*AERA Open*
*January-December 2021, Vol. 7, No. 1, pp. 1–15*
*DOI: 10.1177/23328584211026657*
*Article reuse guidelines: sagepub.com/journals-permissions*


# Testing a Theoretical Assumption of a Learning-Trajectories Approach in Teaching Length Measurement to Kindergartners

**Julie Sarama**
**Douglas H. Clements**
*University of Denver*

**Arthur J. Baroody**
*University of Illinois at Urbana-Champaign*

**Traci S. Kutaka**
*University of Denver*
*University of Virginia*

**Pavel Chernyavskiy**
*University of Virginia*

**Jackie Shi**
**Menglong Cong**
*University of Denver*

*We tested a specific theoretical assumption of a learning trajectories (LTs) approach to curriculum and teaching in the domain of early length measurement. Participating kindergartners (*n *= 189) were assigned to one of three conditions: LT, reverse-order (REV), or business-as-usual (BAU). LT and REV students received one-on-one instruction using the same activities from a length LT, while the REV condition reversed the LT order. At posttest, LT and REV children exhibited significantly greater learning relative to BAU peers. But importantly, LT children outperformed their REV peers. We conclude that instruction following LTs (i.e., providing instruction just beyond a child's present level of thinking, progressing through the levels in order as the child advances) may promote more learning than an equivalent amount of instruction using the same activities but that are not theoretically sequenced.*

Keywords: *achievement, curriculum, early childhood, instructional design/development, learning trajectories, learning environments, mathematics education*

Learning trajectories (LTs) in early mathematics curriculum development and teaching have received increasing attention (Baroody et al., 2019; Clements, 2007; Clements & Sarama, 2021; Maloney et al., 2014; Sarama & Clements, 2009). For example, LTs were a core construct in the National Research Council (2009) report on early mathematics education (subtitled "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 (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). However, little research has directly tested the specific contributions of LTs to teaching compared with instruction provided without LTs (Frye et al., 2013). The goal of the present study was to compare the learning of kindergarteners who received instruction on length measurement following an empirically validated LT to those who received an equal amount of time on the same instructional activities that were not sequenced along the LT's developmental progression.

## *Theoretical Framework and Background*

*Learning Trajectories: Definition and Assumptions.* Our theoretical framework is hierarchic interactionalism (Sarama & Clements, 2009). This term reflects 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). LTs synthesizing these interactions stand at the core of this theory. Different fields name similar constructs differently, such as the use of "learning progressions" in science education.

The structure of LTs is built on the assumption that to be optimally useful to educators, LTs must include and integrate educational standards, empirical research on how children think and learn, as well as teaching strategies (Baroody et al., 2004; Carnine et al., 1997; National Research Council, 2007; Steedle & Shavelson, 2009). Therefore, we define a LT as having three components: a

goal, a developmental progression that describes levels of thinking, and instructional activities (including curricular tasks and pedagogical strategies) designed explicitly to promote the development of each level (Clements & Sarama, 2004; Maloney et al., 2014; National Research Council, 2009; Sarama & Clements, 2009). Definitions of LTs and learning progressions can differ, with some including only LTs' developmental progressions, and others, as also including a sequence of instructional activities. That is, the uniqueness of this view of the LTs construct stems from the inextricable interconnection between the components.

*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 as well as educators (Clements, Sarama, & DiBiase, 2004; Fuson, 2004; Sarama & Clements, 2009; Wu, 2011). Descriptions of the other two components of LTs require more details about hierarchic interactionalism (Sarama & Clements, 2009). Consistent with Vygotsky's (1935/1978) construction of the zone of proximal development (ZPD), hierarchic interactionalism posits that most content knowledge is acquired along developmental progressions, or levels of thinking within a specific topic that are 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, these levels are not stages but probabilistic patterns of thinking through which most children develop (e.g., an individual may learn multiple levels simultaneously; Sarama & Clements, 2009).

Hierarchical interactionalism 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, LTs include instructional tasks explicitly designed to include external objects and actions that mirror the hypothesized mathematical behavior of children as closely as possible. These tasks are sequenced, with each corresponding to a level of the developmental progression, 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 & Sarama, 2004, p. 83; Sarama & Clements, 2009, provides a complete description of hierarchic interactionalism's 12 tenets)

The goals and developmental progressions for many topics have been supported and validated by theoretical and empirical work describing consistent sequences of thinking levels. However, the amount of empirical support differs for different topics and ages (Confrey, 2019; Daro et al., 2011; Gravemeijer, 1994; Maloney et al., 2014; National Research Council, 2009), especially in domains such as the approximate number system and subitizing (e.g., Clements, Sarama, & MacDonald, 2019; vanMarle et al., 2018; J. J. Wang et al., 2016), counting (e.g., Fuson, 1988; Purpura et al., 2013; Spaepen et al., 2018), and arithmetic (e.g., Hickendorff et al., 2010). Furthermore, the application of developmental progressions as curricular guides (e.g., D. M. Clarke et al., 2001) and complete LTs (i.e., Clements et al., 2011; Clements & Sarama, 2008) have been successfully applied in early mathematics intervention projects, with significant effects on teachers' professional development (B. A. Clarke, 2008; Kutaka et al., 2016; Wilson et al., 2013) and children's achievement (D. M. Clarke et al., 2001; Clements & Sarama, 2008; Clements & Sarama, 2008; Kutaka et al., 2017; Murata, 2004; Wright et al., 2006).

*Learning Trajectories: Empirical Evidence.* Despite this research foundation, there are few studies that directly test 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 & Purpura, 2017; Frye et al., 2013), thus suggesting the use of LTs yields benefits, but without identifying their unique contribution (D. M. Clarke et al., 2001; Clements et al., 2011; Clements & Sarama, 2007; Fantuzzo et al., 2011; Gravemeijer, 1999; Jordan et al., 2012). For example, preschoolers who experienced a curriculum specifically designed on LTs demonstrated (a) significantly greater growth in mathematics competencies than those in a business-as-usual (BAU) control group score (effect size = 1.07) as well as (b) greater growth than those who experienced an intervention using a research-based curriculum that followed a sequence of mathematically rational topical units (effect size = 0.47; Clements & Sarama, 2008). 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 with the use of LTs and thus the specific effects of LTs could not be distinguished (Clements & Sarama, 2008).

### *Testing the Theoretical Assumptions of a Learning Trajectories Approach*

The present study is one of multiple experiments rigorously evaluating whether instruction based on LTs for early mathematics (Baroody et al., 2019; Clements & Sarama, 2021; Frye et al., 2013; National Research Council, 2009; Sarama & Clements, 2009) is significantly more efficacious than alternatives. To do so, we need to avoid confounding the essential elements of a learning trajectory with the myriad characteristics of curricula based on LTs. Therefore, we distilled two of the main characteristics of LTs that distinguish their application to curriculum and teaching from alternative pedagogical approaches and designed experimental conditions to rigorously test the efficacy of those two unstated assumptions.

The first assumption is that instruction should move children from their present level to the next higher level and continue in this manner until the instructional goal is reached. A competing approach posits 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; Wu, 2011), potentially obviating the need for potentially slower movement through each level. There is evidence supporting this approach to learning (Borman et al., 2003; Carnine et al., 1997; Clark et al., 2012; Gersten, 1985; Heasty et al., 2012), although the research designs often do not include other research-validated approaches. In contrast, LT approaches justify the assumption that each contiguous level be taught consecutively because LT's developmental progressions are more than linear sequences based on accretion of numerous facts and skills. Each is based on a progression of *levels of thinking* characterized by specific mental actions-on-objects that serve as a *foundation* for successful learning of all subsequent levels. We have tested this assumption in a series of studies that support the LT approach, with children exhibiting significantly greater learning than those taught at the target level for the same amount of time, including performing higher on target-level items, particularly those with low entry knowledge (Clements, Sarama, Baroody, & Joswick, 2020; Clements, Sarama, Baroody, Kutaka, et al., 2020; Clements, Sarama, Baroody, et al., 2019).

The second assumption of an LT approach is that there is a definite *sequence* of such levels of learning and teaching that is determined by research-based developmental progressions and that instruction is more efficacious if it builds each level in turn. Postulating that each level of thinking builds hierarchically on the concepts and processes of the previous levels stands in contrast to some traditional early childhood curricular organizations: theme, project, and emergent approaches (Broderick & Hong, 2020; Edwards et al., 1993; Helm & Katz, 2016; Hendrick, 1997; Katz & Chard, 2000; Tullis, 2011). In these approaches, it is the classroom theme (e.g., "colors"), or a project (visiting an apple orchard and making applesauce or building a bus when children expressed interest in buses spontaneously, reflecting the emergent curriculum approach) that determines the ordering of activities. For example, if the theme is colors, children are asked to sort by color; if it involves apples, children might count the seeds in an apple or cut them and talk about "halves." Thus, the activity is chosen for its fit to the classroom work, which is ostensibly more meaningful and connected for the child and thus will lead to greater learning. The critical question of "Which approach, LT, mathematical-relational, or traditional, results in better mathematical outcomes for preschool children?" has yet to be answered causally. We believe the benefits of LTs' sequencing of levels of thinking outweighs the benefits *this* type of "integration" of math with other themes (and the general philosophy that all early education events should emerge from children's choices). The present study is the first experiment to focus on the second assumption.

### *Present Study*

To evaluate the second, instructional sequence-of-levels assumption, with controlled conditions, we selected the length measurement learning trajectory. Early length measurement is important itself and serves as a critical bridge between geometry and number concepts (Clements & Sarama, 2021; Sarama & Clements, 2009). Research on the development of length-measurement knowledge suggests that young children enter school with a basic idea of how to use rulers and can verbally list or draw its attributes (e.g., lines and numbers, MacDonald & Lowrie, 2011). However, aligning an object at zero and reading the measure of a ruler does not mean that children understand how or why a ruler works. In fact, young children have trouble understanding the relationship between units and how iteration (accurate, repeated placement of a single unit) of discrete standard and nonstandard units produce measures (Lehrer, 2003; National Research Council, 2007). Thus, early measurement is important, but often taught badly, indicating a need for practice-based evidence (Bryk, 2015; Clements & Sarama, 2021).

*Instructional Conditions.* To test the significance of the instructional sequence-of-levels assumption in the context of an experimental study, the primary experimental group (LT) received one-on-one instruction that followed an empirically validated learning trajectory for length (Barrett et al., 2011, 2017; Sarama et al., 2011, 2021; Szilagyi et al., 2013). We compared growth in their length-measurement knowledge to two counterfactual groups. The secondary

experimental condition, reverse-order (REV) condition, provided one-on-one instruction that did not follow the order of the developmental progression. Instruction in this condition instead consisted of the same activities as the LT group, which covered all levels of the trajectory but sequenced in the reverse order of the learning trajectory. This order stood in for the aforementioned traditional approaches for three reasons. First, the traditional approaches would take place over a full year, well beyond the study's time frame, and still would not necessarily include all of the LT activities, which a rigorous test required. Second, the design also required a consistent, well-specified order (a random order would likely mix LT- and non-LT-sequences, as well as cover multiple topic strands). Third, studies suggest that attempting to solve challenging problems first is beneficial, as productive failure can help children make sense of the goals of subsequent instructional sessions (see discussions in Kapur, 2010; Loehr et al., 2014), providing a rationale for the efficacy of the REV condition and thus its usefulness as a counterfactual (albeit direct instruction following failure was not provided). The BAU condition did not receive any one-on-one instruction; of course, they continued with their classroom curriculum as did all children in all groups. This curriculum did not teach length during the semester of the study, so the BAU group served as a nontreatment control for statistical purposes.

*What Can We Learn From These Comparisons?* Comparing BAU growth to that of the LT and REV groups enables us to draw two categories of conclusions. First, we can confirm that "no harm" was done to student learning through pull-out participation in the LT and REV instructional sessions. Second, we can observe whether the LT length-measurement instructional activities benefit learning in the fall of kindergarten, which necessitated the no-treatment group comparison. This is a valid question, especially as even with instruction, children make slow progress even across grades (Barrett et al., 2017).

Comparing the two experimental conditions addresses the main research question, which is to clarify which characteristics of the LTs are "active ingredients" (vs. "inert," Bell et al., 2013) to more clearly describe their contributions to teaching and learning. To this end, we use a dismantling design. In such studies, a full "treatment package" is compared with a (dismantled) treatment condition with the hypothesized active component removed. If the full treatment (instruction for the LT condition) is found to be more effective than the dismantled treatment (instruction for the REV condition), the component that was removed (the order of the LT sequence) can be described as an active ingredient of treatment (Bell et al., 2013). Because the children in the REV group received the same activities (unlike our previous studies addressing the first assumption of LTs), we can attribute any differences in learning to the second assumption: instruction is more efficacious if it supports children to learn each LT level in sequence. Thus, we examine the following research question: Does instruction aligned with an LT's sequence result in greater learning than instruction that uses an LT's sequence in reverse order?

## Method

We used a randomized control trial to compare the three experimental groups: the intervention LT group, the REV group, and the BAU group.

### *Sample*

The intervention took place in a large, urban school district in a Mountain Range state. This school district is racially/ethnically diverse: 53.8% Latinx, 24.7% White, 13.2% African American, 3.2% Asian, 0.7% American Indian, 0.4% Native Hawaiian or Other Pacific Islander, and 4.1% respondents who identified as having two or more races. Additionally, 65% of students qualify for free-/reduced-lunch and 36.3% are English language learners. Preassessments were administered to 187 students; two students moved before the posttest, so 185 students were administered the postassessment.

*Recruitment Process.* Prior to our recruitment, the proposed study was submitted to the district external research review board. On approval, we reached out to multiple elementary schools that had more than one kindergarten classroom. Five school principals and their kindergarten teams agreed to host the study. Participating schools were offered professional development on early mathematics by the principal investigators of the project, as well as given the instructional materials at the end of the intervention. Table 1 contains basic demographic information of each site.

Prior to the study, graduate student instructors volunteered on at least two separate occasions in the participating teachers' classrooms for two reasons. First, we wanted to be seen as friendly adults by students. Second, it was the beginning of the academic year and teachers were still in the process of establishing classroom norms, expectations, and routines. We wanted to make sure that we set and communicated behavior expectations consistent with each classroom teacher.

*Randomization and Assignment to Experimental Condition.* The randomization process began in August 2019 and was completed in September 2019. Within each school, teachers and the research leadership team agreed to the following recruitment and randomization plan. Teachers and the instructional team members worked together to collect as many parental consent forms as possible. Once the teacher confirmed that we had as many signed permission forms we could reasonably expect, we created a list of these children.

TABLE 1
*Demographic Information for Participating Schools*

| Demographic | School 1 | School 2 | School 3 | School 4 | School 5 |
|---|---|---|---|---|---|
| No. of kindergarten classrooms | 3 | 4 | 4 | 1 | 4 |
| No. of classrooms with Spanish instruction | — | 1 | 1 | — | — |
| No. participants/total number of students | 26/405 | 22/226 | 63/583 | 11/177 | 65/457 |
| English language learner status | N/A | 45.1% | 49.9% | N/A | 8.1% |
| Individualized Education Program | N/A | 11.1% | 11.0% | N/A | 9.8% |
| School-level free/reduced-price lunch | N/A | 83.2% | 89.4% | N/A | 21.2% |
| Type | Private | Public | Public | Private | Public |

*Note.* School 2 classrooms are composed of kindergarten and first grade students. N/A = Data not publicly available.

We used a random number generator to assign each student a number from 0 to 1,000 to reorder the class list. Next, using the reordered class list, we used another random number generator, which assigned a 1, 2, or 3 to each student. Students who were assigned a 1 were put in the LT group, 2 in the REV group, and 3 in the BAU group. Frequency analyses were run to ensure that baseline levels of knowledge at pre-assessment (see details in "Assigning Pre-Mastery Levels") were distributed evenly among conditions for the whole sample. Table 2 lists the number of students per experimental condition by classroom and by school.

*English Language Learners.* As can be seen in Table 1, there were two classrooms (one in School 2 and another in School 3) where Spanish is the primary language of instruction. Assessment and instruction were administered by two bilingual instructors and one assessor proficient in Spanish. Additionally, there were three students at School 5 who were Mandarin speakers who were new to the country. Assessment and instruction were administered by one bilingual instructor. The assessment and instructional materials for the Mandarin speakers were translated in advance but did not go through the translation and back-translation process like the Spanish assessment materials.

### *Experimental Conditions*

Students in the LT and REV experimental conditions received ten 12-minute instructional sessions (120 minutes), while the BAU did not receive one-on-one instruction. The data collection period began the first week of September and ended the first week of December. Each LT and REV student had, at minimum, one instructional session per week. However, the pull schedule varied by classroom and reflected the preferences of the teacher. For example, one teacher requested that her students be pulled no more than one time per week, while another teacher in the same school wanted us to pull her students at least twice per week. As per our agreement with school leadership and teachers, we did not remove students from recess, lunch, math blocks, or literacy instruction.

Our rationale for implementing 10 instructional sessions for each group is based on findings from the small-scale pilot study conducted in a single-classroom prior to the larger scale study reported here. During the pilot, we trained instructors and assessors to fidelity in situ, evaluated our assessments for sensitivity, as well as examined the effectiveness of the instructional activities (e.g., added or deleted recommended scaffolding questions). Importantly, we also learned that 10 instructional sessions were necessary to develop a warm and productive relationship with students and to observe student growth, as indicated by a student's transition to $n + 1$ on the developmental progression. Furthermore, the ten 12-minute instructional sessions were negotiated to accommodate the request of participating classroom teachers and school leadership that the intervention minimize the amount of time students were removed from classroom instruction. The following describes each condition and instructor training.

*Learning Trajectories Approach.* There were 71 students in the LT condition (37 girls). This condition was composed of 10 one-on-one instructional sessions. Each instructor had access to a set of instructional activities that aligned with each level of the development progression and selected activities based on the child's present level of thinking and the competences needed to master a particular level. Ultimately, the developmental sequence of each activity prepares the student for the following level.

During the one-on-one sessions, the instructors selected activities based on the student's preassessment performance. The instructors documented and tracked the students' progression throughout the instructional sessions, which informed the selection of subsequent activities in later sessions. As children demonstrated higher levels of thinking, they were encouraged to use more sophisticated strategies, such as iterating with a single unit (instead of using multiple

TABLE 2
*Students per Experimental Condition by School and Classroom*

| | School 1 | | | School 2 | | | | School 3 | | | School 4 | | School 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 | 17 | 18 | Total |
| BAU | 2 | 3 | 3 | 2 | 1 | 2 | 2 | 8 | 2 | 5 | 6 | 4 | 4 | 4 | 5 | 4 | 57 |
| LT | 4 | 2 | 4 | 1 | 1 | 4 | 2 | 4 | 6 | 7 | 4 | 4 | 6 | 6 | 7 | 7 | 69 |
| REV | 2 | 3 | 3 | 1 | 2 | 0 | 3 | 4 | 5 | 4 | 8 | 3 | 8 | 4 | 5 | 4 | 59 |
| Total | 8 | 8 | 10 | 4 | 4 | 6 | 7 | 16 | 13 | 16 | 18 | 11 | 18 | 14 | 17 | 15 | 185 |

*Note*. A total of 187 students were administered pre-assessments and two students moved before the posttest. BAU = business-as-usual; LT = learning trajectory; REV = reverse-order.

units). The instructors provided scaffolds and differentiation throughout instruction based on what was most appropriate for each child, including (but not limited to) providing feedback on correctness of solution and instructor modeling of strategies. Unique to this condition, as necessary, instructors could modify an activity so that it required only the preceding level of thinking in the LT, then return to the original activity structure.

*Reverse Order Group.* There were 59 students in the REV condition (31 girls). This active control condition was composed of 10 one-on-one instructional sessions. Unlike the LT condition, 10 length activities were selected from each level of the developmental progression, which are listed on Table 3 and can be found on the $[LT]^2$ website (LearningTrajetories.org) for in-depth review. Instructors provided activities in a developmental sequence in the length LT in reverse order. Thus, students were exposed to similar activities as the LT condition but began with the most sophisticated level: Level 5 (conceptual ruler measurer).

REV instructors provided feedback about the correctness of children's solutions to questions but did not modify activities structurally to accommodate less sophisticated levels of thinking, as that would have broken the REV sequence. In the event that the child shared an incorrect solution, the instructor would gently let them know they did not get the right answer and would proceed to show them how to solve the problem (e.g., "Hmm, that is not quite right. Here is how I would solve the problem . . ."). The child would then solve the problem themselves before the instructor progressed to the next part of the activity. Children sometimes finished activities before the end of the session.

*Business-as-Usual.* There were 57 students in the BAU condition (31 girls). Students in this condition did not participate in any one-on-one instructional sessions, nor was the topic of length-measurement covered in the general curriculum. However, if one of the students expressed disappointment in not being "picked" to play math games, the instructor asked the teacher for permission to play a 10-minute subitizing or shape composition math game. Each of the instructors at the schools confirmed that kindergarten teachers were not exposed to LT curriculum prior to or during the study.

*Instructor Training.* The instructional team was composed of nine graduate research assistants (GRAs) from programs within the College of Education (others, including the senior authors, taught when needed). All of the GRAs had experience working with young children, and two were certified high school teachers for English literature. The rest of the instructional team members were students from the counseling or research methods and statistics programs. Each of the GRAs were trained by the co–principal investigators (co-PIs) and the project director. The GRAs were trained to implement instruction for the LT and REV conditions.

The training included description of the study design, the theoretical foundations of LTs for length-measurement, and how children advance along the developmental progression. After instructors became familiar with the length-measurement LT, the focus of training shifted to observing and interpreting children's thinking. Additionally, instructors were trained on how to provide appropriate instruction based on their interpretation of the child's thinking. As such, instructors were trained on selecting and implementing appropriate instructional tasks for each child (e.g., modifying activities between sessions to match instructional tasks to developmental levels of individual children) during weekly professional development sessions. In addition, the co-PIs and project directors observed the recorded instructional sessions weekly for each instructor and provided constructive feedback.

Each of the instructors participated in weekly team meetings where the co-PIs and Project Directors provided consultation on student cases. Furthermore, co-PIs and project directors were available, as a form of peer-debriefing, to answer questions and recommend next steps every day. Finally, each of the LT instructors had an in-person midpoint check-in (at or around the fifth instructional session) to determine if any midcourse corrections needed to occur. During this meeting, the graduate instructor consulted with the project director for each individual case.

TABLE 3
*Instructional Sequence for the Reverse Order Experimental Condition*

| Level in the developmental progression | Description | Activity name |
|---|---|---|
| Level 5: Conceptual Ruler Measurer | Possesses an "internal" measurement tool. Mentally moves along an object, segmenting it and counting the segments. Estimates with accuracy. | *Yard stick visualization*<br>"Guess and check" activities with initial estimates followed by measurements help children develop an "internal sense" of a length unit and iteration of that unit mentally. |
| Level 4: Length Measurer | Measures, knowing need for identical units, relationship between different units, partitions of unit, zero point on rulers, and accumulation of distance. | *Taxi Ride*<br>*Wobbly Roads*<br>Activities challenge children to combine the length measurements of a path with angles (bends or turns). Doing so using paths children can walk on, and later "walking with fingers" along paths drawn on paper may help children develop the conception of "connected lengths." |
| Level 3: Length Unit Relator and Repeater | Measures by repeating (iterating) a single unit and understands the need for equal-length unit. Relates the size and number of units (inverse relationship). | *Building Down*<br>*Cubes & Tiles*<br>*Which Bike Path is Shorter?*<br>Children measure by repeating (iterating) a single unit, with the task providing intrinsic feedback as to accuracy. |
| Level 2: End-to-End Length Measurer | Can lay units end to end to measure but may not recognize the need for equal-length units or be able to measure if there are fewer units than needed. | *Fire Trucks*<br>*Actual Size*<br>Children must measure by laying (a sufficient number of) units end to end without gaps or overlaps, with the task providing intrinsic feedback. |
| Level 1: Length Direct Comparer | Physically aligns two objects to determine which is longer or if they are the same length. | *As Long as My Arm*<br>*Measuring Our Shoes*<br>Activities focus on lining up objects at one of their endpoints and then comparing the other endpoints. |

*Note*. Children in the LT condition started one level above their initial developmental level as determined by the pretest and were then taught at successively higher levels. LT = learning trajectory.

Although we did not have the resources to measure fidelity of implementation in every session, we worked to ensure adequate implementation. Part of the LT instructor debrief included watching videos of instruction (see instructor training). This daily debrief served as an opportunity for the project lead to watch videos of sessions with instructors to provide feedback; for example, a reminder to use a scaffold listed in the activity sheet. Moreover, a fidelity evaluator reviewed the videos of two randomly selected students from each instructional team each week and documented whether the instructor Never (0%–25% of the time), Sometimes (26%–50%), Often (51%–75%), or Always (76%–100%): (a) taught in a teaching space that was conducive to student learning, (b) used the $[LT]^2$ write-up to clearly explain the activity directions, (c) positively engaged the student, (d) correctly set up the activity materials outlined in the $[LT]^2$ write-up, and (e) provided feedback about whether student responses were correct or incorrect (and for LT, moved in the developmental progression if indicated; for the REV, did not do so). Fidelity, defined as equal to or greater than 90%, was achieved for all instructors each week of the experiment.

### *Instrument*

The length measurement assessment was composed of 28 items adapted from the Research-Based Early Mathematics Assessment (Clements et al., 2008, 2021) and Cognitively Based Assessment (Battista, 2012) designed to assess length measurement learning for kindergarten students. Items assessed competences from four levels of the length LT, beginning with direct length comparing (e.g., compare the lengths of two objects, presented without alignment) up to length measurer (e.g., measure a 34.5-inch ribbon with a 10-inch ruler). Each item was scored for correctness and strategy sophistication (where 1 = *low*, 2 = *medium*, and 3 = *high*). Rasch scores were constructed (mean of 0, standard deviation of 1) and difficulty parameters confirmed that that beginning items are less difficult compared with the items near the end of the assessment.

*Preassessment.* Past work with kindergarteners (not in the study) indicated that they could not respond successfully to most of the higher level items. Therefore, to avoid undue frustration and possibly attrition, the preassessment adopted

TABLE 4
*Preinstruction Levels by Experimental Condition*

| | Level (Table 2) | | | |
|---|---|---|---|---|
| Condition | 0 | 1 | 2 | 3 |
| BAU | 10 (5.41%) | 39 (20.08%) | 10 (5.41%) | 0 |
| LT | 10 (5.41%) | 54 (29.19%) | 4 (2.16%) | 1 (0.54%) |
| REV | 8 (4.32%) | 49 (26.49%) | 0 | 0 |

*Note*. Level 0 is "length quantity recognizer" in which children recognize length as an attribute, possibly as an absolute descriptor rather than comparative, and distinguish it from other measurable attributes (area, volume). BAU = business-as-usual; LT = learning trajectory; REV = reverse-order.

a stop rule: administration ended when a student made three consecutive mistakes. Consequently, the pre-Rasch scores were constructed on a subset of the earliest items that compose the full-length measurement assessment. The sequencing of items according to Rasch difficulty guaranteed that after three incorrect responses, there was a very low probability that a child would answer any subsequent items correctly. Meanwhile, information—an analog of reliability—generated by the item response theory scores was 3.6 at the sample mean, equivalent to a reliability score of .78.

*Postassessment.* The postassessment was composed of the same items as the preassessment. All items were administered (because analyses were not on pre- to post change, and administrations were the same for all groups, such differences in administration did not affect the rigor of the assessments and data). Similar to the preassessment, Rasch scores were constructed and difficulty parameters suggest that beginning items are less difficult compared with the items near the end of the assessment. Additionally, information—an analog of reliability—generated by the item response theory scores was 8.8 at the sample mean, which is equivalent to a reliability score of 0.90.

*Assigning Starting Points for Instruction.* An initial preinstruction level of thinking in length measurement was assigned to each student. The preinstruction level was determined by correctly answering at least 75% of the items at *n* and all earlier levels. Table 4 contains the students' preinstruction level by condition.

*Covariates.* We tested two child-level covariates for the full sample. Child gender (0 = boy) and school type (public/private, where 0 = public) are coded as binary. During the intervention, no children transferred from one school type to another.

### *Analytic Approach*

The research question was examined within a Bayesian hierarchical linear modeling (HLM) framework using the brms package (Bürkner, 2018) in R 3.6.2 (R Core Team, 2019). Bayesian models more accurately quantify and propagate uncertainty (e.g., Kruschke, 2014) and can be more reliable in cases where traditional HLM methods typically fail (Eager & Roy, 2017).

The baseline model was specified to include preassessment ability, the effect of treatment, and a random intercept for classroom. The covariates we tested included child gender and whether the child attended a public or private school. We did not include a random intercept for the five participating schools since Snijders and Bosker (1993) advise against estimating multilevel models for clusters below 10.

The final models for each research question was selected using Watanabe–Akaike information criteria (WAIC; Watanabe, 2010). Each covariate was added sequentially and tested based on their contribution to model fit (as measured by the WAIC) and compared with the previous, less complex model. We favored parsimonious model solutions: smaller WAIC values to select for robustness and expected out-of-sample predictive performance. As we only have a single quantitative covariate (preassessment Rasch score) we tested a model that specified random slopes that allowed different effects of preassessment Rasch within each classroom. This model was not selected based on a comparison of information criteria.

## Results

### *Descriptive Statistics*

Table 5 contains the pre- and postassessment Rasch scores for each experimental condition. Baseline equivalence was examined and Table 6 contains the differences in preassessment performance, associated effect sizes, and notes whether these effect sizes meets the What Works Clearinghouse (2020) baseline equivalence standard of 0.05 or less in absolute value. The difference in the preassessment scores for the LT and BAU groups meet Institute of Education Sciences baseline equivalence standards. However, there are statistically nonsignificant initial differences between the LT and REV group, as well as the BAU and REV group (Table 5). In accordance with What Works Clearinghouse standards, we include child gender as a

TABLE 5
*Average Rasch Scores for Pretest and Posttest by Intervention Condition With Standard Deviations*

| Condition | | BAU ($n$ = 57) | LT ($n$ = 71) | REV ($n$ = 59) |
|---|---|---|---|---|
| Pretest | $n$ = 185 | 0.03 (0.76) | 0.04 (0.84) | −0.07 (0.62) |
| Posttest | $n$ = 185 | −0.33 (0.72)[a] | 0.31 (1.24)[a] | −0.15 (0.75) |

*Note*. BAU = business-as-usual; LT = learning trajectory; REV = reverse-order.
[a]Paired $t$ test for the means was statistically significant at = .05.

TABLE 6
*Differences in Average Preassessment Scores and Associated Effect Size by Intervention Condition*

| Condition | Theta mean difference (*SE*) | Effect size | WWC standard met? |
|---|---|---|---|
| LT-BAU | 0.01 (0.00) | 0.01 | Yes |
| LT-REV | 0.11 (0.01) | 0.13 | No |
| BAU-REV | 0.11 (0.00) | 0.12 | No |

*Note*. WWC = What Works Clearinghouse; LT = learning trajectory; BAU = business-as-usual; REV = reverse-order.

TABLE 7
*Baseline Effect Size by Private or Public School*

| School type | Theta *SD* | Theta mean | Theta mean difference | Effect size |
|---|---|---|---|---|
| Public | 0.70 | −0.12 | 0.62 | 0.87 |
| Private | 0.73 | 0.50 | | |

covariate, given the slight advantage of girls are pre-assessment (effect size = 0.13). Additionally, there were initial differences between student performance in public ($n$ = 52 students across three schools) versus private schools ($n$ = 135 students across two schools; see Table 7; Hedges $g$ effect size = 0.87).

### *Impact of the Learning Trajectories Approach*

The final model was the parsimonious baseline model (see Table 8). For the $i$th student in the $j$th classroom, the postassessment Rasch score ($\theta\_POST_{ij}$) is modeled as follows:

$$\theta\_POST_{ij} = (\beta_0 + u_j) + \theta\_PRE_{ij} + LT_{ij} + REV_{ij} + GENDER_{ij} + PRIVATE_{ij} + \epsilon_{ij}$$

$$u_j \sim Normal(0, \sigma_u)$$

$$\epsilon_{ij} \sim Normal(0, \sigma_\epsilon)$$

In the equation above, $\theta\_PRE_{ij}$ is the preassessment Rasch score, $LT_{ij}$ and $REV_{ij}$ are dummy variables indicating assignment into LT and REV conditions, $GENDER_{ij}$ is a dummy variable for child sex, $PRIVATE_{ij}$ is a dummy variable for private versus public school, and $u_j$ represent classroom random intercepts. We considered several interactions, but the addition of interactions did not improve the WAIC (see Table 8). This indicates that the effect of the treatment was moderated neither by initial length-measurement competences nor by what kind of school students attended.

HLM parameters are presented in Table 9 and differences between experimental conditions are show in Figure 1. Students in the LT and REV condition outperformed their peers in the BAU condition. Contrasts further reveal that students in the LT group also outperform their REV peers = .32 [0.57, 0.07]. Credible Intervals of 95% were estimated for child gender and whether the child attended a public or private school. However, these intervals include zero and therefore deemed to be statistically nonsignificant.

## Discussion

As one of a set of experiments rigorously testing the efficacy of the educational application of LTs, this study focused on the second assumption of an LT approach: there is a *sequence* of learning and teaching that is determined by a research-based developmental progression. The topic of length was selected as being both important to early mathematics learning and amenable to the LT and counterfactual conditions. The first of these reversed the

TABLE 8
*Fit Indices for Model Selection Based on WAIC and Bayesian $R^2$ (95% Credible Intervals)*

| | WAIC | Effective parameters | Bayesian $R^2$ |
|---|---|---|---|
| Baseline model | 429.6 | 25.6 | 0.51 [0.42, 0.58] |
| Baseline model + condition × prelength | 433.9 | 28.3 | 0.51 [0.43, 0.58] |
| Baseline + condition × private | 435.1 | 28.0 | 0.51 [0.43, 0.58] |
| Baseline + condition × gender | 433.4 | 27.0 | 0.51 [0.43, 0.57] |

*Note*. Models with smaller information criteria indicate lower expected out-of-sample predictive performance; models with larger $R^2$ indicate better within-sample performance. WAIC = Watanabe–Akaike information criteria.

TABLE 9
*Model Parameter Estimates (Posterior Medians) with 95% Credible Intervals for Postlength With Random Effect of Classroom*

| | Estimate | *SE* | 95% CI (lower) | 95% CI (upper) |
|---|---|---|---|---|
| Intercept | −0.49 | 0.22 | −0.92 | −0.07 |
| Prelength | 0.33 | 0.06 | 0.20 | 0.45 |
| Condition–LT (BAU is reference group) | 0.58 | 0.13 | 0.33 | 0.84 |
| Condition–REV (BAU is reference group) | 0.27 | 0.14 | −0.01 | 0.53 |
| Child gender (girls are reference group) | –0.03 | 0.10 | −0.23 | 0.18 |
| Public/private (public is reference group) | 0.43 | 0.40 | −0.36 | 1.22 |
| Intraclass correlation (classroom) | 0.40 | 0.12 | 0.23 | 0.62 |
| Residual error | 0.71 | 0.04 | 0.64 | 0.79 |
| Bayesian $R^2$ | 0.51 | 0.04 | 0.42 | 0.58 |

*Note.* Bayesian $R^2$ is computed using methods specified in Gelman et al. (2019). LT = learning trajectory; BAU = business-as-usual; REV = reverse-order.

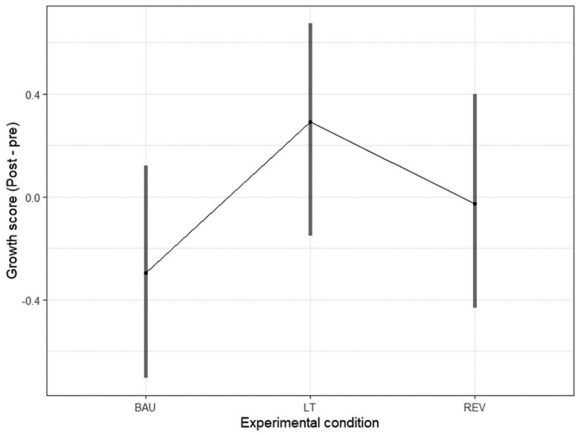


FIGURE 1. *Estimated growth scores (posterior medians with 95% credible intervals) for three experimental conditions.*

sequence of the LT activities to directly test the assumption (REV). The second counterfactual, BAU, served as a passive control.

Students in the LT group outperformed their BAU and their REV peers. The latter contrast, especially, supports the hypothesis that following the developmental progression of an empirically validated LT promotes learning more than the same activities not in that order.

We acknowledge there is a possible alternative explanation for the difference in LT versus REV postassessment performance: the instructor response to student error differed between the LT and REV conditions. Instructors in the LT and REV condition provided feedback about the correctness of solution and modeled how to solve the problem. However, scaffolds provided in response to incorrect answers differed between conditions. Take, for example, the game "Which Bike Path Is Shorter?" In this game, students measure and compare two lengths that cannot be physically compared (length direct comparer level in Table 3) by iterating one or two units—a common cognitive knot.

If a REV student did not endorse the correct answer, the instructor would say, "You're working very hard and I like how you're thinking. But, that's not quite right. Here is how I would do it." The instructor would then provide a demonstration of how to leapfrog two units for the correct answer, narrating their actions in child-friendly language. The child would then be given a turn. If an LT student made an error, the instructor would ask probing questions for diagnostic purposes. A response that was productive for most students was to have the student share their unitized answer and then check their work against the same length, but this time using as many units as they need to cover the distance. The rationale behind the differences in response is that providing an LT-aligned response to REV student errors would have been

incompatible with the nature of the REV condition. To properly address REV student thinking, instructors would have to build skills promoted at a different level of the developmental progression than the intended daily activity.

Students in the REV condition also outperformed their peers in the BAU condition. This indicates that the activities, even when implemented in an order other that of the LT's developmental progression, are still effective. This result is similar to that of the studies testing the first LT assumption, those that had a "teach-to-target" counterfactual (Clements, Sarama, Baroody, & Joswick, 2020; Clements, Sarama, Baroody, Kutaka, et al., 2020; Clements, Sarama, Baroody, et al., 2019). That is, these and the present study suggest teaching each contiguous level in developmental order of a LT is more efficacious and thus more useful than alternatives, but not necessary to facilitate learning in all cases—children experiencing the active counterfactuals also learned, but they learned less. However, note that in previous studies using a teach-to-target approach to test assumption 1 (Clements, Sarama, Baroody, & Joswick, 2020; Clements, Sarama, Baroody, Kutaka, et al., 2020; Clements, Sarama, Baroody, et al., 2019), instruction was at levels $n + 2$ or $n + 3$, avoiding instruction at $n + 1$. In the present study, children in the REV condition experienced activities at each level considered in this study (Table 3) and thus the activities eventually crossed over the child's present level of thinking (including $n - 1$, $n$, $n + 1$, etc.).

We also note an alternative explanation for the findings regarding comparisons to the BAU group: LT and REV students received one-on-one instructional sessions, whereas the BAU students did not. It may be that any one-on-one attention yields greater length-measurement performance at postassessment. Indeed, time spent in domain-specific, rather than general, instruction is associated with higher scores in targeted domains, including mathematics, in children from low-income preschool and kindergarten backgrounds (Votruba-Drzal & Miller, 2016; A. H. Wang, 2010).

Another caveat is that comparisons to the BAU condition necessarily confounded the additional one-on-one sessions with the instructional activities. These comparisons were relevant to an evaluation of the activities' efficacy, but did not address our main question. Finally, the comparison between the LT and REV groups, which differed mainly on the sequence they embodied, is confounded by necessity with the added scaffolding provided children in the LT condition that modified activities at level $n$ to an $n - 1$ structure temporarily.

### Implications for Theory, Research, and Practice

Instruction using activities sequenced according to the levels of an empirically validated learning trajectory was more efficacious than instruction using the same activities for the same amount of time, but not so ordered. This supports the LT assumption that each builds hierarchically on the concepts and processes of the previous levels (e.g., Goodson, 1982; Sarama & Clements, 2009; van Hiele, 1986). That is, each level is characterized by specific concepts (e.g., mental objects) and processes (mental "actions-on-objects"; Clements, Wilson, & Sarama, 2004; Steffe & Cobb, 1988) that underlie mathematical thinking at level $n$ and serve as a foundation to support successful learning of subsequent levels (Sarama & Clements, 2009). However, the learning process is not intermittent and step-like, but rather incremental and gradually integrative. A critical mass of ideas from each level must be constructed before thinking characteristic of the subsequent level becomes ascendant in the child's thinking and behavior (Clements et al., 2001).

Findings are consistent with Vygotsky's construction of the ZPD (Vygotsky, 1935/1978) but add required theoretical, empirical, and practical knowledge. For example, application of ZPD must confront the question of whether any particular competence or activity stands within a child's ZPD (vs. already interiorized or beyond the zone) and must clarify the role and nature of adult guidance for that activity (Wertsch, 1984). The three components of a learning trajectory, instantiated in the theory of hierarchic interactionalism (Sarama & Clements, 2009), provide a research-based structure for children's mathematical knowledge as well as pedagogical tools that enable us to work in congruity with the ZPD theory. That is, goals elucidate the mathematical content, and developmental progressions specify and arrange increasingly sophisticated levels of thinking. These enable us to identify the child's current level of thinking and the following level—exactly the ZPD (the "upper threshold of instruction, Wertsch, 1984). The instruction for that level provides the teacher, Vygotsky's More Knowledgeable Other, with specific teaching activities, as well as a theoretical rationale for why the activity will activate the mental actions-on-objects constituting thinking at that level. LTs also posit the mechanisms that tie a developmental progression levels of thinking to the instructional tasks through the specification of "actions-on-objects."

Also consistent with the ZPD construct (Vygotsky, 1935/1978), the LT approach involves using *formative assessment* (National Mathematics Advisory Panel, 2008; Shepard & Pellegrino, 2018) to provide instructional activities aligned with such empirically validated developmental progressions (D. M. Clarke et al., 2001; Fantuzzo et al., 2011; Gravemeijer, 1999; Jordan et al., 2012) and using teaching strategies that evoke children's natural patterns of thinking at each level, as posited by hierarchical interactionalism (Sarama & Clements, 2009). This approach appears particularly productive for those with the lowest levels of entry competencies. This similarly indicates the importance of supporting children's learning of each level of the LT in order, 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. 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).

Consistent with previous research (Clements, Sarama, Baroody, & Joswick, 2020; Clements, Sarama, Baroody, Kutaka, et al., 2020; Clements, Sarama, Baroody, et al., 2019), teaching each contiguous level in developmental order of a LT is more efficacious and thus useful than alternatives, but not necessary to facilitate learning in all cases—children experiencing the active counterfactuals also learned, just not as much.[1] Thus, selecting activities with traditional approaches (reminiscent of the sequence of instructional activities that defined the REV condition or a random order) has the potential to teach young children mathematics, but not as effectively as selecting activities based on children's present level of thinking.

A caveat regarding implications for practice is that potential pedagogical power of the traditional approaches, that is integration with other activities and domains, could not be realized in the REV condition due to logistical and research design constraints. Furthermore, the REV condition outperformed the BAU condition, showing that the activities were effective even when not aligned with the developmental progression (the relative efficacy of the LT condition may have been attenuated by the recency effect). Therefore, the results support the efficacy of following the developmental progression, but should not be interpreted, for example, as an evaluation of the traditional approaches, which would require full-year research with different curriculum structures. However, the study does imply that theme approaches that considers LTs in their planning would be more effective than those that do not.

## Acknowledgments

Researchers from an independent institution oversaw the research design, data collection, and analysis and confirmed procedures and findings. The authors wish to express appreciation to the school districts, teachers, and children who participated in this research.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grant No. R305A150243. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

## ORCID iD

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

## Note

1. When a level is prerequisite for learning a higher level, and instruction does not include the content of the prerequisite level, the developmental order may, of course, be a necessary condition, such as we found in a related experiment with cardinality Baroody et al. (under review).

## References

Baroody, A. J., Cibulskis, M., Lai, M.-l., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning*, *6*, 227–260. https://doi.org/10.1207/s15327833mtl0602_8

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). Wiley Blackwell. https://doi.org/10.1002/9781119148104

Baroody, A. J., Eiland, M. D., Clements, D. H., & Sarama, J. (2021). *Does a learning trajectory facilitate the learning of early cardinal-number concepts?* Manuscript submitted for publication.

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

Barrett, J. E., Clements, D. H., & Sarama, J. (2017). Children's measurement: A longitudinal study of children's knowledge and learning of length, area, and volume. *Journal for Research in Mathematics Education Monograph Series* (Vol. 16).

Barrett, J. E., Cullen, C. J., Sarama, J., Clements, D. H., Klanderman, D., Miller, A. L., & Rumsey, C. (2011). Children's unit concepts in measurement: A teaching experiment spanning grades 2 through 5. *ZDM–The International Journal on Mathematics Education*, *43*(5), 637–650. https://doi.org/10.1080/10986065.2012.625075

Battista, M. T. (2012). *Cognition based assessment and teaching of geometric measurement (length, area, and volume): Building on students' reasoning*. Heinemann.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology*, *81*(4), 722–736. https://doi.org/10.1037/a0033004

Bereiter, C. (1986). Does direct instruction cause delinquency? Response to Schweinhart and Weikart. *Educational Leadership*, *44*(3), 20–21. https://doi.org/10.1016/0885-2006(86)90036-0

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*(2), 125–230. https://doi.org/10.3102/00346543073002125

Broderick, J. T., & Hong, S. B. (2020). *From children's interests to children's thinking: Using a cycle of inquiry to plan curriculum*. National Association for the Education of Young Children.

Bryk, A. S. (2015). Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189x15621543

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/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. https://doi.org/10.1177/074193259701800201

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). Sense.

Clarke, D. M., Cheeseman, J., Clarke, B., Gervasoni, A., Gronn, D., Horne, M., McDonough, A., Montgomery, P., Rowley, G., & 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). Mathematics Education Research Group of Australasia.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*(1), 35–70. https://doi.org/10.2307/30034927

Clements, D. H., Battista, M. T., & Sarama, J. (2001). *Logo and geometry* (Vol. 10). National Council of Teachers of Mathematics. https://doi.org/10.2307/749924

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*(2), 81–89. https://doi.org/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. https://doi.org/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. https://doi.org/10.3102/0002831207312908

Clements, D. H., & Sarama, J. (2021). *Learning and teaching early math: The learning trajectories approach* (3rd ed.). Routledge. https://www.routledge.com/Learning-and-Teaching-Early-Math-The-Learning-Trajectories-Approach/Clements-Sarama/p/book/9780367521974

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*, *52*(4), 637–648. https://doi.org/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. https://doi.org/10.3102/0002831219842788

Clements, D. H., Sarama, J., Baroody, A. J., Kutaka, T. S., Chernyavskiy, P., Joswick, C., Cong, M., & Joseph, E. (2020). Comparing the efficacy of early arithmetic instruction based on a learning trajectory and teaching-to-a-target. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000633

Clements, D. H., Sarama, J., & DiBiase, A.-M. (Eds.). (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Lawrence 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). Springer. https://doi.org/10.1007/978-3-030-00491-0

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. https://doi.org/10.5951/jresematheduc.42.2.0127

Clements, D. H., Sarama, J., Wolfe, C. B., & Day-Hess, C. A. (2008/2021). *REMA—Research-based Early Mathematics Assessment*. Kennedy Institute, University of Denver.

Clements, D. H., Wilson, D. C., & Sarama, J. (2004). Young children's composition of geometric figures: A learning trajectory. *Mathematical Thinking and Learning*, *6*(2), 163–184. https://doi.org/10.1207/s15327833mtl0602_1

Confrey, J. (2019). *A synthesis of research on learning trajectories/progressions in mathematics*. Organisation for Economic Co-operation and Development. 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., Confrey, J., Daro, V., Maloney, A., Nagakura, W., Petit, M. M., & Sarama, J. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. Consortium for Policy Research in Education.

Edwards, C., Gandini, L., & Forman, G. E. (1993). *The hundred languages of children: The Reggio Emilia approach to early childhood education*. Ablex.

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. https://doi.org/10.3102/0002831210385446

Frye, D., Baroody, A. J., Burchinal, M. R., Carver, S., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide*. National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/PracticeGuide/18

Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-3754-9

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). Lawrence Erlbaum.

Gersten, R. (1985). Direct instruction with special education students: A review of evaluation research. *Journal of Special Education*, *19*(1), 41–58. https://doi.org/10.1177/002246698501900104

Goodson, B. D. (1982). The development of hierarchic organization: The reproduction, planning, and perception of multiarch block structures. In G. E. Forman (Ed.), *Action and thought* (pp. 165–201). Academic Press.

Gravemeijer, K. P. E. (1994). Educational development and developmental research in mathematics education. *Journal for*

*Research in Mathematics Education*, *25*(5), 443–471. https://doi.org/10.2307/749485

Gravemeijer, K. P. E. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, *1*(2), 155–177. https://doi.org/10.1207/s15327833mtl0102_4

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. www.savap.org.pk/journals/ARInt./Vol.2(3)/2012(2.3-47).pdf

Helm, J. H., & Katz, L. G. (2016). *Young investigators: The project approach in the early years* (3rd ed.). Teachers College Press.

Hendrick, J. (Ed.). (1997). *First steps toward teaching the Reggio way*. Prentice Hall.

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. https://doi.org/10.1037/a0018177

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. https://doi.org/10.1037/a0029018

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, *38*, 523–550. https://doi.org/10.1007/s11251-009-9093-x

Katz, L. G., & Chard, S. C. (2000). *Engaging children's minds: The project approach* (2nd ed.). Ablex.

Kutaka, T. S., Ren, L., Smith, W. M., Beattie, H. L., Edwards, C. P., Green, J. L., Chernyavskiy, P., Stroup, W. W., Heaton, R. M., & Lewis, W. J. (2016). Examining change in K-3 teachers' mathematical knowledge, attitudes, and beliefs: The case of primarily math. *Journal of Mathematical Behavior*, *21*, 147–177.https://doi.org/0.1007/s10857-016-9355-x

Kutaka, T. S., Smith, W. M., Albano, A. D., Edwards, C. P., Ren, L., Beattie, H. L., Lewis, W. J., Heaton, R. M., & 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. https://doi.org/10.1177/0022487116687551

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 179–192). National Council of Teachers of Mathematics.

Loehr, A. M., Fyfe, E. R., & Rittle-Johnson, B. (2014). Wait for it . . . delaying instruction improves mathematics problem solving: Classroom study. *Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1166

MacDonald, A., & Lowrie, T. (2011). Developing measurement concepts within context: Children's representations of length. *Mathematics Education Research Journal*, *23*(1), 27–42. https://doi.org/10.1007/s13394-011-0002-7

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

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. https://doi.org/10.1207/s1532690xci2202_2

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards*. http://corestandards.org/

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education, Office of Planning, Evaluation and Policy Development. https://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

National Research Council. (2007). *Taking science to school: Learning and teaching sciences in grades K-8*. National Academies Press. https://www.nap.edu/catalog/11625/taking-science-to-school-learning-and-teaching-science-in-grades

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. National Academies Press. https://doi.org/10.17226/12519

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. https://doi.org/10.1037/a0031753

R Core Team. (2019). *R: A language and environment for statistical computing*. https://www.R-project.org/

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge. https://doi.org/10.4324/9780203883785

Sarama, J., Clements, D. H., Barrett, J. E., Cullen, C. J., & Hudyma, A. (2021). Length measurement in the early years: Teaching and learning with learning trajectories. *Mathematical Thinking and Learning*. Advance online publication. https://doi.org/10.1080/10986065.2020.1858245

Sarama, J., Clements, D. H., Barrett, J. E., Van Dine, D. W., & McDonel, J. S. (2011). Evaluation of a learning trajectory for length in the early years. *ZDM-The International Journal on Mathematics Education*, *43*(5), 667–680. https://doi.org/10.1007/s11858-011-0326-5

Shepard, L. A., & Pellegrino, J. W. (2018). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement*, *37*(1), 52–57. https://doi.org/10.1111/emip.12195

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. https://doi.org/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. https://doi.org/10.1002/tea.20308

Steffe, L. P., & Cobb, P. (1988). *Construction of arithmetical meanings and strategies*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-3844-7

Szilagyi, J., Sarama, J., & Clements, D. H. (2013). Young children's understandings of length measurement: Evaluating a learning trajectory. *Journal for Research in Mathematics Education*, *44*(3), 581–620. https://doi.org/10.5951/jresematheduc.44.3.0581

Tullis, P. (2011). The death of preschool. *Scientific American Mind*, *22*(5), 36–41. https://doi.org/10.1038/scientificamericanmind1111-36

van Hiele, P. M. (1986). *Structure and insight: A theory of mathematics education*. 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. https://doi.org/10.1111/desc.12495

Votruba-Drzal, E., & Miller, P. (2016). Reflections on quality and dosage of preschool and children's development. *Monographs of the Society for Research in Child Development*, *81*(2), 100–113. https://doi.org/10.1111/mono.12244

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. (Original work published 1935)

Wang, A. H. (2010). Optimizing early mathematics experiences for children from low-income families: A study on opportunity to learn mathematics. *Early Childhood Education Journal*, *37*(4), 295–302. https://doi.org/10.1007/s10643-009-0353-9

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. https://doi.org/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.

Wertsch, J. V. (1984). The zone of proximal development: Some conceptual issues. In B. Rogoff & J. V. Wertsch (Eds.), *Children's learning in the "Zone of Porximal Development"* (Vol. 23, pp. 7–18). Jossey Bass.

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. https://doi.org/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*. Sage.

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

WWC. (2000). *Baseline equivalence*. What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf

## Authors

JULIE SARAMA, Kennedy Endowed Chair and Distinguished University Professor, has taught high school mathematics, computer science, middle school gifted mathematics and early childhood mathematics. She directs six projects funded by the National Science Foundation, Institute of Education Sciences, and others (30 total) and has authored over 80 refereed articles, seven books, 60 chapters, and over 100 additional publications.

DOUGLAS H. CLEMENTS is Distinguished University Professor and Kennedy Endowed Chair at the University of Denver, Colorado. Clements has published over 166 refereed research studies, 27 books, 100 chapters, and 300 additional works on the learning and teaching of early mathematics; computer applications, research-based curricula, and taking interventions to scale.

ARTHUR J. BARODDY is a professor emeritus of curriculum and instruction, College of Education, University of Illinois at Urbana-Champaign. His research focuses on the teaching and learning of early number, counting, and arithmetic concepts and skills.

TRACI S. KUTAKA is a research associate at the Marsico Institute for Early Learning at the University of Denver. Her research interests center on early childhood care and education, with an emphasis on mathematics teaching and learning.

PAVEL CHERNYAVSKIY is an assistant professor of biostatistics at the University of Virginia in the Department of Public Health Sciences. His research interests include methods for the analysis of correlated data, as well as Bayesian computational methods.

JACKIE SHI is a PhD candidate in research methods and statistics at University of Denver. Her research interests include applying advanced models in institutional research and public opinions.

MENGLONG CONG is a PhD candidate of the Department of Research Methods and Statistics, Morgridge College of Education, University of Denver. His research interests include child development and mixed-method research.