Al have than Barb?	*larger unknown*  Al has 5 marbles. Barb has 2 more than Al. How many balls does Barb have?

is related to the level of counting and strategic competence (along with other number knowledge, such as subitizing). In Figure S-1, “Levels/Strategies” describes what children know and can do mathematically at a particular point in the developmental progression, while “Mental Actions on

Objects" describes the hypothesized cognitive concepts and processes children deploy as they represent the structure of the different "problem types" enabling them to solve the problems {from \Sarama, 2009 #3380}. The rightmost column describes the Instruction hypothesized to help lower-level children achieve *that* level (not instruction *for* those who have already attained that level).

The research also indicates that the arithmetic LT is interwoven with the counting LT delineated in Figure S-2. That is, increasingly sophisticated arithmetic strategies often depend, at least in part, on increasingly sophisticated counting competences. Children typically start at the **1–Small Number +/-** level. That is, they initially use a concrete counting-all procedure that directly models a change-add-to meaning of addition. (Abstract addition procedures entail verbally counting to represent at least a portion the sum while simultaneously keep tracking track of how much more is being added to the first addend such as 3+5: **3**; 4 [is one more], 5 [is two more], 6 [is three more], 7 [is four more]. 8 [is five more]. Unlike abstract procedures, concrete procedures have a distinct sum count that follows the representation of the addends and thus do not require a keeping-track process.) Given a situation of 3 + 5, children at the **1–Small Number +/-** level count out 3 objects to represent the initial amount of 3 (using the **3–Producer (Small Numbers)** competencies of the counting LT), then count out 5 more items to represent adding 5 more, and finally count all the items starting at "one" to determine the new total "8." Children use such counting methods to solve story situations as long as they understand the language in the story.

Children eventually invent increasingly sophisticated shortcuts. For example, they eventually *count-on*, solving 3 + 5 by counting, "Threeeee… four, five, six, seven, eight!" Starting the with the cardinal term "three" eliminates counting from "one" up to "three" and depends on children achieving level 6 in the counting progress (**Counter from N (N+1, N-1)**) in

Figure S-2. Children eventually invent the relatively efficient abstract *counting-on-from-larger* strategy (e.g., for 3 + 5, starting with "five" and counting on only three more numbers: "5; 6, 7, 8"). See the **4–Counting Strategies +/-** level in Figure S-1.

With subtraction, children also typically start with a direct-modeling strategy, *concrete take-away* (e.g., for 9 – 5, put out nine objects, remove five, and count the remaining four to determine the difference) and, in time, move to *counting-back-from* (e.g., for 9 – 5, "Nine; eight [is one taken away], seven [is two taken away], six [is three taken away], five [is four taken away], four [is five taken away]"). However, counting backwards, especially more than two or three counts, is difficult for most children. Instead, children might learn *counting-up-to* strategy (e.g., for 9 – 5: "5; 6 [is 1 more], 7 is 2 more], 8 [is 3 more]. 9 [is 4 more]).

3. The instructional tasks. As stated, instructional tasks in the learning trajectories are not the only way to guide children to achieve the levels of thinking embedded within the learning trajectories. However, those in the last column of Figure S-1 are specific examples of the type of instructional activity that research indicates helps promote a thinking level {e.g.`, \Clements, 2014 #5679;Clements, 2020 #9997;Gervasoni, 2018 #10190;Murata, 2004 #2571}.

One of the main characteristics of the activities is the type of problem (Fig. 1) that children can solve at each level {Carpenter, 1992 #1921}. Furthermore, in many cases, there is evidence that certain aspects of the instructional tasks are especially effective. For example, research indicates that helping children discover the number-after rule for adding 1 can promote the invention of counting-on (e.g., the sum of 7 + 1 is the number after seven when we count—eight) {Baroody, 1987 #2467;Baroody, 2019 #8346}. The rule serves as a scaffold for counting-on: If 7 + 1 is the number after seven, then 7 + 2 is two numbers after seven (7; 8, 9), 7 + 3 is three numbers after seven (7, 8, 9, 10), and so forth.

Research Questions

With this study, we asked the following research question: Does instruction in which LT levels are taught consecutively (e.g.`, for children at level *n*`, instructional tasks from level *n* + 1`, then *n* + 2) result in greater learning than instruction that immediately and solely teaches the target level, *n* + 3 (aka, the "skip-levels" approach)? We also investigated whether child gender was a significant moderator of differences, due to the conflicting results of differences between girls' and boys' performance in arithmetic problem solving {Fennema, 1998 #2939;Linn, 1989 #652}. Further, given the hierarchical nature of mathematics learning {Sarama, 2009 #3380;Wu, 2011 #3385} and the importance of counting to arithmetic performance, we examined interactions of intervention condition with children's initial competence in counting and arithmetic.

The competing teach-to-target approach requires justification. Theoretically, the hypothesis is that it is more efficient and mathematically rigorous to teach the target level immediately by providing accurate definitions and demonstrating accurate mathematical procedures {see \Bereiter, 1986 #3501;Wu, 2011 #3385}, potentially obviating the need for potentially slower movement through each level. There is evidence supporting this approach to children's learning {Borman, 2003 #2082;Carnine, 1997 #2558;Clark, 2012 #4670;Gersten, 1985 #1327;Heasty, 2012 #4948}, although the research designs do not usually compare to other research-validated approaches . That is, such instruction is deemed more efficient because it skips one or more of a LT's levels (e.g., levels *n* + 1 and *n* + 2) and explicitly focuses on a target competence (*n* + 3) that is assumed to enable the student to perform tasks associated with that and all previous levels. This approach contradicts the implications of the research on learning trajectories, and thus serves as an empirically-based counterfactual for the present study.

Methods

In most of our studies in this series, we conducted pilot studies to enable project leadership to train instructors and assessors to fidelity in situ, as well as evaluate the sensitivity of our assessments (after approval from the institutional review board). Then we implemented a full-scale experiment. Building on the arithmetic pilot {Clements, 2020 #9997}, here we report the larger-scale arithmetic study.

Participants

We received permission forms from 319 students from 16 classrooms in four schools in an urban district in a Mountain West state. Of these, 28 attritted[1]; in decreasing frequency, the reasons for attrition were: the child was non-verbal, moved outside of the district (6 during the study), or demonstrated behavioral issues whereupon the teachers requested they not participate. Thus, 291 students were involved in this study. Table 1 contains school-level demographics.

Table1

Demographics of Participating Schools

School	Number of Students	Non-White Students	Male-Female Ratio	Free and Reduced Lunch	IEP Percentage
School 1	635	28.7%	53:47	3.0%	15.7%
School 2	471	52.6%	49:51	34.7%	21.3%
School 3	508	43.1%	55:44	10.1%	11.8%
School 4	347	35.4%	46:54	43.8%	8.3%

[1] The differential attrition by treatment is 0.09%, suggesting there is no difference in rates of attrition between LT and Skip condition (x^2 *(1) = 0.002, p >0.05*). Differential attrition by child gender is 3.69%, suggesting there is no difference in the rate of attrition between boys and girls (x^2 *(1) = 2.78, p >0.05*).

Intervention Conditions

In the experimental (LT) condition, instruction was based on the learning trajectories for arithmetic and counting. In the comparison (“Skip”) condition, children were presented with the opportunity to solve arithmetic story problems three levels above their level of thinking at the time of pretest ($n+3$, their “target” level). Children were randomly assigned to the LT or Skip group after pretest using a random number generator. We then established baseline equivalence in for pre-counting and pre-arithmetic prior to implementing instruction for each condition.

At least two instructors were assigned to work with children from each classroom. All instructors worked with children in both intervention conditions and (to the extent possible) with the same set of children, maintaining a pace that would enable them to achieve the goal of 15 sessions per child (180+ total minutes) by the end of the intervention. Teacher and instructor schedules required that some children had more than two instructors for a small number of sessions.

LT instruction. Instructors created opportunities for children to represent the objects, actions, and relationships that define the twelve types of arithmetic story problems {Carpenter, 1993 #1098} within the learning trajectories model {Sarama, 2009 #3380;Clements, 2014 #5679}. The intention was to support children’s progression through the arithmetic learning trajectory, with the goal of reaching three levels above each child’s pretest LT level. However, if an LT child attained that level, consistent with the LT approach, instructors presented problems at higher levels. Most sessions started with problems from the level of thinking assumed to be attained by the child (n). If the child had difficulty, more problems of that type were presented; if not, problems progressed to the next level ($n + 1$). Problem types were often presented in the form a story problem using stated interests of the child (e.g., a trip to the grocery or toy store). They provided opportunities for students to practice counting; that is, LT instructors incorporated the counting LT into instruction when children demonstrated gaps in foundational counting skills (e.g., inability to count out, or produce, sets accurately) that negatively impacted their ability to represent, reason about, and solve arithmetic problems. At higher levels of the LT, manipulatives were phased out of instruction to encourage children to use more sophisticated strategies (e.g., counting on or Break Apart to Make Ten). Scaffolds were provided throughout instruction based on what was most appropriate for each child, including (but not limited to) providing feedback, manipulatives, and instructor modeling of solution strategies.

Skip instruction. Similar to the LT instruction, instructors provided children in the Skip group with opportunities to solve story problems, using the stated interests of the child. However, the problem structures were at children's *target* level, defined as three levels higher than the child's initial level of thinking ($n + 3$). For instance, a child demonstrating mastery of the **1–Small Number +/-** LT level at pretest would receive story problems characteristic for the **3b–Find Change +/-** LT level (Fig. S-1). This counterfactual reflects the typical classroom experience during whole-group instruction, which tends to be based on a given set of standards or curricular tasks {often a misunderstanding of the implications of standards`, see \Clements, 2017 #7938}. To ensure instruction at level $n + 3$, children were not provided scaffolding strategies reflecting earlier LT levels; instead, encouragement to solve the problems and feedback, manipulatives, and instructor modeling of solution strategies were provided.

Motivational strategies for all children. Instructors in both conditions had child-friendly images which could be used to build story problems (e.g., farm animal scenario). Children were encouraged to continue working throughout the 15-20-minute session with positive and consistent instructor reinforcement appropriate for the condition. For example, instructors might say "thank you," smile, and ask him or her to explain their thinking in a friendly and conversational tone (e.g., "That's such an interesting way to solve that problem –

can you please show me how you did that with the [manipulatives] again?"). At the end of each session, instructors thanked children for their effort and gave them a sticker of their choosing.

Instructor training. The instructional team was composed of 18 graduate students (GRAs) from the College of Education (others, including the senior authors, taught when needed). GRAs were trained by the co-PIs and the Project Director {Clements, 2020 #9997} to provide instruction for both conditions. Training was comprised of descriptions of the study design and the theoretical foundation of learning trajectories for counting and arithmetic. Instructors participated in regular team meetings where the PIs and Project Directors provided didactic presentations and video clips of activity enactment. Group discussion occurred throughout the trainings to answer questions and clarify misunderstandings about the LTs and the problems that arise in and from practice.

Throughout this study, instructors learned how to use the learning trajectories as a basis for formative assessment, a key to high quality teaching {e.g.`, \National Mathematics Advisory Panel, 2008 #3480}. Formative assessment is particularly difficult for instructors to enact without substantial support {Foorman, 2007 #2806}. Thus, instructors discussed and practiced how to observe and interpret children's thinking as well as select appropriate instructional tasks for each child (e.g., modifying activities between sessions to match instructional tasks to developmental levels of individual children) in weekly professional development sessions. In addition, the PIs and Project Directors observed recorded instructions sessions weekly for each instructor and provided constructive feedback (See Fidelity of Instruction for more details).

Measures

We define counting and arithmetic competence as latent traits within an item response theory framework. Rasch scores were constructed using the R package ltm {Rizopoulos, 2006 #10095}. All items that make up the counting and arithmetic pretest and posttest are ordered by

Rasch item difficulty. All assessments were videotaped; assessment administration and coding were reviewed for accuracy. All discrepancies were resolved with the support of the PIs and Project Directors.

Counting pretest and posttest. The Counting pretest and posttest were composed of eight items. Items adapted from the *Research-Based Early Mathematics Assessment* {REMA`, \Clements, 2008/2019 #8015}and the *Test of Early Mathematics Ability – 3rd Edition* {TEMA-3`, \Ginsburg, 2007 #7304} assessed competences from ten levels of the LT, beginning with **1–Reciter** ("How high can you count? Start at 1 and tell me.") and ending with **9–Counter On Keeping Track** ("Starting at 4, please count 3 more out loud for me").

Although the items were adapted from validated instruments, we applied principal axis factoring (PAF) with varimax rotation to assess dimensionality for this and other measures used in this study. Dimensionality criteria included initial eigenvalues {Kaiser, 1960 #10098}, visual inspection of scree plots {Cattell, 1966 #10096}, variance explained by the factor(s), and parallel analysis {Horn, 1965 #10097}. PAF analysis extracted one factor and Cronbach's $\alpha = 0.78$. Since unidimensionality was established, Rasch scores were constructed. Consistent with the developmental progression, Rasch difficulty parameters suggest that beginning items (designed to measure nascent knowledge and skills) are less difficult relative to items near the end of the assessment (designed to measure more sophisticated knowledge and skills); see Table S-1. Information, an analog of reliability, was above .80 four standard deviations above and below the latent trait continuum.

Arithmetic pretest. The Arithmetic pretest was composed of 21 items similarly adapted from the REMA and TEMA-3. Items assessed competences from ten levels of the LT, beginning with **1-Small Number +/-** ("You have 2 blocks and get 1 more. How many in all?") and ending with **6-Numbers-in-Numbers +/-** ("Cat had some toys. Then she got 4 more. Now she has 12 toys. How many did she have to start with?”).

PAF analysis extracted one factor and Cronbach’s $\alpha = 0.85$. Because unidimensionality was established, Rasch scores were constructed. Information, an analog of reliability, was above .80 four standard deviations above and below the latent trait continuum.

Initial LT levels and instructional assignments. All children were assigned an initial level of thinking in Arithmetic based on accurately answering 75% or more of the items at that (and all earlier) levels. Nearly one-third of children attained the **1–Small Number +/-** level and one-fourth of children were at **3a–Make It N +/-** (Table S-2). Those who did not attain any level were assigned the foundational level in the counting LT.

As stated, the goal was for children to achieve three levels above their initial level (thus, **3b–Find Change +/-**, **6-Numbers-in-Numbers +/-**, and **8-Problem Solver +/-**; see Fig. S-1). These were defined as the *target* levels for Skip instruction and the primary goal for the LT instruction (albeit one that could be surpassed following the LT).

LT instruction necessitated a starting level for instruction. For those LT children who attained a level, instruction was started at the next-higher level (e.g., children who attained **1–Small Number +/-** began instruction at the **2–Find Result +/-** level; see Fig. S-1). However, for those LT children who did not attain a level for the arithmetic LT, instruction began at (a) the lowest arithmetic LT level with both **(1-Small Number +/-)** *and*, at the beginning of the session, (b) one level above the counting LT level following the one they attained at pretest. Most of these LT children were at the **1–Reciter** level (Table S-2) of the counting trajectory, so they began instruction at the next level, **2–Counter (Small Numbers)** (Fig. S-2).

Arithmetic posttest. Thirteen items were added from the REMA and TEMA-3 to the Arithmetic pretest to construct the posttest. Importantly, we included more advanced items from the LT, extending up to **8-Problem-Solver +/-**, multidigit (e.g., “Mary had some marbles. She gave 49 marbles to Mark. Now Mary has 41 marbles. How many marbles did she start with?”).

PAF analysis extracted 2 factors, based on comparison initial eigenvalues with eigenvalues that simulated from parallel analysis. However, we decided to use the unidimensional solution for four reasons. First, visual inspection of scree plots (Fig. S-3) suggests eigenvalues before the “elbow” – or point where values level off – should be considered. Second, the proportion of variance accounted for by the unidimensional model was 27.57%; adding a second factor would only account for an additional 6.97%. Third, the items are derived from assessments where content validity and psychometric functioning is well-documented. Fourth, the parallel analysis is conservative and not the only way to determine the factor solution. Thus, taken together, we decided to go with the unidimensional solution, where Cronbach’s $\alpha = 0.91$. Rasch scores were constructed and again, consistent with the developmental progression, Rasch difficulty parameters suggest that beginning items are less difficult relative to items near the end of the assessment; see Table S-3. Information, an analog of reliability, was above .85 four standard deviations above and below the latent trait continuum.

Procedure

One-on-one sessions were conducted in available spaces based on staff schedules and preferences. Following each instructional session, instructors filled out a tracking file for each child with whom they had worked. Information included (but was not limited to): (a) the content of the lesson; (b) the most sophisticated arithmetic problem type the child was able to solve along with the range of numbers presented (often in the form of an equation; e.g., “$x + 3 = 11$”) and most sophisticated counting skill demonstrated, if addressed in instruction; (c) the child’s accuracy (e.g., correctly answered 3 out of 5 Change add to, result unknown problems); and (d) implications of these for instruction in the subsequent session (e.g., specific arithmetic problems that were “stuck points” or moving to a new level). This allowed LT instructors a way to coordinate instructional content and differentiate support for each child in the LT condition as well as to recall preferences (e.g., for contexts) for all children. It also supported clear communication between instructors.

Fidelity of Implementation

We systematically tracked two components of fidelity of implementation: dosage and adherence. We assessed dosage by documenting the total number of minutes children spent in instruction for each condition (and used as a model covariate). For instance, in a 15-minute session, if a child wanted to share a story about how his/her family celebrated his/her grandmother's birthday for 5 minutes (the intended "saying hello, getting ready" time was 3 minutes), dosage was computed and documented as 10 minutes. At the onset of the intervention, researchers aimed to provide children with 240 minutes of total instruction, which amounted to twenty 15-minute sessions. Due to unplanned circumstances that come with working in schools during an academic year, we were unable to meet this goal for every child. The average number of minutes of instructional time for the LT students was 206 minutes (SD =35.2) in an average of 13.4 sessions (SD = 2.34) and for Skip students, 212 minutes (SD = 34.1) in an average of 14.3 sessions (SD = 2.27). The difference in dosage between the two groups was non-significant. However, students in the Skip condition had 0.9 more instructional sessions on average (95% CI, 1.43, 0.37), which is statistically significant at α = .05.

The extent to which each instructor adhered to the principles of LT and Skip instruction was examined by the PIs and Project Directors every week of the intervention through a review of both videos and an online shared document in which each instructor documented what they did with each child and why. Suggestions or corrections were sent to instructors on e-mail, followed by conversations if requested.

Analytic Approach

Missing data for all variables were unrelated to treatment or control group status. All IRT-scores were grand-mean centered and transformed into a z-score. All models used full information maximum likelihood estimation to adjust for potential bias in the estimates resulting from missing data.

The research question was examined within a Bayesian hierarchical linear modeling (HLM) framework using the brms package {Bürkner, 2018 #10245} in R 3.6.2 {R Core Team, 2019 #10271} Bayesian models have become increasingly popular with the introduction of user-friendly open-source software. Compared to traditional models, Bayesian models provide more information about model parameters by estimating posterior distributions as opposed to only point estimates {e.g.`, \McElreath, 2016 #10272}, correctly quantify and propagate uncertainty {e.g.`, \Kruschke, 2014 #10273}, and are able to estimate models which would otherwise fail {Eager, 2017 #10274}.

We define posttest arithmetic ability, expressed as a Rasch score, as the dependent variable. The baseline model was specified as the effect of treatment (LT versus Skip) as well as pre-counting and pre-arithmetic ability and contained a random intercept for classroom and instruction teams assigned to each school. Demographic metrics were different between schools (e.g., percent of children who qualified for free-/reduced-lunch; see Table 1). However, the number of schools did not justify its inclusion as a random effect given the probability of a negative variance increases if there are too few levels of a variable {Stroup, 2012 #10275}.

Priors used were neither informative nor uninformative but were instead weakly-informative. In selecting weakly-informative priors we deliberately increase the uncertainty in model parameters versus what is known, but avoid priors with infinite variances, as would be typical for uninformative priors, for example. Furthermore, weakly-informative priors have been recommended by several Bayesian practitioners as being an attractive alternative between uninformative and informative priors {e.g.`, \McElreath, 2016 #10272}.

The final model was selected using the Watanabe-Akaike Information Criterion {Watanabe, 2010 #8509}. Child sex and intervention dosage (expressed in minutes) were added to the baseline model and examined to be predictors of arithmetic learning. Each variable was added sequentially and tested based on their contribution to model fit (as measured by the WAIC) compared with the previous, less complex model. We favored parsimonious models with the smallest WAIC to select for robustness and out-of-sample predictive performance. Table 3 depicts the order in which versions of the model were tested, along with WAIC and Bayesian R^2 {Gelman, 2019 #1982}.

Results

Descriptive Statistics

At pretest, LT and Skip children had similar levels of counting competences (Table 2). Additionally, LT children had slightly higher pretest arithmetic scores relative to their Skip peers, although this difference was not significant. The correlation between the pre-Counting and pre-Arithmetic is $r = 0.67$. At posttest, when more arithmetic items were added to preclude a ceiling effect, LT children had higher scores relative to their Skip peers. The difference between these two means measures average growth due to intervention.

Table 2.

Average IRT Scores for Pretest and Posttest Counting and Arithmetic by Intervention Condition

		Counting			**Arithmetic**	
		Pre-	**Posttest**		**Pretest**	**Posttest**
LT Condition	$n = 143$	-0.0138 (0.0687)	0.0696 (0.0674)	$n = 143$	0.0647 (0.0756)	0.3680 (0.0670)

Skip Condition	*n* =148	0.0323 (0.0721)	-0.0765 (0.0780)	*n* = 148	0.0024 (0.0708)	-0.2991 (0.0786)

Baseline equivalence was examined between the LT and Skip groups and was established for both the counting and arithmetic assessments. For counting, Cohen's *d* was an acceptable value of .05; for arithmetic *d* was slightly greater, at .07 (both statistically non-significant), but all analyses employed statistical adjustments required to satisfy baseline equivalence {IES, 2019 #10083}.

Overall Treatment Effects

Table 3.

Fit Indices for Model Selection based on WAIC and Bayesian R^2 (95% Credible Intervals).

	WAIC	**Effective Parameters**	**Bayesian R^2**
Baseline Model	478.1	10.5	0.677 (0.638, 0.708)
Baseline Model + Child Sex	480.5	11.6	0.670 (0.635, 0.709)
Baseline Model + Dosage	470.3	12.4	0.687 (0.650, 0.717)
Baseline Model + Condition x Pre-Arithmetic	473.2	12.3	0.685 (0.645, 0.716)
Baseline Model + Condition x Pre-Counting + Pre-Arithmetic	476.9	11.7	0.681 (0.641, 0.713)
Baseline Model + Condition x Pre-Arithmetic x Pre-Counting	472.0	15.2	0.692 (0.655, 0.722)
Condition x Pre-Arithmetic x Pre-Counting (No random effects)	469.6	10.8	0.689 (0.652, 0.718)

The final model included: pretest counting ability (expressed as a Rasch score), pretest arithmetic (expressed as a Rasch score), treatment condition, and their three-way interaction (see row titled “Condition x Pre-Arithmetic x Pre-Counting – No Random Effects” in Table 3). Notably, the random intercepts for classroom and instructor team were removed from the final model because this lowered the WAIC. A formal comparison of the baseline versus the final model produced $\Delta WAIC = 8.74$ $(SE = 8.24)$, which indicates a one standard error improvement in WAIC from the baseline to the final model.

HLM analyses are presented in Table 4. Although we report the random effects in Table 4, our final model excludes them because intra-class correlations were nearly zero. 95% Credible Intervals were estimated for child gender and dosage (expressed as the number of minutes spent in instruction). However, these were found to include zero, and therefore deemed to be non-significant. The magnitude of the difference between the LT and Skip conditions at posttest is considered large (*d* = 1.20; the main effect of intervention condition in the baseline model).

Table 4

Model Parameters for Post-Arithmetic including Three-Way Interaction with Random Effects for Classroom and Instructor Team

	Est.	**SE**	**95% CI (Lower)**	**95% CI (Upper)**
Intercept	-0.21	0.06	-0.33	-0.09
Pre-Arithmetic	0.61	0.06	0.49	0.73
Pre-Count	0.23	0.06	0.12	0.34
Treatment (Skip is reference)	0.53	0.08	0.37	0.68
Pre-Arithmetic x Pre-Count	-0.12	0.05	-0.21	-0.03
Pre-Arithmetic x Treatment	-0.21	0.09	-0.37	-0.04
Pre-Count x Treatment	0.03	0.09	-0.14	0.20
Treatment x Pre-Arithmetic x Pre-Count	0.17	0.06	0.04	0.30
Classroom Random Intercept (SD)	0.06	0.04	0	0.15
Instructor Team Random Intercept (SD)	0.06	0.05	0	0.19
Residual Error	0.53	0.02	0.49	0.58
R-Squared	0.69	0.02	0.65	0.72

Additionally, the three-way interaction between counting pretest, arithmetic pretest, and treatment condition was statistically significant (95% CI: 0.03, 0.30), averaged across classrooms and instructional teams (Table 4). We disambiguate this interaction in Figure 2, where we show the treatment effect for 9 different values of counting and arithmetic pretest scores. The LT intervention had a positive impact compared to the Skip intervention on posttest arithmetic regardless of baseline knowledge, significantly greater for eight of the nine cells. The exception was the cell of children whose pretest scores were high in arithmetic and low in counting, which showed the smallest treatment effect (95% CI: -0.47, 0.51). In the adjacent cell in Figure 2, children with high pretest arithmetic scores and average scores in counting learned more from the LT approach with a small, yet statistically significant treatment effect (95% CI: 0.04, 0.54).

Five cells had moderate treatment effects ranging from 0.51 (0.24, 0.75) to 0.59 (0.15, 1.01). The final two cells, in the bottom row of Figure 2, showed the greatest impacts: 0.96 (0.74, 1.20) for children who initially had low scores in both arithmetic and counting and 0.78 (0.54, 1.20) for those with low arithmetic and average counting scores.

The Impact of Possible Moderators

Findings did not vary by the assigned instructional team, child gender, or dosage, indicating a robust and general result. Between-classroom and between-instructor team intra-class correlation coefficients were very low: 0.01 (0.00, 0.06) and 0.01 (0.00, 0.08), respectively, further suggesting that posttest scores did not vary with classroom or instructional team. Our final model fits well, explaining 69% (65%, 72%) of the variability in posttest arithmetic scores. To evaluate the robustness of our final model, we performed a prior sensitivity analysis (Table S-4) and a graphical posterior predictive check (Fig. S-4, S-5 and S-6) {Gabry, 2019 #10276}. Our post-hoc analysis revealed no sensitivity to prior specification and no appreciable lack-of-fit either for the sample overall, or by classroom, or by instructor team.

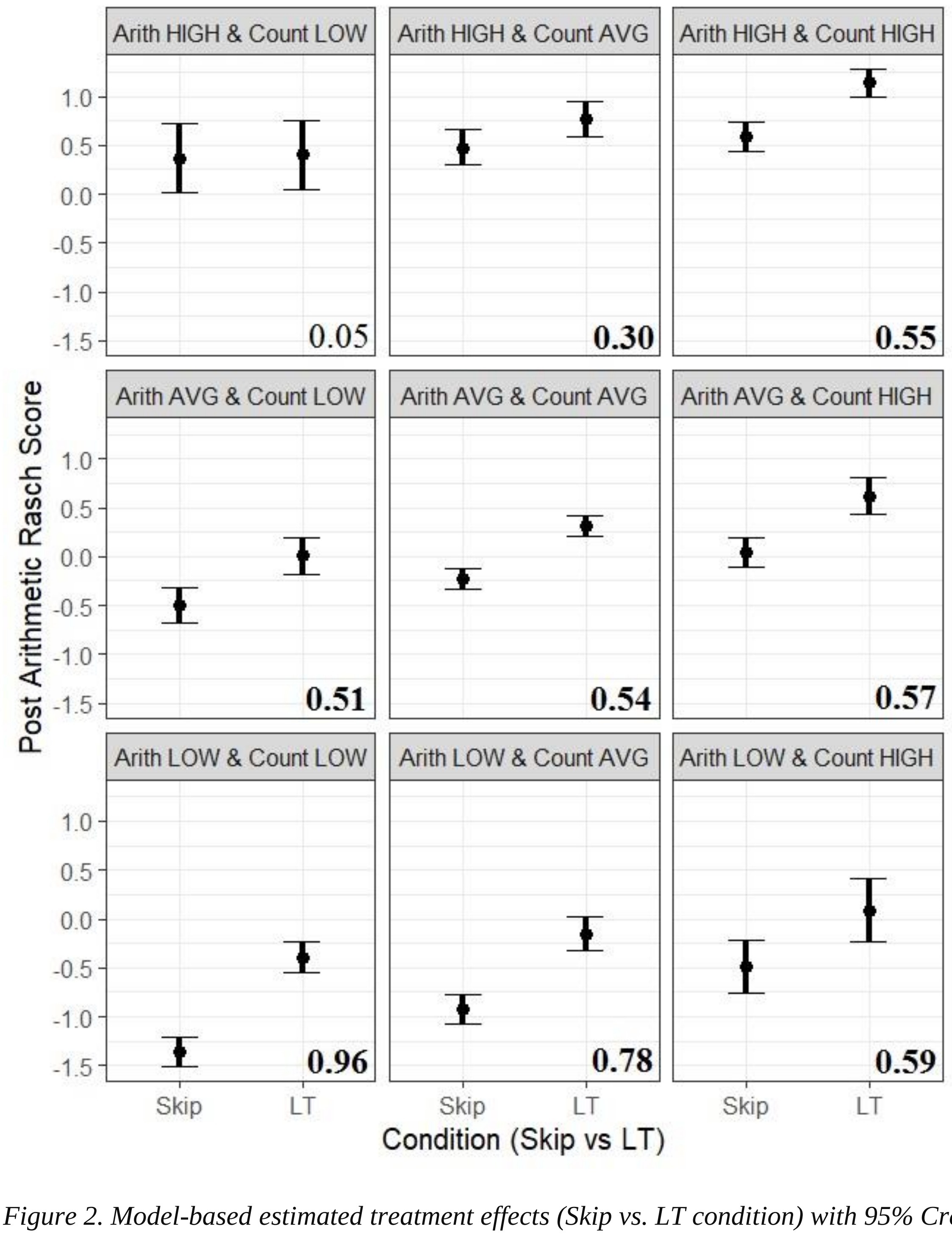


Figure 2. Model-based estimated treatment effects (Skip vs. LT condition) with 95% Credible Intervals at nine levels of baseline knowledge for Arithmetic (Arith) and Counting (Count). HIGH levels of knowledge indicate children are 1 standard deviation above the population average; LOW levels indicate children are 1 standard deviation below the population average; AVG levels are equal to the population average. Each panel is labelled with the treatment effect that is ***bolded*** *if the treatment effect is statistically significant at alpha = 0.05.*

Intervention Impact on the Target Level

Examination of individual items confirmed that the LT group made more completely correct solutions to each and every of the test items compared to the Skip group. At posttest, the LT group (46.18%) had a significant higher correctness rate than SKIP group (30.27%), $\chi^2 = 7.781, df = 1, r = 0.16$ (Campbell, 2007; Richardson, 2011). In fact, the LT group outperformed the SKIP group based on every item: the item mean correctness difference was .16 (range = .01 to .39, range of SD = .08 to .76), with corresponding effect size of .36 (range of .00 to .84). This is notable, as the target level of thinking was achieved more frequently by LT children who experienced *fewer* tasks at that level.

For example, there were 93 LT children and 114 Skip children whose *n*, or level of thinking prior to the intervention, was categorized as the most basic arithmetic level (**1–Small Number +/-**). Given that children in the Skip condition spent their instructional sessions practicing solving $n + 3$ problems, such as change unknown problems (e.g., $4 + x = 7$), we examined performance between the two intervention conditions for this specific problem-type. At posttest, nearly half of children (49.5%, $n = 46$) in the LT condition determined the correct answer relative to 26.3% (n = 30) of their SKIP peers. This difference was significant, $\chi^2 = 11.806, df = 1, p = 0.0006$ (Campbell, 2007; Richardson, 2011).

Discussion

The present study is one of the first to test directly and rigorously the specific contributions of LTs to mathematical learning {e.g.`, \Clements, 2019 #9686;Clements, 2020 #9997}. In this experiment, we designed sequences of instruction that consecutively targeted thinking one level beyond that of a child and evaluated whether this approach is more efficacious relative to instruction that immediately and solely teaches the targeted thinking several levels higher.

Summary

Children benefited from one-on-one instructional sessions, regardless of intervention condition. However, as indicated by a large effect size (*d* = 1.20), LT instruction that occurred one level above a kindergartner's existing level of thinking, determined at each instructional session, yielded greater overall arithmetic learning relative to instruction that occurred three levels above a peer's pretest level (even though random factors lead to the latter getting almost 1 more instructional session on the average).

There was a differential effect of the interventions based on pretest arithmetic and also pretest counting knowledge {the relationship between counting and arithmetic is consistent with the research literature`, e.g.`, \Baroody, 1987 #2467;Carpenter, 1992 #1921;Fuson, 1992 #2147;Sarama, 2009 #3380;Steffe, 1988 #610;Tzur, 2019 #9541}. As indicated by a large effect size, the LT approach had the greatest relative impact for those children who started the intervention with arithmetic and counting competencies one standard deviation below the sample mean. LT children with low initial arithmetic and average counting skills demonstrated significant and (as indicated by the size of the treatment effect) the second-greatest growth compared to their Skip counterparts. As indicated by a moderate treatment effect, LT instruction had a more modest, but still substantial, positive effect on participants with low initial arithmetic and high counting skills and those with average initial arithmetic knowledge regardless of counting skill level.

The impact of LT instruction for those children who started the program with high arithmetic knowledge was mixed. As indicated by a negligible size of the treatment effect, LT children who were low in counting were not different statistically from their Skip peers. As indicated by a small or moderate treatment effect, those who were high in arithmetic and average in counting and those high in both learned more from the LT than the Skip approach. Overall, then, the LT approach—as opposed to moving directly to the target level—appears particularly productive for those with the lowest levels of entry competencies and most in need of remedial instruction on early, foundational levels of thinking.

Within the same starting arithmetic level, why might the relative effect of the LT approach vary by initial counting ability? Among children with low initial arithmetic knowledge, the LT approach may have had a more modest impact with those of high counting ability (than with those of lower levels of counting skill), because the arithmetic and counting LTs are mutually supportive and merge at higher levels. Put differently, Skip children who started with initially low arithmetic and high counting competencies may have used the latter skills to make sense of and solve the more sophisticated problems at their target level. For example, the ability to count from a number other than "one" a specific number of times (e.g., start with five and count four more numbers; level **9–Counter On Keeping Track)** is a necessary component of solving to solve arithmetic problems by means of abstract counting-on in levels at and above **{Level 4–Counting Strategies +/- by counting on \Clements, 2014 #5679;Sarama, 2009 #3380}**. Such connections are substantiated by empirical results showing counting is a strong predictor of later arithmetic {Kolkman, 2013 #5145;Koponen, 2013 #5390} and cultivated through the development of counting {Friso-van den Bos, 2018 #10279;Le Corre, 2007 #3759;Lipton, 2005 #2834}.

There may also be good reasons why the effects of the LT intervention on children with high starting arithmetic scores were mixed. The LT instruction of low counters focused on counting competencies, so less time was available for arithmetic concepts and procedures. The starting competencies in counting and arithmetic of LT children with moderate counting ability allowed their instructors to move more quickly through the developmental levels and spend more time on arithmetic instruction. This would be especially true of LT children with high initial achievement in both counting and arithmetic, who then received problems at higher (n + 4 levels). (A caveat must be noted: It is possible that some classroom teachers using instruction similar to our Skip intervention would also notice children had achieved these targets and would present more challenging problems. This raises the possibility that the finding for this cell may be partially an artifact of our research design, which taught the target level consistently.)

For all these analyses, results did not vary by the assigned instructional team, child gender, or dosage. This indicates robust and general results.

Beyond growth in children's knowledge, the interventionists' qualitative field notes show a clear indication that the Skip group expressed more counter-productive frustration than the LT group. This may indicate that instruction several levels beyond a child's current developmental level is not only less effective, but also counter-productive as it may increase a child's aversion to mathematics.

Limitations

The findings from this study should be interpreted in light of six limitations. First, a convenience sampling approach was used, such that the selected school district solicited interested administrators who volunteered their staff to participate in the study. Future research might target a nationally representative sample.

Second, dosage by student (the unit of randomization) varied. Analyses indicated that students who fell into the low arithmetic/low counting group at pretest received the most instruction on average (i.e., 14.9 sessions, 228.5 minutes for 16 children) compared to all other groups, particularly the high arithmetic/high counting group (i.e., 11.8 sessions, 183.0 minutes for 22 children). Although results indicated that these differences did not significantly impact the efficacy of the intervention, is it possible that with equal instructional time across schools and students, we may have seen more growth in those students who performed well at pretest.

Third, we administered a mid-assessment to children in the LT condition to assess student progress in the LT condition and determine instructional needs for the second half of the intervention. However, we did not administer the mid-assessment to children in the Skip condition because this would present problems at all levels. The items that made up the mid-assessment were the same as the posttest although they contained start and stop rules (e.g., stop after 3 incorrect responses). Administering a mid-assessment served to determine whether the updated version of the assessment was (a) sensitive to growth, as well as (b) contained items difficult enough to prevent a ceiling effect. However, some students in the LT condition at School A received a mid-assessment only a handful of weeks before receiving the posttest at the request of the school. As a result, LT students were exposed to some of the assessment items one more time compared to Skip students. However, these items were quite similar to the intervention items all children received.

Fourth, based on the findings from the pilot study in Fall 2017, training for instructors focused on the earlier developmental levels in the arithmetic LT. More specifically, training emphasized instruction for the following LT levels: **1–Small Number +/-** through **6–Numbers-in-Numbers +/-** within 30. However, once we pre-assessed students, we found that a portion of students in the LT condition already were demonstrating mastery at the higher LT levels. As can be seen in Table S-2, 26% of children had pre-mastery levels at **3a–Make It N +/-**. Consequently, instructors needed to implement more advanced instruction for which they did not necessarily have initial training. Therefore, the principal investigators provided training as needed for those specific instructors.

Fifth, we were not able to examine whether there was a differential impact of intervention efficacy by child- or family-level demographics. Although child sex did not interact with the effect of treatment, other studies suggest demographic characteristics, parental characteristics, and the home environment to be potentially moderating covariates on academic outcomes {e.g.`, \Bradley, 2001 #10277}. District leadership changed (unexpectedly) and we were no longer granted the same level of access to child and family demographic information.

Sixth, instruction was one-on-one. Although the same for both treatment groups, generalization to classroom instruction should be made with caution.

Implications for Theory, Research, and Practice

Teaching contiguous levels of a learning trajectory was more efficacious than the teach-to-the-target (Skip) approach. This supports the LT assumption that there are valuable learnings in each level of a developmental progression that best not be skipped and that each level is built upon the foundation of the earlier levels of thinking. Consistent with Vygotsky's construction of the zone of proximal development {Vygotsky, 1935/1978 #2610}, the LT approach involves using *formative assessment* {National Mathematics Advisory Panel, 2008 #3480;Shepard, 2018 #8673} to provide instructional activities aligned with such empirically-validated developmental progressions {Clarke, 2001 #2057;Fantuzzo, 2011 #4529;Gravemeijer, 1999 #1412;Jordan, 2012 #5144} and using teaching strategies that evoke children's natural patterns of thinking at each level, as posited by hierarchical interactionalism {Sarama, 2009 #3380}. This approach appears particularly productive for those with the lowest levels of entry competencies, specifically for children with low initial arithmetic and either low or average initial counting competencies. This similarly indicates the importance of supporting children's learning of each level of the LT, as children may not be able to make sense of tasks from higher levels if they have not built the concepts and procedures that constitute prior levels of thinking, supporting the tenets of hierarchical interactionalism. Children with low entering competencies may be especially at risk of learning only to apply rote, prescribed procedures {"reduction of level" according to \van

Hiele, 1986 #39} to sophisticated problems under teach-to-target instruction.

The results have additional implications regarding exposure and the amount of exposure. Specifically, the overall results call into question a basic assumption of the teach-to-target instruction approach. According its proponents, such instruction is more effective and efficient because targeting high-level concepts and skills enables a student to learn those of earlier levels as well {e.g.`, \Carnine, 1997 #2558;Clark, 2012 #4670;Clements, 2014 #5679;Wu, 2011 #3385}. In fact, the LT participants who were exposed to a greater variety of levels (e.g., problem types and number ranges), including those below target-level instruction, performed significantly and substantively better than Skip children. In brief, although some students—especially those with high levels of relevant knowledge already—may spontaneously learn non-targeted lower concepts and skills, it cannot be taken for granted that many or even most students will do so.

When instruction is meaningful (i.e., ensures and builds on more basic knowledge), the amount of exposure needed for learning can be less than instruction that does not do so. At posttest, a greater proportion of LT children responded correctly to target-level problem types despite less exposure than the Skip participants. These results provide particularly cogent support our hypothesis: instruction that helps children learn each successive level of thinking along a research-based developmental progression is more efficacious than instruction that directly teaches a target level without addressing intermediate levels, even on the teach-to-target's problem types.

Therefore, the findings have several implications for practice. All children benefited somewhat from one-on-one instructional sessions, both those receiving learning trajectories-based (LT) instruction and teach-to-target (or "skip-levels") instruction. However, LT instruction led to greater learning of arithmetic overall and on find change problems (targeted by both interventions) in particular. This finding is significant not only because these problem types are linguistically more complex, but also because LT children spent significantly *less* time working with these problem types during instruction. This finding mirrors previous findings of this project {Clements, 2019 #9686;Clements, 2020 #9997}, supporting the LT approach as opposed to an ostensibly more "efficient" approach of directly teaching target skills. As noted, a limitation is that one-on-one instruction might not generalize to classroom instruction; however, this is a theoretical study that suggests what characteristics of LT instruction may account for the success of multiple classroom LT interventions {e.g.`, \Clarke, 2001 #2057;Clements, 2008 #2785;Clements, 2011 #4177;Kutaka, 2017 #8189;Murata, 2004 #2571;Wright, 2006 #2868}.

The findings also indicated that LT instruction had the greatest relative positive impact for those children who started the intervention with the lowest counting and arithmetic skills or had average counting skills. LT instruction had a lesser (but still positive) relative impact for those children who started the intervention with high arithmetic knowledge. Although following a development progression is not necessary–children in both groups learned– the LT approach appears beneficial for most students and strongly indicated for those with lower entry levels in both counting and arithmetic compared to a teach-to-target approach.

Finally, future research could use other designs, such testing this key assumption of the LT approach while controlling for exposure or practice by comparing LT (n + 1 training of a n-level child) with Skip training a control child who started at n – 1 or lower.

References

Alonzo, A. C., & Gotwals, A. W. (2012). Leaping forward: Next steps for learning progressions in science. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning Progressions in Science* (pp. 475–490): Sense Publishers.

Artut, P. D. (2015a). Preschool children's skills in solving mathematical word problems. *Educational Research and Reviews, 10*(18), 2539-2549. doi: 10.5897/ERR2015.2431

Artut, P. D. (2015b). Preschool children's skills in solving mathematical word problems. *Educational Research and Reviews, 10*(18), 2539–2549. doi: 10.5897/ERR2015.2431

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*, 141–157.

Baroody, A. J., Clements, D. H., & Sarama, J. (2019). Teaching and learning mathematics in early childhood programs. In C. P. Brown, M. B. McMullen & N. File (Eds.), *The Wiley Handbook of Early Childhood Care and Education* (1st ed., pp. 329–353). Hoboken, NJ: Wiley Blackwell Publishing.

Baroody, A. J., & Purpura, D. J. (2017). Number and operations. In J. Cai (Ed.), *Handbook for research in mathematics education* (pp. 308–354). Reston, VA: National Council of Teachers of Mathematics (NCTM)

Bereiter, C. (1986). Does direct instruction cause delinquency? Response to Schweinhart and Weikart. *Educational Leadership*, 20–21.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125–230. doi: 10.3102/00346543073002125

Bradley, R. H., Burchinal, M. R., & Casey, P. H. (2001). Early intervention: The moderating role of the home environment. *Applied Developmental Science, 5*(1), 2-8.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R Package brms. *The R Journal, 10*(1), 395-411. doi: 10.32614/RJ-2018-017

Carnine, D. W., Jitendra, A. K., & Silbert, J. (1997). A descriptive analysis of mathematics curricular materials from a pedagogical perspective: A case study of fractions. *Remedial and Special Education, 18*(2), 66–81.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E. H., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education, 24*(5), 428–441. doi: 10.2307/749152

Carpenter, T. P., & Fennema, E. H. (1992). Cognitively guided instruction: Building on the knowledge of students and teachers. *International Journal of Educational Research, 17*(5), 457–457. doi: 10.1016/S0883-0355(05)80005-9

Carr, M., & Alexeev, N. (2011). Developmental trajectories of mathematic strategies: Influence of fluency, accuracy and gender. *Journal of Educational Psychology, 103*, 617–631.

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 6–11.

Clarke, B. A. (2008). A framework of growth points as a powerful teacher development tool. In D. Tirosh & T. Wood (Eds.), *Tools and processes in mathematics teacher education* (pp. 235–256). Rotterdam: Sense.

Clarke, D. M., Cheeseman, J., Clarke, B., Gervasoni, A., Gronn, D., Horne, M., . . . Sullivan, P. (2001). Understanding, assessing and developing young children's mathematical thinking: Research as a powerful tool for professional growth. In J. Bobis, B. Perry & M. Mitchelmore (Eds.), *Numeracy and beyond (Proceedings of the 24th Annual Conference* *of the Mathematics Education Research Group of Australasia, Vol. 1)* (pp. 9–26). Reston, Australia: MERGA.

Clements, D. H., Fuson, K. C., & Sarama, J. (2017). The research-based balance in early childhood mathematics: A response to Common Core criticisms. *Early Childhood Research Quarterly, 40*, 150–162.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*, 81–89. doi: 10.1207/s15327833mtl0602_1

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*(2), 136–163. doi: 10.2307/748360

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443–494. doi: 10.3102/0002831207312908

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach* (2nd ed.). New York, NY: Routledge.

Clements, D. H., & Sarama, J. (2020). Learning and teaching with learning trajectories ([LT]^2). Retrieved from Marsico Institute, Morgridge College of Education, University of Denver website: www.learningtrajectories.org

Clements, D. H., Sarama, J., Baroody, A. J., & Joswick, C. (2020). Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction. *ZDM Mathematics Education*. doi: 10.1007/s11858-019-01122-z

Clements, D. H., Sarama, J., Baroody, A. J., Joswick, C., & Wolfe, C. B. (2019). Evaluating the efficacy of a learning trajectory for early shape composition. *American Educational Research Journal, 56*(6), 2509–2530. doi: 10.3102/0002831219842788

Clements, D. H., Sarama, J., & DiBiase, A.-M. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Erlbaum.

Clements, D. H., Sarama, J., & MacDonald, B. L. (2019). Subitizing: The neglected quantifier. In A. Norton & M. W. Alibali (Eds.), *Constructing number: Merging perspectives from psychology and mathematics education* (pp. 13–45). Gateway East, Singapore: Springer.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127–166. doi: 10.5951/jresematheduc.42.2.0127

Clements, D. H., Sarama, J., Wolfe, C. B., & Day-Hess, C. A. (2008/2019). *REMA—Research-based Early Mathematics Assessment* Denver, CO: Kennedy Institute, University of Denver.

Confrey, J. (2019). A synthesis of research on learning trajectories/progressions in mathematics. Retrieved from Organisation for Economic Co-operation and Development website: http://www.oecd.org/education/2030/A-Synthesis-of-Research-on-Learning-Trajectories-Progressions-in-Mathematics.pdf

Daro, P., Mosher, F. A., Corcoran, T. B., Barrett, J., Battista, M. T., Clements, D. H., . . . Sarama, J. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. New York, NY: Consortium for Policy Research in Education (CPRE).

Dowker, A. (2007). What can intervention tell us about the development of arithmetic? *Educational & Child Psychology, 24*(2), 64–82.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. Retrieved from https://arxiv.gg363.site/pdf/1701.04858.pdf

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start Children. *American Educational Research Journal, 48*, 763–793.

Fennema, E. H., Carpenter, T. P., Franke, M. L., & Levi, L. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher, 27*, 6–11.

Foorman, B. R., Santi, K. L., & Berger, L. (2007). Scaling assessment-driven instruction using the Internet and handheld computers, Volume 2. In B. Schneider & S.-K. McDonald (Eds.), *Scale Up in Practice* (pp. 69–89). Lanham, MD: Rowman & Littlefield.

Friso-van den Bos, I., Kroesbergen, E., & Van Luit, J. E. H. (2018). Counting and number line trainings in kindergarten: Effects on arithmetic performance and number sense. *Frontiers in Psychology*, *9*, 975. doi: 10.3389/fpsyg.2018.00975

Frye, D., Baroody, A. J., Burchinal, M. R., Carver, S., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag.

Fuson, K. C. (1992). Research on learning and teaching addition and subtraction of whole numbers. In G. Leinhardt, R. Putman & R. A. Hattrup (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 53–187). Mahwah, NJ: Erlbaum.

Fuson, K. C. (2004). Pre-K to grade 2 goals and standards: Achieving 21st century mastery for all. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 105–148). Mahwah, NJ: Erlbaum.

Fuson, K. C. (2018). Building on Howe's three pillars in kindergarten to grade 6 classrooms. In Y. Li, W. J. Lewis & J. J. Madden (Eds.), *Mathematics matters in education: Essays in honor of Roger E. Howe* (pp. 185–207). Cham: Springer International Publishing.

Fuson, K. C., & Li, Y. (2014). Learning paths and learning supports for conceptual addition and subtraction in the US common core state standards and in the Chinese standards *Mathematics Curriculum in School Education* (pp. 541–558): Springer.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Royal Statistical Society Series A (Statistics in Society), 182*, 389-402. doi: 10.1111/rssa.12378

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. doi: 10.1037/a0025510

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician, 73*(3), 307-309.

Gersten, R. (1985). Direct instruction with special education students: A review of evaluation research. *The Journal of Special Education, 19,* 41–58.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematical difficulties. *Journal of Learning Disabilities, 38,* 293–304.

Gervasoni, A. (2018). The impact and challenges of early mathematics intervention in an Australian context. In G. Kaiser, H. Forgasz, M. Gravenm, A. Kuzniak, E. Simmt & B.

Xu (Eds.), *13th International Congress on Mathematical Education* (pp. 115-133). Cham: Springer International Publishing.

Ginsburg, H. P. (1977). *Children's arithmetic*. Austin, TX: Pro-ed.

Ginsburg, H. P., & Baroody, A. J. (2007). *Tema 3 Test de Competencia Matemática Básica (Test of Early Mathematics Ability)* (3rd ed.). Madrid, Spain: TEA Ediciones, S. A.

Gravemeijer, K. P. E. (1994). Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education*, *25*, 443–471.

Gravemeijer, K. P. E. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, *1*, 155–177.

Heasty, M., McLaughlin, T. F., Williams, R. L., & Keenan, B. (2012). The effects of using direct instruction mathematics formats to teach basic math skills to a third grade student with a learning disability *Academic Research International*, *2*(3), 382–387.

Hickendorff, M., van Putten, C. M., Verhelst, N. D., & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology, 102*(2), 438–452. doi: 10.1037/a0018177

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica, 30*(2), 179–185.

IES. (2019). Baseline equivalence. *WWC Standards Brief*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology, 104*(3), 647–660. doi: 10.1037/a0029018

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95–103.

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162–175. doi: 10.1037/a0029285

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial introduction with R* (2nd ed.). Burlington, MA: Academic Press.

Kutaka, T. S., Ren, L., Smith, W. M., Beattie, H. L., Edwards, C. P., Green, J. L., . . . Lewis, W. J. (2016). Examining change in K-3 teachers' mathematical knowledge, attitudes, and beliefs: The case of primarily math. *Journal of Mathematical Behavior, 19*(5). doi: 0.1007/s10857-016-9355-x

Kutaka, T. S., Smith, W. M., Albano, A. D., Edwards, C. P., Ren, L., Beattie, H. L., . . . Stroup, W. W. (2017). Connecting teacher professional development and student mathematics achievement: A 4-year study of an elementary mathematics specialist program. *Journal of Teacher Education, 68*(2), 140–154. doi: 10.1177/0022487116687551

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395–438.

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher, 18*(8), 17–27.

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*, 978–988.

Maloney, A. P., Confrey, J., & Nguyen, K. H. (Eds.). (2014). *Learning over time: Learning trajectories in mathematics education*. New York, NY: Information Age Publishing.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press.

Murata, A. (2004). Paths to learning ten-structured understanding of teen sums: Addition solution methods of Japanese Grade 1 students. *Cognition and Instruction, 22*, 185–218.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

National Research Council. (2007). *Taking science to school: Learning and teaching sciences in grades K-8*. Washington, DC: National Academies Press.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academy Press.

NGA/CCSSO. (2010). *Common core state standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*(2), 453-464. doi: 10.1037/a0031753

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Shepard, L., & Pellegrino, J. W. (2018). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and Practice*, *37*(1), 52–57.

Spaepen, E., Gunderson, E. A., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2018). Meaning before order: Cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering. *Cognition, 180*, 59–81. doi: 10.1016/j.cognition.2018.06.012

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching, 46*(6), 699–715. doi: 10.1002/tea.20308

Steffe, L. P., & Cobb, P. (1988). *Construction of arithmetical meanings and strategies*. New York, NY: Springer-Verlag.

Stroup, W. W. (2012). *Generalized linear mixed models: Modern concepts, methods and applications*. Boca Raton, FL: CRC press.

Tzur, R. (2019). Hypothetical learning trajectory (HLT): A lens on conceptual transition between mathematical “markers”. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 56–74). Rotterdam, The Netherlands: Sense Publishers.

van Hiele, P. M. (1986). *Structure and insight: A theory of mathematics education*. Orlando, FL: Academic Press.

vanMarle, K., Chu, F. W., Mou, Y., Seok, J. H., Rouder, J., & Geary, D. C. (2018). Attaching meaning to the number words: Contributions of the object tracking and approximate number systems. *Developmental Science, 21*(1), e12495. doi: 10.1111/desc.12495

Vygotsky, L. S. (1935/1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *J Exp Child Psychol, 147*, 82–99. doi: 10.1016/j.jecp.2016.03.002

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*, 3571–3594.

Wilson, P. H., Mojica, G. F., & Confrey, J. (2013). Learning trajectories in teacher education: Supporting teachers' understandings of students' mathematical thinking. *The Journal of Mathematical Behavior, 32*(2), 103–121. doi: 10.1016/j.jmathb.2012.12.003

Wright, R. J., Stanger, G., Stafford, A. K., & Martland, J. (2006). *Teaching number in the classroom with 4-8 year olds*. London: Sage.

Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics*. Providence, RI: American Mathematical Society.