*American Educational Research Journal*
*December 2019, Vol. 56, No. 6, pp. 2509–2530*
*DOI: 10.3102/0002831219842788*
*Article reuse guidelines: sagepub.com/journals-permissions*


# Evaluating the Efficacy of a Learning Trajectory for Early Shape Composition

**Douglas H. Clements**
**Julie Sarama**
*University of Denver*
**Arthur J. Baroody**
*University of Illinois at Urbana-Champaign*
**Candace Joswick**
*University of Denver*
**Christopher B. Wolfe**
*Saint Leo University*

*Although basing instruction on learning trajectories (LTs) is often recommended, there is little direct evidence regarding the premise of a LT*

Douglas H. Clements is the Kennedy Endowed Chair in Early Childhood Learning and Distinguished University Professor at the University of Denver Morgridge College of Education, Katherine A. Ruffatto Hall 154, 1999 East Evans Avenue, University of Denver, Denver, CO 80208-1700; e-mail: *douglas.clements@du.edu*. His research interests include the learning and teaching of early mathematics; computer applications; creating, using, and evaluating research-based curricula; and taking interventions to scale.

Julie Sarama is the Kennedy Endowed Chair in Innovative Learning Technologies and Distinguished University Professor at the University of Denver Morgridge College of Education. Her research interests include young children's development of mathematical concepts and competencies, implementation and scale-up of educational reform, professional development models and their influence on student learning, and implementation and effects of software.

Arthur J. Baroody is Professor Emeritus of Curriculum & Instruction at College of Education, University of Illinois at Urbana-Champaign. His research focuses on the teaching and learning of number, counting, and arithmetic concepts and skills in early childhood.

Candace Joswick is a postdoctoral research fellow at the Marsico Institute of Early Learning and James C. Kennedy Institute for Educational Success at the University of Denver's Morgridge College of Education. Her research interests include learning trajectories and progressions for teaching and learning mathematics from birth through Grade 12, technology for teaching and learning mathematics, and mathematical classroom discourse.

Christopher B. Wolfe is an assistant professor of education psychology at Saint Leo University. His research interests include the development of reading and mathematics, curricula effects, and multilevel modeling.

*approach—that instruction should be presented (only) one LT level beyond a child's present level. We evaluated this hypothesis in the domain of early shape composition. One group of preschoolers, who were at least two levels below the target instructional LT level, received instruction based on an empirically validated LT. The counterfactual (skip-levels) group received an equal amount of instruction focused only on the target level. At posttest, children in the LT condition exhibited significantly greater learning than children in the skip-levels condition, mainly on near-transfer items; no child-level variables were significant moderators. Implications for theory and practice are discussed.*

KEYWORDS: achievement, curriculum, early childhood, instructional design/development, learning trajectories, learning environments, mathematics education, two-dimensional shape composition

The use of learning trajectories (LTs) in early mathematics instruction has received increasing attention from policy makers, educators, curriculum developers, and researchers (Baroody, Clements, & Sarama, 2019; Clements & Sarama, 2014a; Maloney, Confrey, & Nguyen, 2014; Sarama & Clements, 2009) and are generally deemed as a useful tool for guiding standards, instructional planning, and assessment (Frye, Baroody, Burchinal, Carver, Jordan, & McDowell, 2013; National Research Council, 2009). Despite these recommendations, little research has directly tested the specific contributions of LTs to children's learning (Frye et al., 2013). The primary goal in the present study was to compare the learning of preschool children who received instruction on shape composition based on an empirically validated LT to those who received an equal amount of instruction that focused only on the target goal.

## Background and Theoretical Framework

### The Nature of Learning Trajectories and an LT for Two-Dimensional Shape Composition

Building on Simon's (1995) original formulation, we conceptualize LTs 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 align with each level (Clements & Sarama, 2004; Maloney et al., 2014; National Research Council, 2009). In the remainder of this section, we discuss the components, illustrating each with the LT for composition of two-dimensional geometric figures.

#### *Goals*

The learning goals of LTs are based on standards grounded in research (National Governors Association Center for Best Practices, Council of Chief

State School Officers [NGA/CCSSO], 2010). Such goals, then, consider the expertise of mathematicians, social needs, and research on children's thinking about and learning of mathematics (Clements, Sarama, & DiBiase, 2004; Fuson, 2004; National Research Council, 2009).

For example, shape composition, the ability to describe, use, and visualize the effects of composing geometric regions, is an important construct because the concepts and actions of creating and then iterating units in the context of constructing patterns, measuring, and computing are established bases for mathematical understanding and analysis (Clements, Battista, Sarama, & Swaminathan, 1997; NGA/CCSSO, 2010; Reynolds & Wheatley, 1996; Steffe & Cobb, 1988). Additionally, there is suggestive evidence that this type of composition corresponds with, and may support, children's ability to compose and decompose numbers (Clements, Sarama, Battista, & Swaminathan, 1996). Thus, the goal of our LT for shape composition is children can accurately and with anticipation compose two-dimensional shapes to create composite shapes (i.e., planning to create a superordinate figure by combining two or more shapes).

### *Developmental Progressions of Levels of Thinking*

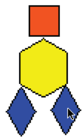
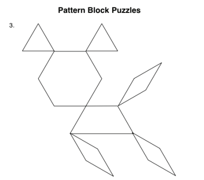
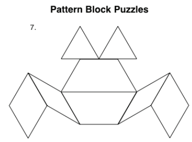
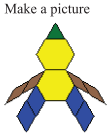
LTs' developmental progressions are more than linear sequences based on accretion of numerous facts and skills or a "progression" of assessment tasks. For example, some have confused LTs with sequences based solely on the structure of mathematics content or with "stages," such as Piaget's (see also Lesh & Yoon, 2004; Resnick & Ford, 1981). Similar to learning progressions (Alonzo & Gotwals, 2012) or developmental sequences (Mueller, Sokol, & Overton, 1999), LTs include sequences of levels of thinking, each more sophisticated than the last, through which children develop on their way to achieving the mathematical goal. Each level is characterized by specific concepts (e.g., mental objects) and processes (mental "actions-on-objects"; Clements, Wilson, & Sarama, 2004) that underlie mathematical thinking at level $n$ and serve as a foundation to support successful learning of subsequent levels (Steffe & Cobb, 1988). Specification of these actions-on-objects allows a degree of precision not achieved by previous theoretical or empirical works. Furthermore, LTs address both thinking and learning—that is, transitions between levels are central to effective teaching and learning (Steffe, Thompson, & Glasersfeld, 2000). In this approach, effective instruction involves more than teaching a specific lesson or concept (such as "today we are focusing on counting objects") because such an approach does not account for levels of development, individual differences in children's abilities, or the connectedness of mathematical knowledge. Instead, instruction must also focus on the growth children experience in their progress toward the goal.

The developmental progression for shape composition was developed and validated over multiple studies (Clements, Wilson et al., 2004; Clements & Sarama, 2007/2013). Born in observations of kindergartners composing physical and computer shapes (Sarama, Clements, & Vukelic, 1996), we combined these observations with related observations from other researchers (Mansfield & Scott, 1990; Sales, 1994) and some elements of psychological research (e.g., Vurpillot, 1976) to create the initial developmental progression. We then engaged in cycles of observations and analysis to refine the developmental progression (and begin to develop instructional activities; Clements & Sarama, 2007/2013) including collaborative action research with eight teachers. This version of the developmental progression was subjected to a wide variety of empirical tests including qualitative and quantitative techniques, from clinical interviews with 72 children aged 3 to 7 years to Rasch analyses, using confidence intervals to detect segmentation and developmental discontinuity (Clements, Sarama, & Liu, 2008).

The resultant developmental progression advances through levels of thinking from trial and error, to partial use of geometric attributes, and then mental strategies to synthesize shapes into composite shapes (see the left column of the LT in Figure 1). As an example of the mental actions-on-objects, children at the *Piece Assembler* level intuitively recognize a manipulative shape that corresponds to a distinct outlined shape in a puzzle. With continuous perceptual support, they can use trial and error as they apply slide and turn motions to match the shape to the puzzle outline. The *Piece Assembler*'s recognition of the final composite is based on a provided visual gestalt and is post hoc (Sarama & Clements, 2009). The *Picture Maker* can use a general configuration by mentally filling in one or two missing components of a shape's outline to complete puzzles in which several shapes combine to make a semantic part of a puzzle (e.g., the body of the wagon in Figure 1). When such a gestalt is unavailable, but with consistent perceptual supports, children can maintain an approximate visual image of a side length, using this to choose a shape that matches the side of another shape or one line segment of an outline. This is shown in the right column of Figure 1, which illustrates children choosing a square on the basis of side length and general configuration, then, finding it does not fit the nonsquare region, choosing another shape randomly until it fits. The *Shape Composer* has constructed the figural concept of both side length and angle size and can build, maintain, and manipulate mental images of the shapes, allowing advance planning of the selection and placement of shapes when solving a puzzle.

## *Instruction*

What distinguishes LTs from learning progressions or developmental sequences is that an LT's goals and developmental progressions are inextricable interconnected with instruction (Clements & Sarama, 2014b).

| Developmental Progression | Example Behavior | Instructional | Tasks |
|---|---|---|---|
| **Piece Assembler** Makes pictures in which each shape represents a unique role (e.g., one shape for each body part) and shapes touch.<br>For this study, Target Level – 2, *n* 2 | Make a picture<br><br>Solve a puzzle<br>Fills simple puzzles such as those at the right using trial and error. | In the first “Pattern Block Puzzles” tasks, each shape is not only outlined, but touches other shapes only at a point, making the matching as easy as possible. Students merely match pattern blocks to the outlines.<br><br> | Then, the puzzles moved to those that combine shapes by matching their sides, but still mainly serve separate roles.<br><br> |
| **Picture Maker** Puts several shapes together to make one part of a picture (e.g., two shapes for one arm). Uses trial and error and does not anticipate creation of new geometric shape. Chooses shapes using “general shape” or side length. | Make a picture<br> | The “Pattern Block Puzzles” at this level start with those where several shapes are combined to make one “part,” but internal lines are still available. | Later puzzles in the sequence require combining shapes to fill one or more regions, without the guidance of internal line segments. |

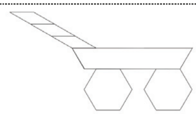
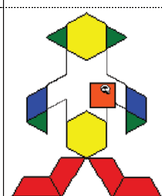
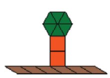
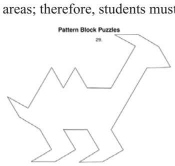
| Developmental Progression | Example Behavior | Instructional | Tasks |
|---|---|---|---|
| For this study, Target Level – 1, *n* 1 | Solve a Puzzle<br>Fills easy puzzles that suggest the placement of each shape (but note to the far right that they student is trying to put a square in the puzzle where its right angles will not fit—this remains a level of “trial and error” strategies). |  |  |
| **Shape Composer**. Composes shapes with anticipation (“I know what will fit!”). Chooses shapes using angles as well as side lengths. Rotation and flipping are used intentionally to select and place shapes.<br>For this study, Target Level, *n* | Make a picture<br><br>Solve a Puzzle<br>Solves puzzles using side and angle recognition and matching are correct | The “Pattern Block Puzzles” have no internal guidelines and larger areas; therefore, students must compose shapes accurately.<br><br> | |
| **Substitution Composer** Makes new shapes out of smaller shapes and uses trial and error to substitute groups of shapes for other shapes to | Make a picture with intentional substitutions | At this level, students solve “Pattern Block Puzzles” in which they must substitute shapes to fill an outline in different ways. | |

*Figure 1.* **Relevant levels from the learning trajectory for composition of geometric shapes (adopted from Clements & Sarama, 2014a; Sarama & Clements, 2009).**

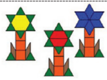
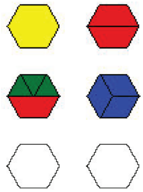
| Developmental Progression | Example Behavior | Instructional Tasks |
|---|---|---|
| create new shapes in different ways.<br>For this study, this is one beyond the target level. |  |  |

*Figure 1.* **(continued)**

Instructional tasks and pedagogical strategies are designed for each level to support children's construction of the mental actions-on-objects underlying that level's pattern of thinking. The tasks include external objects and actions that mirror the hypothesized mental actions-on-objects as closely as possible.

For example, the sequence of instructional tasks for Shape Composition (right column of Figure 1) requires children to solve shape puzzles, the structures of which correspond to the levels of the developmental progression. The mental *objects* are the two-dimensional shapes and the *actions* include creating, copying, comparing, uniting, and disembedding both individual units and composite units. Thus, to progress from the *Piece Assembler* to the *Picture Maker* level, a puzzle might be presented with every internal line drawn except one, which could be missing or only partially drawn. Once the child succeeds, more internal lines would be faded with the expectation that children would incrementally construct the ability to complete known shapes imagistically, disembedding it from the puzzle and understanding how (in this scaffolded context) shapes placed sequentially, usually linearly, unite to create a semantic component of the puzzle. As another example, from the *Picture Maker* to the *Shape Composer* level, puzzles progress to have corners of different angle sizes (at first, salient differences, such as 90° vs. 30°, then less difference) and increase in number of shapes needed to fill regions with no internal lines. Furthermore, the progression is accompanied by simple teaching strategies intended to increase visualization and anticipation, such as "Can you *see* what shapes will fit?"

## Rationale for the Present Research

### *Choice of Topic: Shape Composition*

Given that there have been few other studies examining instruction of shape composition, this study not only provides a foundation for evaluating

LTs but also contributes novel insights into intervention on an understudied aspect of mathematics. We chose to evaluate the efficacy of the shape composition LT because it is important to children's mathematical development as described previously and yet is a topic that receives little instruction in schools.

*Previous Research on LTs*

In a review of methodologically sound evaluations of mathematics curricula, Frye et al. (2013) concluded that interventions with LTs as one component are (as a whole) more efficacious in promoting numeracy than curricula that do not (Frye et al., 2013). For example, Clements and Sarama (2008) found that preschoolers who experienced a curriculum specifically designed on LTs increased significantly more in mathematics competencies than those in a business-as-usual control group (effect size, 1.07) and more than those who experienced a curriculum structured into topic-based units rather than developing all topics (LTs) across the year (effect size 0.47; Clements & Sarama, 2008). Given that the mathematical content of the LT and topical units curricula were quite similar, the difference in efficacy 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 misconceptions or less developed knowledge).

However, although LT-based interventions

> were informed by a developmental progression, no study specifically examined how a teacher's use of a developmental progression affected children's performance on math assessments compared with children who might be taught similar content by a teacher not following a developmental progression. (Frye et al., 2013, p. 84)

That is, previous evaluations or interventions involving LTs (e.g., Clarke et al., 2001; Clements & Sarama, 2007; Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Fantuzzo, Gadsden, & McDermott, 2011; Gravemeijer, 1999; Jordan, Glutting, Dyson, Hassinger-Das, & Irwin, 2012) did not isolate the variable or variables that produced the statistically significant or (as measured by effect size) practically substantially important differences. That is, the studies could not identify the unique contribution of LTs because their impact was confounded by other differences in instructional practices (e.g., the amount of progress monitoring, math talk, or time dedicated to math). For instance, the three curricula evaluated by Clements et al. (2008), the LT curriculum (*Building Blocks*), business as usual (locally developed curricula), and topical-units intervention (*Preschool Mathematics Curriculum*) also differed in organization (e.g., LTs for each topic interwoven throughout the year vs. the other two using separate topical units) and in specific activities used. Therefore, the specific effects of LTs could not be distinguished. To evaluate whether instruction based on LTs is significantly more efficacious than plausible alternatives, we must avoid confounding assumptions of an accepted approach to implementing LTs—using formative assessment to provide instructional activities aligned with empirically validated developmental progressions (Clarke et al., 2001; Clements & Sarama, 2014a; Gravemeijer, 1999; Jordan et al., 2012; Maloney et al., 2014; National Research Council, 2009) with various other instructional factors.

### *Unique Assumptions of an LT Approach in Need of Evaluation*

This widely accepted approach to LT-based instruction has two assumptions that distinguish it from alternative pedagogical approaches.

1. Consistent with Piaget's (1964) principle of assimilation and moderate novelty principle, the first assumption is that instruction should move children from their present level of thinking to the following level, and so forth to the target level. The competing hypothesis is that it is more efficient and mathematically rigorous to provide accurate definitions and demonstrate accurate mathematical procedures using direct instruction, obviating the need for potentially slower movement through each level approach (see Carnine, Jitendra, & Silbert, 1997; Clark, Kirschner, & Sweller, 2012; Wu, 2011). An approach involving direct instruction is popular among practitioners (e.g., more than 50 teachers at various conferences have told us that their principals insist that they teach only end-of-the-year standard skills). That is, direct instruction might efficiently skip one or more of an LT's levels and explicitly teach a target competence (e.g., directly teaching level $n + 2$ procedures to a child operating at level $n$ or even earlier levels). In contrast, LT-based approaches justify the assumption that each contiguous level be taught consecutively because each level is characterized by actions-on-objects that hypothetically must be built at level $n$ as a foundation for effective learning of level $n + 1$ (and thus, if skipped, leave gaps that impede learning).
2. The second assumption, that follows from the first, is that sequencing activities aligned with the developmental progression of a LT results in greater learning than instruction that uses the same activities but sequences them differently. A counterfactual for those studies is a theme-based approach that uses the same activities but in which sequencing is viewed as arbitrary or less important than embedding them in meaningful projects or contexts, such as playing a "pizza game" on the day the class is making pizza (Helm & Katz, 2016; Katz & Chard, 2000; Tullis, 2011).

No research of which we are aware directly tests the two theoretical assumptions of our LTs. The present study serves to rigorously test the first assumption; subsequent studies will test the second assumption. Specifically, we addressed 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 focuses on target level $n + 2$ (the "skip-levels" approach)? We also examined whether child-level variables, such as age, gender, and ethnicity, were moderators on outcomes.

## Method

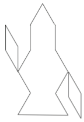
### Participants

Participants were enrolled in a large public school district with a diverse population of elementary school children. Parental consent was obtained for 152 children in 15 prekindergarten (pre-K, 4-year-olds, a.m. and p.m.) classrooms. Of these children, one child scored at the target level on the pretest and was removed before assignment to groups was conducted. An additional six participants (two from the LT intervention group and four from the Skip Levels comparison group) were assigned to condition but did not have valid posttests scores (two left school, the remainder would not provide assent to assessments on three different occasions). The final 145 participants included 82 in the LT intervention group and 63 in the skip-levels comparison group. These children were, on average, 4.62 years old ($SD$ = 0.59; range = 3.38–5.80). Approximately 57% of participants were male; 58% Caucasian, 14% African American, 12% Hispanic, 7% Asian, 3% Indian/Pacific Islander, and 6% other/not reported.

### Measures

Pretest and posttest were a subtest from the REMA (Clements, Sarama, Wolfe, & Day-Hess, 2008/2019) that were designed and verified as assessing the different levels of the developmental progression for shape composition (Clements, Sarama, & Liu, 2008; Clements, Wilson, et al., 2004). For the purposes of this study, we grouped the items into three categories relative to their similarity to the target training tasks. This included both the level of the items in the developmental progression and additional demands the items might include relative to the training tasks. Near transfer items asked children to solve puzzles using manipulatives (e.g., Item 1, see Figure 2). Although the puzzles and shapes differed, these were otherwise isomorphic to the target instructional activities, the *Shape Composer* level. Medium transfer items posed tasks with additional requirements, such telling how many of each component shape would be needed to complete a puzzle or having to fill a puzzle using different shapes (see *Substitution Composer* in Figure 1). Far transfer items were those that had similar additional requirement and also did not provide manipulatives but required children to use mental imagery to compose or decompose shapes (e.g., "How many of which of several drawn figures could be used to make a large figure?").

Give the child the set of pattern blocks, randomly mixed in front of them, and the picture of a puzzle (right). Say: "**Use pattern blocks to fill this puzzle. Put them together with full sides touching.**"



Code 1A (*Very small gaps or misalignments that can be attributed to fine motor limitations are acceptable*)
- 0 = incorrect (placed no shapes *or* placed shapes but not one "fit" the puzzle form, where *fit = at least one side aligned, with no "hangover" outside the puzzle.)*
- 1 = "partially correct" (one or more shapes "fit" but there were one or more gaps or "hangovers")
- 2 = correct (completed puzzle accurately; no gaps or "hangovers")
- NR = no response

Code 1B *For all but 1-2 of the shapes,*
- 0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
- 1 = was hesitant or not systematic (e.g., used cycles of trial and error)
- 2 = completed the puzzle correctly, systematically, but may be "halting"
- 3 = completed the puzzle correctly, immediately, and confidently
- 9 = NR (no response)

Code 1C *For all but 1-2 of the shapes,*
- 0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
- 1 = turned shapes after placing on puzzle in an attempt to get them to fit
- 2 = turned shapes into correct orientation prior to placing them on the puzzle
- 9 = NR

Code 1D *For all but 1-2 of the shapes,*
- 0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
- 1 = tried out shapes by picking them seemingly at random, then putting them back if they did not look right, so seemingly trial and error
- 2 = appeared to search for "just the right shape" that they "know will fit" and then finding and placing it.
- 9 = NR

*Figure 2.* **Item 1 of the shape composition test.**

Graduate research assistants acting as assessors and interventionists had to be certified in pilot administrations to be involved in data collection. Individual child measures were calculated using both the correctness and strategy components of the REMA. Dichotomous correctness responses involved accuracy (such as Code 1A in Figure 2, with NR recoded to zero). Strategy responses included recoding of solution behaviors (such as 1B, 1C, and 1D in Figure 2), for those items that included such codes, along four levels of sophistication ranging from inappropriate/incorrect to very sophisticated. The latter rankings, for example, included observed solution behaviors best suited to solve the problem quickly and correctly. These codes provide greater detail on the processes that children used in solving the problems and allow more accurate assessment of children's thinking (Clements et al., 2008/2019; Clements, Wilson, et al., 2004). Especially because items were constructed and previously validated to assess different levels of the LT (cf. Wilson, 2009) and within a comprehensive assessment (Clements, Sarama, et al., 2008), responses were submitted to Rasch analysis to yield a coherent, unidimensional latent trait (Bond & Fox, 2007; Wright & Stone, 1979).

Equation 1 represents the mathematical formula used in the Rasch–Masters partial credit model (Masters, 1982), expresses the probability, $P_{nij}$, that person $n$ of ability measure $B_n$ is observed in category $j$ of a rating scale specific to item $I$ of difficulty measure of $D_i$ as opposed to the probability $P_{ni(j-1)}$ of being observed in category $(j - 1)$ of a rating scale with categories $j = 0$ (Linacre, 2014).

$$\log_e\left(P_{nij}/P_{ni(j-1)}\right)=B_n - D_{ij} \tag{1}$$

Fidelity of implementation was measured by coding a teaching session according to a rubric. Unacceptable fidelity was coded if an interventionist repeatedly used an incorrect puzzle (for a skip-levels group, a puzzle other than one at the *Shape Composer* level in Figure 1; for the LT group, a puzzle other than one level above the children's operating level) or similarly gave incorrect assistance (for a skip-levels group, modifying a *Shape Composer* level puzzle by drawing internal lines or providing similar gestures; for the LT group, neglecting to modify as necessary for the child's instructional level). Acceptable fidelity was coded if no such errors occurred; Acceptable-with-Corrections was coded if one such error was made.

## Interventions

We developed an elaborated, scripted instructional unit on shape composition following the LT (Figure 1). Instruction was straightforward: children were invited to solve puzzles. A variety of puzzles at the appropriate level were offered to promote child choice and maintain interest. The LT group was offered puzzles and provided scaffolding at the level directly following the level at which they had evinced competence (i.e., $n + 1$, adjusted dynamically). For example, if a child could not solve a problem from a newly introduced level, the interventionist might draw one internal line as a scaffold, then another if needed. The skip-levels group was given puzzles at the target level (*Shape Composer*) without scaffolding that might reduce the level of the task. Both groups were provided encouragement and praise for effort and allowed to switch to a new puzzle (at the appropriate level) if frustrated.

## Procedures

We trained the interventionists to deliver the activities. Interventionists piloted these activities and video recordings of their instruction were reviewed by the authors using the fidelity measure, with feedback given to interventionists individually throughout the intervention. They also recorded the level of thinking they believed the children exhibited and whether they were engaged or showed signs of frustration.

We preassessed all children for whom we had obtained consent and examined the resulting data to determine initial instructional level (leading to elimination of one child). Children within each classroom were randomly assigned to small groups, and then the two to four groups in each classroom were randomly assigned to condition. This design provides control for variance due to community, school, and teachers. In summary, we implemented a three-level randomized block design with fixed effects.

Interventionists then implemented the respective treatments. The authors checked the fidelity of each interventionalist's instruction on all sessions for the first two weeks and 10% of subsequent lessons for each using the fidelity measure, always offering feedback for "fine-tuning" instruction. Fidelity measures revealed adequate fidelity for all but one interventionist (graduate research assistant), and this interventionist's instruction implementation was ultimately deemed acceptable (improving after feedback on the first two sessions), so all data were maintained. Interventionists rated the children's level along the LT's developmental progression after each session. We successfully implemented 1 to 10 days ($M$ = 8.07 days, $SD$ = 1.51) for the 5-week shape composition instructional unit lasting an average of 8.59 minutes per session (including introduction, activities, and transitions). After the instructional period was completed, we posttested all children remaining in the study at the end of the instruction.

## Analytic Procedures

We used a cluster randomized trial design, with children embedded within groups, which are embedded within classrooms. One threat to the validity of the design is contamination across groups within the same classroom. We minimized this through careful separation of groups, parallel administration of the treatment of these groups, and explicit agreement on the part of all teachers that the topic of the treatments would neither be discussed nor dealt with in any way during the intervention period. Randomizing within blocks via the randomized block design (in our case randomizing groups) is more powerful than just randomizing blocks (e.g., classrooms), even if there is very substantial contamination (Rhoads, 2010).

We first computed inferential statistics that account for the original nested structure of the data via multilevel models using M*plus* (Version 7.3, Muthén & Muthén, 1998/2014), which provide correct estimates of effects and standard errors when the data are collected at several levels. This also permits examination of the degree to which child-level relationships vary across schools. We maximized statistical power by controlling for characteristics that may help to explain variability in outcomes, specifically, by using strategic covariates such as baseline (pretest) values of the

*Table 1*
**Means and Standard Deviations of Rasch Measures on Correctness**

| | Learning Trajectories ($n$ = 82) | | Skip-Level Condition ($n$ = 63) | |
|---|---|---|---|---|
| Condition | *M* | *SD* | *M* | *SD* |
| Pretest | −1.89 | 1.81 | −2.14 | 2.17 |
| Posttest | −0.99 | 1.41 | −1.92 | 2.30 |

outcome measures and child-level moderators. We complemented these comparisons with descriptive comparisons of children's correctness and use of processes on every assessment item, using classical scoring.

We first assessed baseline equivalence using a two-level fixed-effect model with pretest measure as the dependent variable and condition at Level 2. Next, we evaluated the unconditional model with posttest mathematical performance as the dependent variable and no included predictors. To evaluate the effect of the LT intervention on children's posttest, compared with the skip-levels children's mathematics performance, we entered the pretest mathematics achievement measure centered at the group level as well as the intervention indicator at the child level. This basic model allows for an examination of the treatment impact alone. We then added child-level covariates including age, race, gender, and time in intervention (measured in minutes). This model used Equation 2.

$$POSTTEST_{ij} = \gamma_{00} + \gamma_{01} * CONDITION_j + \gamma_{02} * GROUPPRE_j + \gamma_{10} * AGE_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * RACE_{ij} + \gamma_{40} * PRETEST_{ij} + \gamma_{50} * TIME_{ij} + u_{0j} + r_{ij} \quad (2)$$

## Results

This first set of analyses was conducted using the Rasch measures that were based on both correctness and process behaviors (e.g., strategies). Means and standard deviations by group are presented in Table 1. The two-level fixed-effect model indicated that the two conditions were not significantly different at pretest ($\beta$ = .089; $p$ = .828), supporting baseline equivalence. The unconditional model indicated that about the majority of the variance (intraclass correlation coefficient = 24%) in the posttest measures lay between groups ($r^2$ = .975, $p$ = .016; $g$ = .24).

The basic model comparing the LT and skip-levels interventions indicated that the pretest is a significant predictor of the posttest measure ($\beta$ = .807, $p < .0001$). The treatment indicator was also significant ($\beta$ = .481, $p$ = .003). The difference between the treatment and control group represents a substantial effect ($g$ = .55).

The model that added child-level covariates including age, race, gender, and time in intervention indicated that only treatment group remained a significant predictor of posttest outcomes ($\beta$ = .500; $p$ = .007). Specifically, no impact for gender ($\beta$ = .298; $p$ = .515), age ($\beta$ = −.084; $p$ = .834), ethnicity ($\beta$ = −1.59; $p$ = .767), or time on task ($\beta$ = −.332; $p$ = .428) was found on posttest measures controlling for pretest measures at both the school and child level.

We then explored the differences of the two groups on each item. We first present the results of a single item, #1 (see Figure 2, ideally solved with target-level competencies) in detail. On the pretest, the two groups were balanced across the correctness measure (A) as well as the other codes. In comparison, at posttest, only 3 (4% LT) compared with 9 (13% Skip) children were completely incorrect at posttest; and 44 (59%) compared with 23 (35%) were completely correct. The process codes tell a similar story. Pretest distributions are similar, but posttest distributions differ; the LT group showed greater frequency of the more sophisticated strategies than the skip-levels group.

Table 2 includes means and standard deviations for all items categorized according to levels of transfer. For the near transfer items, both groups made gains on all items with consistent differences in favor of the LT group on correctness and the sophistication of their solution processes. For the medium transfer items, gains of both groups were smaller. Relative gains (or performance on the posttest-only items) in favor of the LT group were similar for items (4 and 6) that used the same shapes that children used in the training sessions, but smaller on (near zero for two of the three) items that used different shapes (5, 7, and 8). For the far transfer items, gains were negligible and there were no reliable differences between the groups, with the skip-levels group making slightly greater gains on one of the two items (9). Finally, the interventionists' qualitative field notes show a clear indication that the skip-levels group expressed more counterproductive frustration than the LT group, including on target-level tasks.

## Discussion

Although LTs have received increasing attention from policy makers, educators, curriculum developers, and researchers and are deemed as a useful tool for guiding standards, instructional planning, and assessment (Frye et al., 2013; National Research Council, 2009), little research has directly and rigorously tested the specific contributions of LTs to student learning. For example, even successful projects based on LTs (e.g., Clarke et al., 2001; Clements et al., 2011; Clements & Sarama, 2008; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Murata, 2004; Wright, Stanger, Stafford, & Martland, 2006) confound the use of LTs with other factors and thus cannot identify the unique contributions of the LTs per se. Our research design

*Table 2*

**Means and Standard Deviations of Correct (A Codes) and Process (B, C, D Codes) for Items Categorized by Level of Transfer**

| | | Learning Trajectory Condition | | | | Skip-Level Condition | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pretest | | Posttest | | Pretest | | Posttest | |
| Item | Description | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| | Near transfer | | | | | | | | |
| 1A | See Figure 2. | 0.93 | 0.64 | 1.59 | 0.57 | 0.89 | 0.60 | 1.21 | 0.73 |
| 1B | | 1.05 | 0.75 | 1.70 | 0.77 | 0.90 | 0.78 | 1.48 | 0.97 |
| 1C | | 0.95 | 0.65 | 1.47 | 0.62 | 0.87 | 0.71 | 1.30 | 0.75 |
| 1D | | 1.08 | 0.74 | 1.57 | 0.60 | 0.95 | 0.73 | 1.37 | 0.77 |
| 2 | Similar to Item 1 | 0.53 | 0.55 | 0.87 | 0.72 | 0.53 | 0.56 | 0.71 | 0.68 |
| 2B | | 0.55 | 0.58 | 1.05 | 0.92 | 0.54 | 0.65 | 0.85 | 0.91 |
| 2C | | 0.56 | 0.60 | 0.88 | 0.70 | 0.51 | 0.60 | 0.79 | 0.82 |
| 2D | | 0.65 | 0.74 | 1.05 | 0.84 | 0.61 | 0.71 | 0.84 | 0.86 |
| 3 | Similar to Item 1 | 0.59 | 0.52 | 0.87 | 0.60 | 0.57 | 0.50 | 0.76 | 0.62 |
| 3B | | 0.45 | 0.61 | 0.87 | 0.78 | 0.51 | 0.65 | 0.73 | 0.83 |
| 3C | | 0.45 | 0.61 | 0.83 | 0.68 | 0.53 | 0.58 | 0.64 | 0.71 |
| 3D | | 0.49 | 0.73 | 0.90 | 0.82 | 0.57 | 0.71 | 0.69 | 0.79 |
| | Medium transfer | | | | | | | | |
| 4 | Fill identical puzzles in different ways. | 0.07 | 0.30 | 0.28 | 0.58 | 0.05 | 0.21 | 0.11 | 0.36 |
| 5 | Use 4 of 6 shapes to fill puzzle. | 0.20 | 0.40 | 0.34 | 0.48 | 0.14 | 0.35 | 0.25 | 0.44 |
| 6 | How many of one shape will fill another. | 0.09 | 0.29 | 0.20 | 0.40 | 0.16 | 0.37 | 0.17 | 0.38 |
| 6B | From trial and error to immediate correct answer. | 0.14 | 0.48 | 0.47 | 0.84 | 0.30 | 0.66 | 0.25 | 0.65 |
| 7 | How many of which shapes needed to fill puzzle. | — | — | 0.00 | 0.00 | — | — | 0.02 | 0.13 |
| 8 | How many of which shapes needed to fill puzzle. | — | — | 0.18 | 0.39 | — | — | 0.10 | 0.30 |
| | Far transfer | | | | | | | | |
| 9 | Choose shape created by composing shapes. | 0.32 | 0.4 | 0.34 | 0.48 | 0.29 | 0.46 | 0.38 | 0.49 |
| 10 | Choose shapes created by decomposing shape. | 0.17 | 0.38 | 0.20 | 0.40 | 0.16 | 0.37 | 0.17 | 0.38 |

allowed us to test a key assumption of a widely used implementation approach to LTs (Clarke et al., 2001; Clements & Sarama, 2014a; Gravemeijer, 1999; Jordan et al., 2012; Maloney et al., 2014; National Research Council, 2009) by creating a counterfactual that alters only one factor. Specifically, in the present study, we assessed one such assumption, designing sequences of instruction that follow the levels of a LTs developmental progression, by evaluating the efficacy of instruction in which LT levels are taught consecutively (for children at level $n$, instructional tasks from level $n + 1$, then $n + 2$) compared with instruction that immediately and solely targets level $n + 2$ (the "skip-levels" approach). Thus, although both interventions had the same target instruction (goal and instruction), the LT intervention embodied a key assumption of our LT approach whereas the counterfactual did not: In the LT intervention, children were taught levels consecutively, whereas in the skip-levels condition, participants were taught the target level exclusively.

Critical for this evaluation was the use of a theoretically and empirically supported LT including three, interrelated components: goal, developmental progression, and instruction. This allows the research question and procedures (e.g., assessment, teaching) to be based on clear conceptual foundation. We used an LT with extensive support in the literature (Casey, Erkut, Ceder, & Young, 2008; Clements et al., 2011; Clements, Wilson, et al., 2004; Mansfield & Scott, 1990; Sales, 1994; Sarama et al., 1996; The Spatial Reasoning Study Group, 2015). The mathematical topic, the composition of shape, is significant in that the concepts and actions of creating and then iterating units and higher order units in the context of constructing patterns, measuring, and computing are established bases for mathematical understanding and analysis (Sarama & Clements, 2009). Additionally, there is evidence that this type of composition corresponds with, and may support, other mathematical competencies (Clements et al., 1996; Razel & Eylon, 1990, 1991; Reynolds & Wheatley, 1996; Steffe & Cobb, 1988; The Spatial Reasoning Study Group, 2015).

Although instruction was brief, consisting of an average of a little more than eight 9-minute sessions over 5 weeks, we found that the LT treatment was more effective than skip-levels treatment. Using Rasch measures that incorporated use of processes as well as correctness, the effect size for the difference between groups was .55. There were no significant differences on outcomes for the variables of gender, age, ethnicity, or time on task, indicating a robust and general result.

Examination of individual items confirmed that the LT group made more completely correct solutions to the assessment items and used strategies at higher levels of sophistication than children in the skip-level group. These effects were especially pronounced on tasks similar to the target level (level $n + 2$), that is, on near transfer tasks. This is notable, as the target level was achieved more frequently by children in the LT group who experienced *fewer* tasks and less instructional time at that level ($n$ + 2) than those in the skip-levels group who spent all their time on tasks at level $n$ + 2.

However, the benefits of the LT treatment did not extend to all medium transfer items. These items posed tasks with additional requirements, such as naming how many of each component shape would be needed to complete a puzzle or to fill a puzzle using different shape combinations. The LT group made greater gains on some of these items, but only when the shapes used were the same as in the training. When other shapes were used, differences between the groups were small and usually inappreciable. Thus, the LT group evinced more transfer, but to a limited degree.

The effects of the LT treatment did not extend to far transfer (on which the groups performed similarly on one, but the skip-levels gained a bit more than the LT group on the other). Far transfer items did not use manipulatives but required children to use mental imagery to combine shapes or decomposition. It is possible that the target-level tasks presented to the children in the skip-levels group stimulated them to use spatial imagery; that is, these children may have more frequently attempted to visualize where shapes would fit in the challenging puzzles, leading to an increase in spatial imagery. However, a conservative interpretation is that neither treatment affected performance on far transfer items. Given the modest amount of time to learn the target level, near but not far transfer might be expected. Future research should investigate if a greater number of sessions will promote medium and far transfer and if either the LT or skip-levels approach promotes far transfer more than the other.

We also investigated whether entering knowledge or other individual child-level factors were significant moderators of differences. None of the child-level moderators, including age, gender, ethnicity, or time in intervention sessions were significant nor were the group-level moderator, the interaction of intervention condition by group pretest significant, when entered together or separately.

Beyond growth in children's knowledge, the skip-levels group expressed more counterproductive frustration than the LT group. This may indicate that instruction that was provided beyond a child's level is not only ineffective, but also counterproductive as it may increase a child's aversion to mathematics. In future research, we intend to code such affect responses systematically.

Several caveats should be noted. First, instruction was provided by trained interventionists to small groups, not teachers to full classes. Future research could check our theoretically motivated results with studies using entire classrooms as the unit of analysis. In a similar vein, it could be argued that there are many other approaches to teaching the topic at hand, and that the comparison intervention was artificial. However, our goal was to provide a clear, precise test one of two main assumptions of our LT approach, rather than to find an "ideal" approach. Also, our counterfactual is one that has been theoretically and practically justified (recall Clark et al., 2012; Carnine et al., 1997; Wu, 2011, and the many teachers who are asked to teach target-level skills only). A second caveat is that results are limited to one domain of mathematics; future research must involve other domains, as it is possible that the more effective method of instruction varies by topic. Third, although we assessed the effect of several possible moderators, it is possible that effects would be different for populations with different inter- or intraindividual differences. Our future studies will investigate some of these issues, but much work remains to be done.

Although the results of this study will have implications for the use of LTs across multiple domains (e.g., Alonzo & Gotwals, 2012; National Research Council, 2007), the domain of early mathematics is particularly important and fecund for this research. LTs have played a substantial role in mathematics education (Simon, 1995). They were the explicit core construct in the NRC (National Research Council, 2009) report on early mathematics (note the subtitle: "Paths toward excellence and equity"), played a similar role in writing standards (e.g., National Council of Teachers of Mathematics, 2006; NGA/CCSSO, 2010), and have been successfully applied in early mathematics intervention projects (e.g., Clarke et al., 2001; Clements et al., 2011; Clements & Sarama, 2008; Cobb et al., 2003; Murata, 2004; Wright et al., 2006).

This first experiment provides a rigorous evaluation on one critical research question concerning the relative effectiveness of a LT versus a target-only, or skip-levels, approach. Findings indicate that teaching each contiguous level of a LT is more efficacious and thus useful but not necessary. However, because we do not know if this result will generalize to other defining assumptions of this approach to LTs or to other topics or ages of children, we will continue to conduct a series of studies. This study clearly shows that children learned more about levels $n$ to $n + 2$ overall and achieved greater transfer with the target level (level $n + 2$) in particular by following a LT than by focusing solely on the target level of thinking.

## Note

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. Although the research is concerned with theoretical issues, not particular curricula, a small component of the intervention used in this research have been published by some of the authors, who thus could have a vested interest in the results. 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, children who participated in this research, graduate research assistants who helped implement it, Douglas Van Dine, who oversaw initial data collection, and David Purpura, who helped with initial analyses.

## 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). Rotterdam, Netherlands: Sense. doi:10.1007/978-94-6091-824-7_20

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.), *Handbook of early childhood care and education* (pp. 329–354). Hoboken, NJ: Wiley Blackwell.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum. doi:10.1111/j.1745-3984.2003.tb01103.x

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*, 66–81.

Casey, B. M., Erkut, S., Ceder, I., & Young, J. M. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology*, *29*, 29–48.

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*, 6–11.

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 Incorporated* (Vol. 1, pp. 9–26). Reston, Australia: MERGA.

Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *Elementary School Journal*, *98*, 171–186.

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*, 136–163.

Clements, D. H., & Sarama, J. (2007/2013). *Building blocks, Volumes 1 and 2*. Columbus, OH: McGraw-Hill Education.

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

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

Clements, D. H., & Sarama, J. (2014b). Learning trajectories: Foundations for effective, research-based education. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 1–30). New York, NY: Information Age.

Clements, D. H., Sarama, J., Battista, M. T., & Swaminathan, S. (1996). Development of students' spatial thinking in a curriculum unit on geometric motions and area. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the 18th Annual Meeting of the North America Chapter of the International Group for the*

*Psychology of Mathematics Education* (Vol. 1, pp. 217–222). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

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., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Maths Assessment. *Educational Psychology*, *28*, 457–482. doi:10.1080/01443410701777272

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*, 127–166. doi:10.5951/jresematheduc.42.2.0127

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

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

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*, 9–13.

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.

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. (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.

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

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

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*, 647–660. doi:10.1037/a0029018

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

Lesh, R. A., & Yoon, C. (2004). Evolving communities of mind—in which development involves several interacting and simultaneously development strands. *Mathematical Thinking and Learning*, *6*, 205–226.

Linacre, J. M. (2014). *A user's guide to Winsteps/Ministep Rasch-model computer program*. Chicago IL: Winsteps.com.

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

Mansfield, H. M., & Scott, J. (1990). Young children solving spatial problems. In G. Booker, P. Cobb, & T. N. deMendicuti (Eds.), *Proceedings of the 14th annual conference of the Internation Group for the Psychology of Mathematics*

*Education* (Vol. 2, pp. 275–282). Oaxlepec, Mexico: International Group for the Psychology of Mathematics Education.
Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.
Mueller, U., Sokol, B., & Overton, W. F. (1999). Developmental sequences in class reasoning and propositional reasoning. *Journal of Experimental Child Psychology*, *74*, 69–106.
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.
Muthén, L. K., & Muthén, B. O. (1998/2014). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: Author.
National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards*. Washington, DC: Author.
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 Academies Press. doi:10.17226/12519
Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered* (pp. 7–20). Ithaca, NY: Cornell University.
Razel, M., & Eylon, B.-S. (1990). Development of visual cognition: Transfer effects of the Agam program. *Journal of Applied Developmental Psychology*, *11*, 459–485.
Razel, M., & Eylon, B.-S. (1991, July). *Developing mathematics readiness in young children with the Agam Program*. Paper presented at the meeting of the Fifteenth Conference of the International Group for the Psychology of Mathematics Education, Genova, Italy.
Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.
Reynolds, A., & Wheatley, G. H. (1996). Elementary students' construction and coordination of units in an area setting. *Journal for Research in Mathematics Education*, *27*, 564–581.
Rhoads, C. H. (2010). The implications of "contamination" for experimental design in education. *Journal of Educational and Behavioral Statistics*, *36*, 76–104. doi:10.3102/1076998610379133
Sales, C. (1994). *A constructivist instructional project on developing geometric problem solving abilities using pattern blocks and tangrams with young children* (Unpublished master's thesis). University of Northern Iowa, Cedar Falls.
Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.
Sarama, J., Clements, D. H., & Vukelic, E. B. (1996). The role of a computer manipulative in fostering specific psychological/mathematical processes. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the 18th Annual Meeting of the North America Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 567–572). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*, 114–145. doi:10.2307/749205

The Spatial Reasoning Study Group. (2015). *Spatial reasoning in the early years: Principles, assertions, and speculations*. New York, NY: Routledge.
Steffe, L. P., & Cobb, P. (1988). *Construction of arithmetical meanings and strategies*. New York, NY: Springer-Verlag.
Steffe, L. P., Thompson, P. W., & Glasersfeld, E. V. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267–306). Mahwah, NJ: Erlbaum.
Tullis, P. (2011). The death of preschool. *Scientific American Mind*, *22*(5), 36–41.
Vurpillot, E. (1976). *The visual world of the child*. New York, NY: International Universities Press.
Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*, 716–730.
Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
Wright, R. J., Stanger, G., Stafford, A. K., & Martland, J. (2006). *Teaching number in the classroom with 4-8 year olds*. London, England: Paul Chapman/Russell Sage.
Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics*. Providence, RI: American Mathematical Society.

Manuscript received January 28, 2018
Final revision received January 31, 2019
Accepted March 18, 2019