Journal for Research in Mathematics Education
2013, Vol. 44, No. 3, 581–620

Young Children's Understandings of Length Measurement: Evaluating a Learning Trajectory

Janka Szilágyi
The College at Brockport, State University of New York

Douglas H. Clements and and Julie Sarama
University of Denver

This study investigated the development of length measurement ideas in students from prekindergarten through 2nd grade. The main purpose was to evaluate and elaborate the developmental progression, or levels of thinking, of a hypothesized learning trajectory for length measurement to ensure that the sequence of levels of thinking is consistent with observed behaviors of most young children. The findings generally validate the developmental progression, including the tasks and the mental *actions on objects* that define each level, with several elaborations of the levels of thinking and minor modifications of the levels themselves.

Key words: Children's strategies; Clinical interviews; Conceptual knowledge; Early childhood; Item-response theory; Measurement

Research on learning trajectories (LT) has the potential to help connect the processes of teaching and learning. One promising approach to the creation of a LT begins by building a developmental progression, or sequence of levels of thinking, through which most children progress in the learning of a particular mathematical topic. Developmental progressions can guide teachers in making sense of their students' understandings of a concept (Confrey, 1990) and in choosing appropriate instructional activities (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Fennema et al., 1996) that help children move along these progressions (Clements, Sarama, & DiBiase, 2002; Sarama & Clements, 2009). To accomplish this, developmental progressions must accurately and clearly describe the levels of thinking through which most children proceed. Therefore, the

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305K05157, "Scaling Up TRIAD: Teaching Early Mathematics for Understanding With Trajectories and Technologies" and in part by the National Science Foundation through Grant No. DRL-0732217, "A Longitudinal Account of Children's Knowledge of Measurement." The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or the National Science Foundation. This research was initiated when the authors were affiliated with the University of Buffalo, The State University of New York.

validation and refinement of such developmental progressions is a critical component of such research-and-development work. In this study, we evaluated and elaborated the developmental progression of a hypothesized learning trajectory for the early learning of length measurement.

Learning Trajectories and Their Developmental Progressions

There are three components of a LT: "the learning goal, the learning activities, and the thinking and learning in which students might engage" (Simon, 1995, p. 133). Although Simon initially emphasized the individual teacher's construction of a LT for a particular teaching episode, other researchers (Baroody, 2004; Battista, 2004; Clements, Wilson, & Sarama, 2004) subsequently emphasized the importance of general LTs that may provide a foundation for common, systematic teaching practice (Raudenbush, 2009) and curriculum design (Clements, 2007).

Such shared LTs should be based on a generalizable model of learning. Existing models have had diverse sources, including historical developments of mathematics, thought experiments, the practices of "successful" teachers, and wide-ranging theoretical positions (Clements, 2008). Clements and Sarama (2004b) argued that generalizable models should be based, whenever possible, on empirical research. They defined LTs as having research-based developmental progressions at their core:

> We conceptualize learning trajectories as 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, created with the intent of supporting children's achievement of specific goals in that mathematical domain. (p. 83)

In the theory of *hierarchical interactionalism* (Clements & Sarama, 2007a, 2009; Sarama & Clements, 2009), of which such learning trajectories are a core component, levels of thinking are coherent and characterized by increased sophistication, complexity, abstraction, and generality. However, the learning process is viewed not as intermittent and tumultuous but as more incremental, with knowledge becoming integrated slowly. In this way, various models and types of thinking grow in tandem to a degree, but a critical mass of ideas from each level must be constructed before the thinking characteristic of the subsequent level becomes ascendant in the child's thinking and behavior (Sarama & Clements, 2009). These ideas can be characterized by specific mental objects (e.g., concepts) and actions (processes; Clements et al., 2004; Steffe & Cobb, 1988). These actions-on-objects are children's main way of operating in the domain. Different developmental courses are possible within those constraints, depending on individual, environmental, and social confluences (Clements, Battista, & Sarama, 2001; Confrey & Kazak, 2006). The differences within and across individuals create variation that is the wellspring of invention and development. At a group level, however, these variations are not so wide as to vitiate the theoretical or practical usefulness of the tenet of developmental progressions; for example, in a class of 30, there may be only a handful of different solution strategies (Murata & Fuson, 2006), many of which represent different levels along the developmental progression (for a complete explication, see Clements & Sarama, 2009; Sarama & Clements, 2009).

Information about the developmental progressions of LTs is not only helpful in understanding how children advance but also provides guidance for teachers in choosing appropriate activities to help children move along the progressions (Clements et al., 2002). Similarly, LTs can serve as the foundation of research-based curriculum development (Clements, 2007, see also Bredekamp, 2004; Clements, Sarama, et al., 2002; National Association for the Education of Young Children & National Council of Teachers of Mathematics, 2002). The current study evaluated a developmental progression for length measurement, previously crafted as a foundation for standards, assessments, and curricula, such as the *Building Blocks* mathematics curriculum (Clements & Sarama, 2004a, 2007c, 2009; Sarama & Clements, 2003, 2009).

Length Measurement

Length is a comparative property of objects that embodies the amount of one-dimensional space between endpoints of the objects, which can be compared or quantified (measured). Piaget, Inhelder, and Szeminska (1960) defined the concept of length measurement as the synthesis of subdivision and change of position, which involves taking one part out of the whole and iterating that unit along the whole. Although useful in analyzing higher level strategies of length measurement, this definition fails to acknowledge the richness of understandings that young children possess. To analyze strategies at different levels, researchers have posited various ideas and competencies as necessary for a partial or full understanding of length measurement. These include awareness of the attribute, equal units, unit–attribute relations, partitioning, unit iteration, origin (zero point), transitivity, conservation, accumulation of distance, proportionality, and the relation between number/arithmetic and measurement (Clements & Stephan, 2004; Lehrer, 2003; Piaget et al., 1960; Stephan & Clements, 2003).

Piaget et al. (1960) and Piaget and Szeminska (1952) inspired considerable research on such reasoning abilities as transitivity and conservation of length (Bearison, 1969; Braine, 1959; Kidder & Lamb, 1981; Miller & Baillargeon, 1990; Sawada & Nelson, 1967; Schiff, 1983), and the relationship between those reasoning abilities and specific measurement concepts (Boulton-Lewis, 1987; Hiebert, 1981; Petitto, 1990). Other studies examined developmental sequences for length measurement competencies of various types (Boulton-Lewis, Wilss, & Mutch, 1996; Clements, 1999b; Kamii & Clark, 1997).

These research corpi, along with research on instructional sequences (Clarke, Cheeseman, McDonough, & Clark, 2003; McClain, Cobb, Gravemeijer, & Estes, 1999; Nunes, Light, & Mason, 1993; Outhred, Mitchelmore, McPhail, & Gould, 2003; Stephan, Cobb, Gravemeijer, & Estes, 2001), were reviewed and synthesized to create a LT for length measurement for the Building Blocks project, as depicted

Developmental progression	Actions on objects	Representative task
Length Quantity Recognizer Identifies length/distance as an attribute. May understand length as an absolute descriptor (e.g., all adults are tall), but not as a comparative (e.g., one person is taller than another). "I'm tall, see?" May compare noncorresponding parts of a shape in determining side length.	Action schemes are connected to length vocabulary. In some situations, such vocabulary is connected to categories of linear extent, such as "tall/long" or "short." In others, action schemes are used to compare lengths—one object is longer if a scan lasts perceptibly longer than the scan of another object. Thus, intuitive comparisons are made on direct perceptual, normative (one object can be a class standard stored in memory, such as a doll's length), or functional (if guided/prompted; e.g., "Is this block long enough?") bases. However, in some situations salient differences at one end of the objects are substituted for a scan (potentially leading to inaccuracies if the other endpoints are not aligned). Also, irrelevant details such as the shape of objects can affect these categorizations and comparisons.	Length Puzzle Give the child a rectangular foam sheet that has one rectangular opening in the middle, and a foam strip that is exactly as wide as the opening, but has a shorter length than that of the opening. Ask the child to put the strip in the opening to solve the puzzle, and ask whether the strip fits well in the opening. Children at this level will indicate that the strip does not fit well, referring to length in some way in their explanations.
Length Direct Comparer Physically aligns two objects to determine which is longer or whether they are the same length. Stands two sticks up next to each other on a table and says, "This one's bigger." 	The scheme addresses length as the distance between the endpoints of a path. Shape of the objects and path can affect the application of the scheme. With perceptual support, objects can be mentally and then physically slid and rotated into alignment and their endpoints compared.	Two Pencils Show the child two colored pencils, the shorter one horizontal, and the longer one vertical, in a —| shape. Ask the child to show which of the pencils is longer. Children at this level will compare the length of the two pencils by aligning their endpoints.

Figure 1 (*continued*)

Developmental progression	Actions on objects	Representative task
Indirect Length Comparer Compares the length of two objects by representing them with a third object. Compares length of two objects with a piece of string. When asked to measure, may assign a length by guessing or moving along a length while counting (without equal length units). Moves finger along a line segment, saying 10, 20, 30, 31, 32. May be able to measure with a ruler, but often lacks understanding or skill (e.g., ignores starting point) Measures two objects with a ruler to check whether they are the same length, but does not accurately set the "zero point" for one of the items.	A mental image of a particular length can be built, maintained, and (to a simple degree) manipulated. With the immediate perceptual support of some of the objects, such images can be compared. For some, explicit transitive reasoning may be applied to the images or their symbolic representations (i.e., object names). If asked to measure, a counting scheme operates on an intuitive unit of spatial extent or amount of movement, directing physical movements (or, less frequently, eye movements) along a length while counting (resulting in a trace-and-count or point-and-count strategy). The sensory-concrete mental actions require the perceptual support of the object to be measured.	Picture of Two Pencils Give the child a picture of two pencils, as well as a foam strip that is longer than each of the pencils. Ask which of the pencils is longer. Children at this level will use the strip well to compare the length of the two unmovable pencils, as well as explain their strategy. 

Figure 1. The developmental progression and mental actions-on-objects for the learning trajectory for length measurement (Clements & Sarama, 2004a, 2007c, 2009; Sarama & Clements, 2003, 2009).

Figure 1 (*continued*)

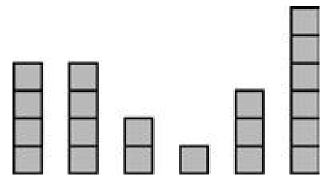
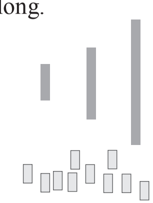
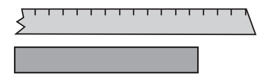
Developmental progression	**Actions on objects**	**Representative task**
Serial Orderer to 6+ Orders lengths, marked in 1 to 6 units. (This develops in parallel with "End-to-end Length Measurer.") Given towers of cubes, puts in order, 1 to 6.	Scheme is organized in a hierarchy, with the higher order concept a (possibly implicit) image of an ordered series. Ability to estimate relative lengths (driving a trial-and-error approach) is eventually complemented by a scheme that considers each object in such a series to be longer than the one before it and shorter than the one after it (resulting in a more efficient strategy).	Order Towers Give the child six premade connecting cube towers of different lengths. Ask the child to put them in order from shortest to longest. Children at this level will use a meaningful strategy to correctly align the towers. 
End-to-end Length Measurer Lays units end to end. May not recognize the need for equal-length units. The ability to apply resulting measures to comparison situations develops later in this level. (This develops in parallel with "Serial Orderer to 6+.") Lays nine 1-inch cubes in a line beside a book to measure how long it is.	An implicit concept that lengths can be composed as repetitions of shorter lengths underlies a scheme of laying lengths end to end. (This scheme must overcome previous schemes, which use continuous mental processes to evaluate continuous extents, and thus are more easily instantiated.) This initially applied only to small numerosities (e.g., 5 or fewer units). Starting with few restrictions (i.e., only weak intuitive constraints to use equal-size units or to avoid gaps between "units"), the scheme is enhanced by the growing conception of length measuring as covering distance (or composing a length with parts) with further application of these constraints.	Blue Strips Give the child a 2-, a 4-, and a 7- inch-long blue foam strip, and 10 yellow strips that are each 1 inch long. Ask which of the three strips is the same length as four of the yellow strips. Children at this level will correctly use the units to decide which of the longer strips is 4 yellow strips long. 

Figure 1 (*continued*)

Developmental progression	**Actions on objects**	**Representative task**
Length Unit Repeater Measures by repeated use of a unit (but initially may not be precise in such iterations).	Action schemes include the ability to iterate a mental unit along a perceptually available object. The image of each placement can be maintained while the physical unit is moved to the next iterative position (initially with weaker constraints on this placement).	Spaghetti Give the child a picture of spaghetti and one foam strip. Ask the child how long the spaghetti is when measured with the strip, which is 1 inch long. Children at this level will correctly iterate the unit to get the measure; some imprecision may occur due to use of finger.
Length Unit Relater Relates size and number of units explicitly (but may not appreciate the need for identical units in every situation). "If you measure with centimeters instead of inches, you'll need more of them, because each one is smaller." Recognizes that different units will result in different measures and that identical units should be used, at least intuitively and/or in some situations. Uses rulers with minimal guidance. Measures a book's length accurately with a ruler.	With the support of a perceptual context, scheme can predict that fewer larger units will be required to measure an object's length.	Different Feet Introduce two friends, who have visibly different-size feet. Tell the child that when one of the friends measured the rug with his feet, he found that the rug was 4 feet long, and when the other friend measured, the rug was 9 feet long. Ask the child how it was possible that their results were different when they measured the same rug. Children at this level will correctly point out the difference in the size of feet, and explain the result of measuring with different units.

Figure 1 (*continued*)

Developmental progression	Actions on objects	Representative task
Length Measurer Measures, knowing need for identical units, relationship between different units, partitions of unit, zero point on rulers, and accumulation of distance. Begins to estimate. "I used a meter stick three times, then there was a little left over. So, I lined it up from 0 and found 14 centimeters. So, it's 3 meters, 14 centimeters in all."	The length scheme has additional hierarchical components, including the ability to simultaneously imagine and conceive of an object's length as a total extent and a composition of units. This scheme adds constraints on the use of equal-length units and, with rulers, on use of a zero point. Units themselves can be partitioned, allowing the accurate use of units and subordinate units.	Broken Ruler Show the child a ruler that is broken before the 2-inch mark, and a 5-inch-long foam strip. Ask the child to use the broken ruler to measure the foam strip.  Children at this level will use and explain a correct strategy to get the correct measure.
Conceptual Ruler Measurer Possesses an "internal" measurement tool. Mentally moves along an object, segmenting it, and counting the segments. Operates arithmetically on measures ("connected lengths"). Estimates with accuracy. "I imagine one meter stick after another along the edge of the room. That's how I estimated the room's length is 9 meters."	Interiorization of the length scheme allows mental partitioning of a length into a given number of equal-length parts or the mental estimation of length by projecting an image onto present or imagined objects.	Length of Sign Show the child a picture of a building with a sign. Ask the child how long the sign is if the length of the building and the lengths on the two sides of the sign are known. Children at this level will use the operations of addition and subtraction to calculate the length of the sign.

in the first two columns of Figure 1 (Clements & Sarama, 2007b, 2007c; Sarama & Clements, 2002; Sarama & DiBiase, 2004). As stated, the foundation of such LTs is a research-based developmental progression. In this study, we applied both Rasch measurement and qualitative methods to evaluate the developmental progression component of the LT created as a foundation for the research-based development of the *Building Blocks* early childhood mathematics curriculum (Clements & Sarama, 2004a, 2007c, 2009; Sarama & Clements, 2003, 2009). There were two main research questions:

1. Do the levels of the developmental progression provide a valid description of most young children's acquisition of concepts and strategies related to length measurement? That is, is the Rasch measurement-based empirical hierarchy of the assessment items consistent with the hypothesized hierarchy?
2. What mental actions-on-objects constitute each level of thinking?

Method

Participants

A convenience sample of 121 children was recruited from three schools in two countries. We used participants from two countries to ensure the developmental progression was not bound to one culture. The three schools involved in the study included one public school in a low-socioeconomic rural area in western New York, and a preschool and an elementary school[1] in a small town in eastern-central Hungary. The U.S. participants were 18 prekindergartners, 18 kindergartners, 19 first graders, and 25 second graders; the Hungarian participants were 8 prekindergartners, 16 first graders, and 17 second graders.[2] All participants returned signed consent forms and completed all interview assessment tasks.

Assessment Tasks

Tasks were selected or designed to elicit responses reflective of children's thinking and understanding in terms of the developmental progression component of the *Building Blocks* learning trajectory for length (Clements & Sarama, 2007c). Items were selected from existing literature whenever available, and additional tasks were written by the researchers to complete the assessment of each level of thinking (see the third column in Figure 1 for a detailed description of the items). Two tasks were included for each level of thinking (see Figure 1). The initial instrument was piloted with a small number of children (not participants in the final

[1] In Hungary, formal schooling starts with first grade, which children enter at the age of 6. Before first grade, children in Hungary attend a mandatory school-preparatory program for a year, which is hosted by preschools (although children in these programs match U.S. kindergartners in age, the program is different in nature from U.S. kindergartens). Preschools in Hungary also provide programs for younger age groups.

[2] The prekindergartners, the first graders, and the second graders in this study were comparable in age in both countries, respectively. No kindergartners from Hungary participated in the study.

study), and the wording and the materials used for some of the items were altered to increase the understandability and usefulness of each item (see Szilágyi, 2007).

Because the tasks involved have a crucial role in giving meaning to the developmental progression and in shaping the hypothesis of the learning process, it is important to note that the proposed task sequence is simply one possible plausible path out of many (Clements & Sarama, 2004b).

Procedures

Regardless of age, all children were administered all tasks. The interviewer moved on to the next task even if students did not complete a task accurately. Each assessment was delivered in one sitting.

All participants were individually interviewed by one of the researchers, and all interviews were videotaped. The task-based interviews (Goldin, 1997, 2000) began with a script, but they were also enhanced with Piaget's method of clinical interviewing (Ginsburg, 1997). To ensure that open-ended interactions did not influence the children's responses to other tasks, clinical probes were administered only after the end of the scripted interview. Such mixed-method interviewing has been used successfully to validate developmental progressions (Clements et al., 2004).

The assessment period took approximately 5 weeks in the United States, and 3 weeks in Hungary. The goal was to complete the interviews close together in time to minimize potential instruction during the course of the interviewing period. All interviews in a particular classroom were done within 1 or 2 days, and all the interviews were completed during the months of May and June, shortly before the summer break started for the children in both countries.

Data Analyses

Data sources included video recordings of the interviews and the notes taken by the interviewer during the assessments. As stated, we designed each assessment item to elicit behaviors representative of a particular level of the developmental progression (see Figure 1). Children's problem-solving behaviors were the basis for their score for each item. A score of 1 represented the correct use of a strategy and/or complete understanding of the target concept, and a score of 0 was assigned if such a strategy or understanding was partially or completely incorrect. Only children's initial responses to each item were scored. Data resulting from subsequent clinical probes on an item were analyzed qualitatively.

To prepare qualitative data for analysis, the researcher who conducted the interviews viewed the entire video of each interview, simultaneously reading the handwritten notes taken at the time of the interviews. The results of this process were transcripts of conversations between the interviewer and the children, as well as carefully written detailed notes of the children's answers and behaviors in the context of each task. The transcripts of the Hungarian interviews were translated into English by the first author. To validate the translations, the quotations used in this article were translated back into Hungarian by a Hungarian-English bilingual colleague.

To validate the assessment instrument, we analyzed the data using Rasch measurement (e.g., Bond & Fox, 2001; Hawkins, 1987; Snyder & Sheehan, 1992; Wright & Mok, 2000) and qualitative methods. The WINSTEPS Rasch modeling computer program (Linacre, 2006) was used to analyze quantitative data. The reliability of the instrument was evaluated based on the error of item and person estimates and the reliability indices. Items were evaluated by their item difficulties and their infit and outfit values. Fit statistics between 2 and −2 indicated that the assessment items each contributed to the measurement of a single latent trait, and that the participants responded in ways that are explicable by theory (i.e., the ordered levels of the developmental progression). We analyzed whether both the tasks and the children's abilities show acceptable fit with the ideal latent trait. The developmental progression can be used meaningfully only if most respondents follow a similar pathway in their development. Persons and items with standardized fit statistics larger than 2 or smaller than −2 were examined closely for reason(s) for misfit. For each misfitting item, the strategies of children who scored differently than expected by the model[3] were examined to determine what might have caused the items' divergence from the model. Therefore, fit statistics were used as a guide for qualitative analysis to reveal possible reasons for the unacceptable fit value of specific tasks and children, and to further examine the strength of the assessment items in evaluating length measurement ability.

We evaluated and elaborated the developmental progression for length measurement using both quantitative and qualitative analyses. The Rasch model provided an equal-unit-scale representation of the degree-of-difficulty ranking for the assessment items, and a degree-of-ability ranking for the children to whom the instrument was administered, thus providing empirical evidence regarding the hierarchy of the items and each participant's placement on that hierarchy. Confidence intervals were used to detect segmentation and developmental discontinuity. Nonoverlapping confidence intervals were interpreted to suggest the possible distinctness of contiguous levels of development. Because such gaps also could have resulted from characteristics of items not related to the hypothesized concepts and processes they were designed to measure, each gap was examined in detail qualitatively to determine if the gaps resulted from veridical differences in levels of thinking. Nevertheless, it remains possible that different items could have difficulty levels within the gaps.

Qualitative analyses also were used to elaborate or alter the relevant levels. Children's behaviors were described in detail for each task, to inform the descriptions of the developmental level measured by the task, providing data to evaluate and extend the hypothesized mental actions on objects that are requisite for successful solution of the task and that cognitively define the level (for descriptions of the actions at each level, see Sarama & Clements, 2009, pp. 289–292). Qualitative data were coded to differentiate between behaviors manifested by the

[3] This included children with ability estimates higher than the difficulty of the item who were not successful at solving the task, and children with ability estimates lower than the difficulty of the item who were successful at solving the task.

children when solving specific tasks. This process involved evaluating the strategies used by the children based on detectable similarities and differences, and the patterns that were found in the children's thinking and behavior generated the codes. Coding the data in such a manner resulted in a set of categories, each representing different behaviors that reflected the children's various levels of understanding of length measurement.

Informed by the Rasch-estimated abilities of the children, the researchers looked for the emergence of different strategies based on the children's abilities and established an order in which the behaviors triggered by a particular activity seemed to emerge. The different strategies that resulted based on the children's different levels of thinking were described in detail for each task, to inform how the particular developmental level measured by the task may evolve among children with different abilities. Based on an evaluation of the leap between unsuccessful and successful strategies, this allowed the identification of the specific mental actions on objects that are representative of each developmental level. In this way, we built a model of the processes that take place in young children's minds when thinking about length measurement. Further, for each inconsistency between the quantitative results and the developmental progression, the qualitative results were consulted before any changes were made to the developmental progression. Only concurrent quantitative and qualitative results suggested we alter the order of levels or collapse levels (e.g., combining contiguous levels into a single level). The result of these analyses was a modified developmental progression for the measurement of length.

Results and Discussion

Psychometrics of the Assessment Instrument

Table 1 presents the Rasch statistics for each item on the modified instrument[4] and means and standard deviations for those. Four items were excluded from the original instrument based on the results of qualitative analyses (which will be described in subsequent sections) substantiated by unacceptable infit t-values: Snakes 2 (−2.5), Ruler (3.1), Make a Road (4.2), and Missing Tower (−2.9).

Figure 2 presents the construct, or persons and items, map. The vertical axis is expressed in logits.[5] The person distribution is on the left, with the # symbol representing two children and the dot representing one. The distribution of children approximated a normal curve, with an arithmetic mean (−1.92 logits) below the arithmetic mean of the items, which is located at 0 by default. This resulted in larger

[4] Additional information on the results of the first calibration involving all initial items is reported in Szilágyi (2007). Most of the results reported in this paper are the result of a second calibration conducted after removing the four items.

[5] Difficulty and ability estimates are represented on an equal-interval scale, called the *logit* (log odds unit) *scale,* in which the relative distances between the scores are meaningful in that they express relative differences in ability and/or difficulty. Item difficulty and person ability estimates are placed on the logit scale so that there is a 50% probability that a person gets an item right with a difficulty that matches his or her ability on the scale.

Table 1
Descriptive and Rasch Statistics for Assessment Tasks

Task	Task difficulty estimate	*SE*[a]	Infit MNSQ	Infit ZSTD
Long Ribbon	7.82	1.11	0.51	−0.5
Snakes 1	5.77	0.66	0.67	−0.8
Broken Ruler	5.77	0.66	1.05	0.3
Length of Sign	4.41	0.53	1.03	0.2
Two Roads	4.15	0.51	0.52	−1.9
Spaghetti	0.94	0.36	0.84	−0.7
Different Feet	0.20	0.34	0.85	−0.8
Different Units	−0.67	0.32	0.68	−2.0
Add Strips	−0.78	0.32	0.99	0.0
Ladybugs	−1.77	0.31	0.92	−0.4
Blue Strips	−2.82	0.31	0.90	−0.5
Order Towers	−3.20	0.31	0.81	−1.1
Picture of Two Pencils	−4.95	0.31	0.92	−0.5
Length Puzzle	−7.14	0.41	1.14	0.7
Two Pencils	−7.72	0.47	1.12	0.5
Mean	0.00	0.46	0.86	−0.5
Standard deviation	4.64	0.21	0.19	0.7

[a] See Figure 3 for a visual representation of the error associated with the difficulty estimate of each task, which is represented by the radius of the circle.

estimation errors for the hardest items, with the item Long Ribbon having the largest error score of 1.11. Nevertheless, the task reliability index (0.99) suggested that the resulting order of tasks is highly replicable. The high estimation error for the children (mean = 1.14, *SD* = 0.24) implied that the ability estimates involve a certain amount of imprecision. However, the child reliability index (0.90) suggested that the estimated order of the children is reliable. Therefore, it was meaningful to proceed with further analyses regarding the levels of the developmental progression based on the empirical order of the tasks representing those levels.

The Developmental Progression for Length Measurement

Figure 3 shows the empirical order of the interview tasks. The circle around each task represents its confidence interval. Where the confidence intervals around the tasks overlap vertically, development between the levels represented by the particular items cannot be considered distinct. The lack of segmentation may mean

```
<more>|<rare>
8 # +
| Long Ribbon
. |
7 +
|
|
6 # +
| Broken Ruler Snakes 1
|
5 .# T+
|S
| Length of Sign
4 ### + Two Roads
|
|
3 +
|
### |
2 +
S|
|
1 ##### + Spaghetti
|
| Different Feet
0 #### +M
|
.### | Add Strips Different Units
-1 +
|
###### | Ladybugs
-2 M+
.### |
| Blue Strips
-3 +
.##### | Order Towers
|
-4 +
.######## |
|S
-5 + Picture of 2 Pencils
S|
|
-6 .######### +
|
|
-7 + Length Puzzle
|
.## | Two Pencils
-8 .## +
<less>|<frequ>
EACH '#' IS 2.
```

Figure 2. Person–item map for the measurement assessment.

that the items do not accurately represent the particular level of thinking, that the difficulty of the item was affected, that the theory was flawed, or some combination of these (Wilson, 1990).

Where the confidence intervals around the tasks on Figure 3 (on p. 600) do not overlap, development between the levels represented by the particular items may be considered distinct. This, therefore, provides initial evidence for the existence of developmental levels, which was further examined using qualitative analyses. The four horizontal lines in Figure 3 indicate possible points of segmentation in the developmental progression.

The hypothesized order of the levels of the developmental progression and the tasks, and the empirical order of those based on the difficulty estimates resulting from Rasch analysis (second calibration after removing the four items) are shown in Table 2.

Table 2 also presents the levels of the modified developmental progression proposed based on the findings of this study. The levels from the developmental progression are listed alongside the tasks designed to measure each level. To the right of these are the same columns for the modified developmental progression inferred from the analyses. The difficulty estimates, expressed in logits, are included in parentheses after each task.

Table 2 indicates that our analyses support the developmental progression in general; however, the orderings of some items (and therefore possibly the corresponding levels) are inconsistent with the predictions. The following sections describe these mismatches and the children's strategies, to undergird a revised developmental progression of levels and to expatiate on the mental objects and actions on them available to the children for each level of the developmental progression.

Length Quantity Recognizer

Children at the *Length Quantity Recognizer* level (see Figure 1) acknowledge length relationships between pairs of objects that are already aligned (parallel along their lengths with endpoints on a line perpendicular to the lengths), but knowledge regarding the need for such alignment in establishing length relationships is absent or unstable.

The confidence intervals provided by the Rasch model for items Two Pencils (see Figure 1) and Length Puzzle (see Figure 1) did not indicate a difficulty gap between the two items, and the empirical order of the items was the opposite of the order expected by theory. Therefore, the Rasch order did not substantiate the distinctness of the two levels—Length Quantity Recognizer and Length Direct Comparer—represented by the items (see Figure 1). However, qualitative analysis of those items placed the items' validity into question, so the existence and sequencing of the levels was neither accepted nor rejected. Given findings that suggested two developmentally distinctive levels of thinking, length as a property in isolation and length as a comparative quantity, the two levels were tentatively left in the developmental progression, but further research is needed regarding these two levels.

Qualitative data suggested that Length Puzzle did not provide a meaningful representation of the children's actual abilities, as hypothesized. To solve Length Puzzle, when asked if the foam strip fit well in the opening (see Figure 1 for more

Table 2
Summary of Findings for the Order of Tasks and Levels of Development

Hypothesized levels	Tasks	Modified levels	Tasks (difficulty in terms of logits)
Length Quantity Recognizer	Length Puzzle	Length Quantity Recognizer	Length Puzzle (−7.14)
Length Direct Comparer	Two Pencils Snakes 1 Snakes 2	Length Direct Comparer	Two Pencils (−7.72)
Indirect Length Comparer	Picture of Two Pencils Two Roads Make a Road	Indirect Length Comparer—Nascent	Picture of Two Pencils (−4.95)
Serial Orderer to 6+	Order Towers Missing Tower	Serial Orderer to 6+	Order Towers (−3.20)
End-to-end Length Measurer	Blue Strips Ladybugs	End-to-end Length Measurer	Blue Strips (−2.82) Ladybugs (−1.77)
Length Unit Iterater	Spaghetti Ruler	Length Unit Relater and Iterater	Add Strips (−.78) Different Units (−.67) Different Feet (.20)
Length Unit Relater	Different Units Different Feet		Spaghetti (.94)
		Path Measurer	Two Roads (4.15) Length of Sign (4.41) Broken Ruler (5.77) Snakes 1 (5.77)
Length Measurer	Broken Ruler Long Ribbon		Long Ribbon (7.82)
Conceptual Ruler Measurer	Add Strips	Conceptual Ruler Measurer	

details on the task), 7 children struggled with pushing the strip into the opening. Because they were distracted by the difficulty of placing the strip in the opening, they did not take the length of the strip into consideration, but they commented on the fact that eventually the strip fit well. Therefore, this item probably misdiagnosed the ability of these 7 children, establishing an incorrect empirical order of difficulty. We suggest that the Length Quantity Recognizer level might require the development of a different kind of item to adequately measure it.

The qualitative findings suggested that length as an attribute is understood by the children in two developmentally distinctive ways: length as a property of an object in isolation from other objects, and length as a quantity that allows comparisons between objects. Children initially view length as a property that objects possess due to their shape (e.g., "long," or extending along one dimension substantially more than the other two). When presented with a box with a circular base and a piece of a chenille stick that was shorter than the diameter of the box,[6] 47 children said that the chenille stick was long, when it was clearly shorter than the box. Many children correctly indicated that the box was long, and some of them also remarked that "the worm is long . . . but it's shorter." The children who did not establish a length relationship between the two objects either did not understand length as an abstract comparative attribute, or they simply did not use that knowledge to compare objects. The round shape of the box might have favored the "length-as-a-shape" conceptualization.

Children who begin to understand length as a comparative property are able to detect differences in the length of two objects when they are aligned, without necessarily appreciating the need for such alignment, and often without being able to verbally explain their thinking. When asked which one of two pencils not in alignment was longer (for more about item Two Pencils, see Figure 1), some children with ability estimates lower than or close to the difficulty of the task chose both of the pencils to be longer because they both looked long. Others chose one of the pencils to be longer, but without physically moving the pencils or being able to explain their choice.[7] Even when these children were prompted to show it, they insisted that they could "just see it." Therefore, these children evaluated length based on gross visual clues alone. Their concept of length was nonrelational in the sense that they had not yet developed the need to establish one-to-one correspondence between pairs of points—one from each object—that are equidistant from aligned ends of the objects to see which one of them has leftover length.

Research suggests that even young babies have the ability to (intuitively) discriminate length as a continuous extent. Seven-month-old babies are sensitive to contour length and they use that information to differentiate between small sets of objects (Clearfield & Mix, 1999; Feigenson, Carey, & Spelke, 2002). At the age of 16 months, they use distance information in simple problem-solving situations

[6] Note that this task was not included in the quantitative analysis, but it was included in qualitative analyses to enrich the data about children's thinking.

[7] It is important to note that the difference in length between the two pencils was not visibly detectable based on the way they were set up (see Figure 1 for detail), which should have encouraged a need to check formally.

to estimate the location of an object hidden in a sandbox (Huttenlocher, Newcombe, & Sandberg, 1994). However, the findings of this study suggest that these early abilities to detect length are converted into explicit understanding of the attribute and then comparative understandings only over periods of years. That is, before achieving these later developmental levels, children do not use length information readily in situations involving length comparison. Even so, the fact that length information is discriminated by the children makes a variety of comparative length experiences meaningful for them. During the explicating and verbalizing process, which involves becoming explicitly aware of the mathematical concepts and processes and connecting this knowledge with language, children develop a limited understanding of the meaning of length-related words such as “long.” The reason for that might be experiences emphasizing the use of the word “long” to describe straight objects. Focus on experiences involving comparison between different objects based on length might promote the emergence of an increasingly sophisticated connection between children’s existing knowledge and language.

Length Direct Comparer

Children at the *Length Direct Comparer* level (see Figure 1) know that relative length decisions need to be made based on direct alignment between the objects being compared. Although this knowledge may not be readily available in every context, these children know to put two objects in direct alignment to make the difference in length visible.

Three tasks were included in the assessment instrument to detect children’s skills related to this developmental level: Two Pencils, Snakes 1, and Snakes 2. Although the unacceptably low infit t-value of −2.5 for item Snakes 2 indicates that responses to this item are overdetermined in measuring the underlying construct (i.e., the task is “too good”), qualitative findings showed that some students counted the segments of the snakes to compare length, conceptualizing this task to be a number comparison task. Therefore, the item was excluded from the instrument. For the item Snakes 1, although it was intended to measure this level of thinking, both quantitative and qualitative data suggested that it assessed thinking at a level of ability further along the developmental progression. Therefore, it will be discussed in the section on the *Path Measurer* level.

Qualitative data suggested that at early levels, children do not believe that they need to operate physically on objects in making a comparative length decision, even though adults judge that the difference in length cannot be justified based on visual clues. Children initially are content to rely on those cues to prove a visually based decision. Only later do they believe that a direct comparison strategy is needed to make reliable decisions. When asked to solve the item Two Pencils (see Figure 1), 82 of the 110 children who successfully solved this task chose one of the pencils to be longer without moving the pencils. Some of these children were aware of the strategy of direct alignment, and they used it upon being asked to prove their answer. However, they did not yet seem to develop a need to put two objects in alignment before making a decision about their length.

Indirect Length Comparer

Children at the *Indirect Length Comparer* level (see Figure 1) can compare the length of two objects by representing them with a third object. Although children can learn to use the behavioral technique of indirect comparison without an understanding of the principle of transitivity (Boulton-Lewis, 1987; Hiebert, 1981), the present study included behaviors signifying such understanding as a criterion for thinking at this level.

The assessment instrument included three tasks related to this developmental level: Picture of Two Pencils (see Figure 1), Two Roads, and Make a Road. However, the unacceptably high infit t value of 4.2 for the item Make a Road, informed by qualitative data,[8] resulted in its exclusion from the instrument. Regarding the item Two Roads, although it was intended to measure this level of thinking, both quantitative and qualitative data suggested that it required thinking at a level of ability further along the developmental progression. Therefore, it will be discussed under the Path Measurer level.

The confidence intervals provided by the Rasch model for the items Two Pencils and Picture of Two Pencils, represented in Figure 3, indicated a difficulty gap between the two items. Consequently, the development between levels Length Direct Comparer and Indirect Length Comparer was considered distinct. It was concluded that children initially develop the ability to compare objects directly and only later do they acquire the skills and knowledge necessary for indirect comparison.

To solve indirect comparison situations (involving two nonmovable objects) using a third object and transitive reasoning, children need to understand that the length of one object can be represented by the length of another and how to create such a representation. The representative role of the third item (the foam strip) in measurement may become meaningful as length becomes a mental object, replacing the notion of being physically connected to the object. Qualitative data suggested that children develop through several phases as they learn to abstract length as a mental object. At first, children are not able to think of length as an abstract concept that can be mentally extracted from an object and further manipulated, either physically through representation with another object or in thought. When asked to use a foam strip to compare two unmovable lengths on the item Picture of Two Pencils (see Figure 1), 27 children were not able to infer a comparative length relationship between the two pencils based on how they each related to the length of the strip, even though they correctly aligned the strip with each pencil and concluded that the strip was longer than both pencils.

Later, children appear to develop an implicit, or *theorem in action* (Vergnaud, 1982),

[8] Seventeen of the 89 children whose abilities were lower than the difficulty of this item were able to solve this task correctly. Qualitative data suggested that these children misunderstood the question, and instead of making the pipe cleaner the same length as the zigzag road on the picture they made the pipe cleaner the same shape, which eventually resulted in correct measuring. These same children indicated that the length of the zigzag road was the distance between the two endpoints, not the length of the path.

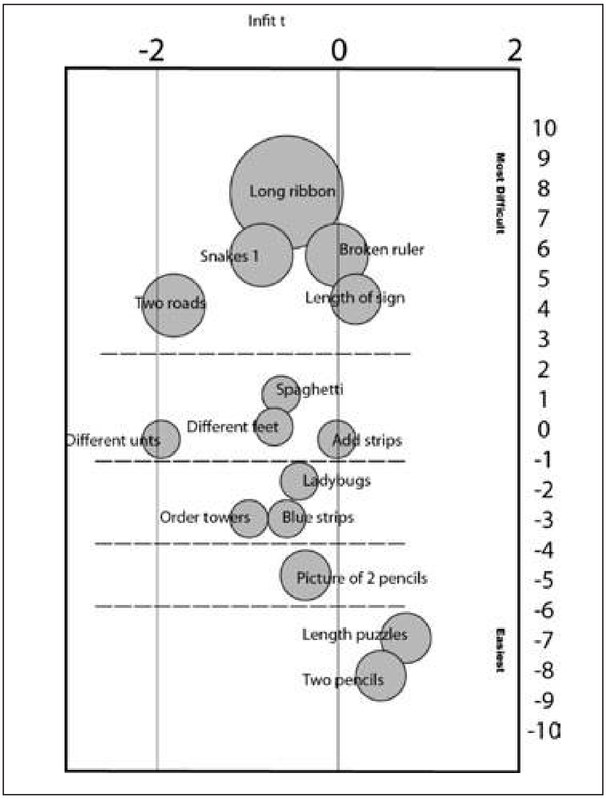


Figure 3. Segmentation in the developmental progression.

understanding of transitivity. Children with ability estimates close to the difficulty estimate of the item Picture of Two Pencils were able to use the strip well to compare the length of the two pencils. However, they were not able to explain their strategy. Data suggest that these children did not lack the logical operation involved in transitivity, but rather they lacked the mental model for length as an abstraction on which to mentally carry out operations involving comparison and transitivity. One child, for example, traced both pencils with his finger as he laid the strip along them, and he correctly concluded which of the two pencils was longer. This may have indicated that he was not ready to consider the strip as a representation of the length of the pencil, and the movement of his finger helped him to make the process physically concrete. The length of an object did not seem to be meaningful to this child without a physical (and in this child's case, possibly motion-based) representation.

Qualitative data suggested that children misused the third object, the foam strip, in many ways on the item Picture of Two Pencils. For example, they would cover each pencil with the third object without ascertaining where the endpoints of the pencils aligned with it. Some would therefore incorrectly conclude that the pencils were the same length, because both could be covered completely by the paper.

Although these children showed some signs of transitive reasoning, they did not manage the materials in this task, perhaps indicating an unstable or incomplete understanding of establishing length representations. These children appeared to need a more robust abstraction of length as a mental object, along with competence in establishing accurate length representations and in transitive reasoning, to achieve generalizably accurate and meaningful indirect comparison strategies. The ability to mentally abstract length may not constitute conservation in the full Piagetian sense, but a nonverbal conservation, similar to Vergnaud's (1982) *theorem in action*. Although they may not be explicitly aware of the conservation concept, these children operate with the implicit assumption that the length of an object is maintained when it is moved. Similarly, these children demonstrate behaviors consistent with transitive thinking. However, their knowledge of transitivity may not be conscious, given that they cannot correctly explain their transitive reasoning. This is in accordance with Boulton-Lewis' (1987) differentiation between ability to use and to explain transitive thinking.

These results therefore lead us to alter the definition and description of the Indirect Length Comparer level. We define it as initial, implicit competence in the use of a third object to compare the length of two other objects, with considerable development in length representation and explicit transitive reasoning occurring in conjunction with the development of subsequent levels of thinking.

For children at this redefined Indirect Length Comparer level, a number assigned to the length of an object as a representative makes sense. However, unlike children at the Length Unit Relater and Repeater level, they cannot operate on that number. When asked to measure, a child may assign a length by guessing or moving along a length while counting (without equal length units). The child may also move a finger along a line segment, saying "10, 20, 30, 31, 32." These "trace-and-count" and "point-and-count" strategies help the children maintain the physical connection between the length of the object and the number that expresses it. It is noteworthy that children in this study appeared to need physical objects *to assign numbers* (not to use indirect length comparisons, which may have components of mental imagery, as stated previously) even if inaccurately, reflecting the "sensory-concrete" level of thinking that requires the use of—or at least direct reference to—sensory material to make sense of a concept or procedure (Clements & Sarama, 2009).

Serial Orderer to 6+

The *Serial Orderer to 6+* level of thinking (see Figure 1) requires that children understand the complete order of objects as one mental object, for which every object in the series is longer than each of the preceding ones (Piaget & Inhelder, 1967). This understanding seems to rely on the successive application of transitivity.

Two tasks were included in the assessment instrument to detect children's skills related to this developmental level: Order Towers (see Figure 1) and Missing Tower. However, the unacceptably low infit *t*-value of −2.9 for the item Missing Tower, informed by qualitative analysis,[9] resulted in its exclusion from the instrument.

The development between the levels Indirect Length Comparer and Serial Orderer to 6+ was considered distinct based on the confidence intervals provided by the Rasch model for the items Picture of Two Pencils and Order Towers, represented in Figure 3. It was concluded that children develop the indirect comparison strategy before they can order six or more objects based on their length. Although six children were able to seriate six objects before they could compare two objects indirectly, that might have been a result of specialized experience (or measurement error). The findings of Johnson's (1974) study suggest that children can be trained to seriate six sticks based on length before they are able to make transitive inferences about the length of those sticks. Our findings suggest that transitive thinking may be required for meaningfully seriating objects based on length, at least with explicit understanding of the process. That is, the difference in these findings may have been a result of the fact that in the current study emphasis was placed on children's understanding of their actions.

Qualitative data from this study suggest that, at first, children's ability is limited to establishing one or two length relations at a time. When asked to put six premade connecting cube towers in order from shortest to longest (i.e., Order Towers; see Figure 1), many children were observed to make ordered groups of two or three towers. When asked to put all the towers in order, they were not able to coordinate the groups to establish an overall order. The ability to see two neighboring relations of length at a time allows only for putting three objects in order.

Children who were able to correctly place the six towers in order saw the series as a whole and meaningfully coordinated more than two length relations, using different strategies. Some children carefully compared each tower pairwise to find the shortest tower, and they repeated that strategy until they found the next tower in line. Others used the number clue embedded in the towers, and counted the number of cubes in each to decide the order.

End-to-end Length Measurer

At the *End-to-end Length Measurer* level (see Figure 1), children's ability to form a mental image of length as distance covered (or "scanned") constrains their placement of units to eliminate gaps, thus making the covering continuous. This implies that they have represented and maintained the length of the object, although they have not necessarily applied the same mental operations to the parts or units imposed on that length. This intuitive understanding of the additivity of length, reflected by children's ability to count the number of units into which a length is subdivided, does not necessarily allow for taking the size of those units into consideration.

The two items designed to measure this level of thinking were Blue Strips (see Figure 1) and Ladybugs (see Footnote 14 on p. 607). Although the Rasch model

[9] Qualitative findings indicated that this item may be more related to the construct of coordinating multiple relations of length at a time than to that of length measurement (see Szilágyi, 2007, for more detail).

did not indicate segmentation between these two items, the confidence intervals for the items Order Towers and Blue Strips indicated an overlap between the two items (see Figure 3). Therefore, the Rasch model did not indicate the distinctness of the levels Serial Orderer to 6+ and End-to-end Length Measurer. Although it may be that items Order Towers and Blue Strips represent the same level of development, it may also be that the two abilities represented by the two tasks develop parallel to each other instead of consecutively, hence the overlap between the confidence intervals (thus, future studies may need to use more complex models than the Rasch model). Similarly, one or more of the items may not accurately represent the abilities related to the levels, or their difficulty may have been influenced (Wilson, 1990). For instance, it is possible that the item Blue Strips' complex nature may make it a less accurate assessment of the End-to-end Length Measurer level of thinking. When asked which one of the three blue strips was the same length as four of the yellow unit strips (see Figure 1 for Blue Strips), one child, after putting two yellow strips along the 2-in. blue strip, agreed that it was about the same length as 4 yellow strips. However, when at the end of the interview the researcher put two yellow strips along the 2-in. blue strip and asked her how long it was, she was able to conclude that it was 2 yellow strips long. When the researcher asked if it was 4 yellow strips long, she disagreed, and she put 4 yellow strips along the 4-in. blue strip, claiming that was the one that was 4 yellow strips long.

Based on our findings, we did not find sufficient evidence to modify the progression. Therefore we propose that the two levels of development, Serial Orderer to 6+ and End-to-end Length Measurer, exist, but we must leave it to future research to ascertain the nature and relationship of these levels. We also suggest that additional tasks be designed to provide more data regarding the development of measurement using multiple copies of a unit.

Length Unit Relater and Repeater

This level of thinking involves the development of three different subdomains of knowledge: the additivity of length, the relationship between unit size and number, and unit iteration. Contrary to our initial hypothesis, Rasch analysis suggested that the items Add Strips (see Footnote 12 on p. 605), Different Units (see Footnote 13 on p. 605), Different Feet (see Figure 1), and Spaghetti (see Figure 1) represented a single level of thinking. Therefore, two hypothesized levels, Length Unit Repeater and Length Unit Relater, were combined into one level: Length Unit Relater and Repeater. Although this study supported the parallel development of the three abilities representative of this level of thinking, further, deeper analyses are needed to explore the true nature and pattern of development regarding three posited sublevels:[10] Additive Length Composer, Length Unit Relater, and Length Unit Iterater.

[10] The term *sublevel* is used because qualitative data suggested developmental ordering.

Additive Length Composer. Children who possess this competence can express the length of an object by a number, and they can operate on that number. Therefore, they are able to carry out a simple addition of two measures to obtain the measure of their combined length. These children may see the need for equal and universal units, although they may not be able to apply their knowledge in all situations.

Item Add Strips is representative of the newly proposed sublevel *Additive Length Composer*. The confidence intervals provided by the Rasch model for the items Ladybugs and Add Strips indicate segmentation between the two items. Based on this finding, we concluded that children learn to add up the number of units used in an end-to-end fashion to measure length before they take into consideration the length of the units being combined. That, however, does not imply that they also understand the inverse unit-size/number-of-units relationship.

Based on qualitative data, children at first develop an intuitive sense of additivity. For the item Add Strips,[11] to determine the length of the combined strip, some children who were able to use multiple copies of a unit to measure the length of an object (so they demonstrated understanding of additivity in a context that involves unit parts) used a "point-and-count" or a "trace-and-count" strategy. These children knew that length can be expressed by counting parts, but they were not able to extend that understanding to a complex situation that involved unequal parts. Children's understanding of the additivity of length becomes more sophisticated when they add the measure of the two parts to compose the measure of the connected strip. These children, however, were not able to subdivide a length according to the number of units expressed by the measure. Therefore, for them, the additivity of length was not reversible.

Length Unit Relater. At this sublevel, children's conceptions of length units are conserved (Piaget et al., 1960) both in recognizing that the use of different-size units yields different measures, and in the ability and preference to maintain the length, or extent, or mental images of units. Thus, length conservation is generalized to the whole and the parts, although this may be only intuitive, and may be constrained to situations in a single dimension. Children no longer need the physical connection between the object and its length ensured by the "point-and-count" strategy, therefore working toward "integrated-concrete" knowledge (interconnected structures that are "concrete" at a higher level because they are connected to other knowledge, both physical knowledge that has been abstracted and thus distanced from concrete objects and abstract knowledge of a variety of types; see Clements & Sarama, 2009).

The items Different Units and Different Feet were designed to measure a child's ability to acknowledge and explain the result of measuring an object with two different-sized units. The confidence intervals provided by the Rasch model for items Add Strips and Different Units were overlapping. Therefore, it was proposed

[11] In this task, children were asked how long the combined strip will be when a 7-inch-long and a 4-inch-long strip are taped together end to end.

that the two abilities represented by the two items, the understanding of the additivity of length and the recognition of the unit-size/number-of-units relationship, develop parallel to each other. Some children learn to compose lengths before they relate unit length to unit size, whereas other children comprehend the impact of unit size on the number of units needed first. For both abilities, however, the child has to work at least at the End-to-end Length Measurer level, which may be a prerequisite for these two abilities. It is proposed that the existence of both these sublevels is the result of the children's ability to conserve length. These two abilities later pave the way to the development of a need for using a standard unit when measuring, and the recognition that comparison to that standard unit lies at the heart of measurement.

Understanding of the unit-size/number-of-units relationship appears to rely on an ability to coordinate multiple pairwise relationships. When asked to predict if fewer, the same number, or more of one unit than another would be needed to measure a long strip to solve the item Different Units,[12] many children made a decision based on the relationship between only two strips; they failed to take into consideration the relationship between all three strips.

Qualitative data revealed that children are aware of the fact that different units result in different measurements before they are able to generalize the correct relationship between the size of a unit and the number of the units needed to measure (limited here to an understanding that it is inverse in direction, not that it is a specific multiplicative inverse). When asked why two friends might have gotten different results when they measured the same rug to solve item Different Feet (see Figure 1), 82 children were able to point out that their feet were different sizes, but 47 of them said that measuring with longer feet resulted in a larger number. Although these children were not able to predict correctly that longer feet would result in a smaller count, some of them were able to provide a correct answer to the item Different Units. The reason for that may have been that modeling a problem situation is easier than mentally representing or generalizing it. One child said, "It would be more blue strips . . . let me try!" He then used iteration to correctly measure the green strip with the blue and the yellow, and he explained, "The yellow is more 'cause the blue is longer . . . the blue is big so the blue will take more room and the yellow will take less room."

The ability to recognize and generalize the idea that unit size has an impact on the number of units needed does not automatically lead to the applicability of the idea. Qualitative analysis for item Ladybugs[13] revealed that when comparing the

[12] For this task, children were told that the measure of a green strip was 14 yellow strips long. They were shown a blue strip that was double the size of the yellow strip, and they were asked to infer whether the same number, fewer, or more blue strips would be needed to measure the green strip. They were allowed to manipulate the stips.

[13] For that task, children were provided with a sheet of paper that had two straight chenille sticks (9 and 10 inches long) glued on it; the sticks were neither aligned nor parallel. Children were then asked to determine which chenille stick ("worm") was shorter using multiple copies of 1.5-inch and 1-inch discrete units.

length of two objects, some children with demonstrated knowledge of the unit-size/number-of-units relationship mixed units between objects, but not within objects. Other research similarly showed that children with demonstrated understanding of the unit-size/number-of-units relationship are not able to apply their knowledge when comparing the length of two objects (Carpenter & Lewis, 1976).

According to the current study, none of the children who mixed units within objects were aware of the unit-size/number-of-units relationship. Based on this, it can be concluded that comprehension of the effect of using different-size units is needed for true understanding of the need for equal units when measuring. Children may know to put purely equal units along an object when measuring its length. However, they understand the need for equal units only when they understand the impact of using different-size units. This is in accordance with Petitto's (1990) findings regarding the development of the need for equal-length units, which she reported to be based on the development of conservation of length. According to Hiebert (1981), the inverse relationship between the size of a unit and the number of units needed also depends on the ability to conserve. Similarly, the need for universal units also may develop as a result of a clear insight into the unit-size/number-of-units relation, and learning to apply knowledge of that relationship in various contexts.

Length Unit Repeater. Children at this sublevel are able to iterate a single copy of a unit along an object to measure its length. They are able to physically (marking the end of the unit every time it is moved) and/or mentally subdivide an object, to conceptualize it as made up of multiple copies of the unit. They count the number of units that can fit along the length of an object and assign that number to express the length of the object. At this level of thinking, children start laying the ground for developing the Conceptual Ruler Measurer level of thinking.

Two items were originally designed to measure this level of development: Spaghetti and Ruler. The item Ruler was eliminated from the assessment instrument because of its unacceptably high infit *t*-value of 3.1. Because the confidence intervals provided by the Rasch model for the items Different Feet and Spaghetti were overlapping, segmentation between the two items was not indicated. However, quantitative and qualitative data suggested that most children develop at least some understanding of the unit-size/number-of-units relationship before they are able to iterate a unit. Only 2 children were able to iterate without being able to solve either of the items Different Feet or Different Units. All the other children who were able to iterate were able to solve one or both problems that measure the sublevel Length Unit Relater. Moreover, 12 children who were not able to iterate solved the items Different Feet and Different Units. Therefore, it was concluded that at least some insight regarding the unit-size/number-of-units relation can develop before unit iteration does. In addition, that understanding seems to be contextually generalizable before a demonstrated ability to iterate, as shown by the results for the two different contexts used in this study. A clear understanding of the unit-size/number-of-units relationship may actually foster the development of unit iteration through promoting the understanding of the requirement of equal units.

According to Hiebert (1981), children can learn to use the technique of iteration correctly without understanding the Piagetian logical operations of conservation and subdivision embedded in iteration-related tasks. However, this study shows that when learned in a rote manner, the context or the relative size of the unit may have an influence on children's use of the unit-iteration strategy. When the children's answers on the items Spaghetti and Long Ribbon[14] were compared, it was found that some children did not demonstrate any knowledge regarding unit iteration for the item Spaghetti, but they iterated a longer unit, a ruler, perfectly well for the item Long Ribbon. That the ruler was more conducive to unit iteration than regular objects may be due to children's preference for using the ruler in measurement (Boulton-Lewis et al., 1996).

Some children's use of the unit iteration strategy suggests that a true understanding of unit iteration relies on coordinating the strategy with the understanding of the need for equal units and the understanding of additive length composition. Two patterns of strategies were relevant. One strategy, used by 17 children, involved iterating by marking the endpoint of the unit with a finger, but without considering the width of the finger. These children focused on the need for equal units, but they failed to take all the parts into consideration. The second strategy involved "eye-iteration" to correctly measure the length of the spaghetti. These 5 children all iterated well, but without physically marking the endpoints of the unit. They all explained that they "remembered with their eyes" where the unit ended. Three children used their fingers to mark off imaginary units that they counted to get the measure. These children saw the whole as the sum of its parts. However, they did not have a reliable method to ensure that the subdivisions were equal. Some of these children might not have known (or might not have been able to apply the idea) that measurement involves exact comparison to the unit.

Path Measurer

This level of development is related to children's ability to compose and decompose lengths. Children at this level of thinking understand the mutual relationship between lengths and parts of lengths. Their concept of the additivity of length is reversible; they understand that the whole length is (or can be) divided or decomposed into sections, and that when the lengths of the sections are added, the whole is composed. Their understanding of the conservation of length with respect to parts put together to form a whole is not only conscious (they are able to verbalize that knowledge), but they also use that knowledge in comparative length decisions. All these skills gradually become complete, generalizable, and interrelated.

Although the Rasch model established a clear order of the items Two Roads (see Footnote 16 on page 608), Length of Sign (see Figure 1), Broken Ruler,

[14] To solve the item Long Ribbon, children were asked to measure the length of a 43-inch ribbon using a 1-foot ruler.

Snakes 1,[15] and Long Ribbon (see Footnote 14 on p. 607), it did not verify that uniquely different levels of thinking differentiated any of these items. However, due to the fact that none of the children solved Long Ribbon before Snakes 1, supplemented by qualitative data regarding behaviors reflecting children's mental actions on objects when solving these problems, we suggested that the item Long Ribbon was representative of a sublevel of thinking that is qualitatively different from the sublevel represented by the other items related to the Path Measurer level. The difficulty gap between the items Spaghetti and Two Roads was the greatest along the whole developmental progression. Therefore, children in this study developed the ability to iterate before they demonstrated abilities related to the Path Measurer level. However, because the lack of participants with abilities at the higher end of the difficulty-ability continuum resulted in unusually large confidence intervals, thus increasing the chance of overlaps, further research is needed to explore the nature and relationship of the abilities represented by the Path Measurer level of thinking and its two sublevels.

Path Measurer. Children are able to verbalize their understanding of the conservation of length with respect to parts put together to form a whole. For example for the item Two Roads[16] they knew that the zigzag road can be longer than the straight road, even if the distance between its endpoints was shorter. They were able to use a third object, a chenille stick, to correctly represent the length of the zigzag road, and they showed that it was longer than the straight road by straightening it out and directly comparing it to the straight road. These children clearly understood that the length of the zigzag road is not the distance between its endpoints, but the sum of the lengths of its parts. They were able to mentally abstract the length of a path, through mentally straightening it out, maintaining (at least an approximation of) its length. Therefore, the mental representation of length that these children possessed appears to be flexible, involving the whole as the sum of its subdivisions, versus length as a rigid distance. Their understanding of the conservation of length seemed to apply not only to length as an indivisible but additive whole (that kind of conservation was needed for the Length Unit Relater and Repeater level, along with a true understanding of the technique of indirect length comparison) but also to length as an entity subdivided into parts. These children were aware that length does not change as a result of rearranging the position of parts.

Some children at lower levels of thinking were constrained by the component of movement along the paths. Many children realized that the length of the zigzag road was longer, even though it "didn't stretch as far as the straight road." Many of the children actually walked the cow down both paths to prove the straight road

---

[15] This task presented the children with two "snakes" made of craft sticks that were fastened together end to end so the pieces could be turned. The "snakes" were folded in a zigzag shape so that one looked longer and had more turns than the other, but because its sections were shorter, it was the shorter snake.

[16] In this task, children were shown the picture of a pond and two roads, one straight and the other zigzag, that both lead to the pond. They were asked to compare the length of the two roads to help a cow take the shorter road to the pond.

was shorter, because it was "quicker." When the ingredient of movement was no longer involved, these children's intuition regarding length was undermined.

Most children's concept of the additivity of length was reversible at this sublevel of thinking. They understood that the whole length is divided or decomposed into sections and that the sections are added up to compose the whole. When asked how wide a sign is if the length of the building and the lengths on the two sides of the sign are known, to solve item Length of Sign (see Figure 1), these children were able to calculate the length of the sign based on the numerical information that was provided, using a combination of the operations addition and subtraction. We propose that their extended understanding of the conservation of length and the concomitant logical operations allowed them to reverse the operation of additivity.

Many children at this sublevel of thinking can measure length accurately with a ruler, understanding the process and the properties on which it is based. They see the ruler as composed of equal-sized units, understanding that the numbers on the ruler represent the number of those units. Therefore, they also comprehend the importance and meaning of the zero point on the ruler. Children at this level of thinking used two correct strategies for the item Broken Ruler (see Figure 1). Some of these children counted how many units on the ruler covered the full length of the strip, whereas others checked the number on the ruler at both ends of the strip, and they simply subtracted the smaller number from the larger one. The item Broken Ruler involves units common to both the object being measured and the object used to measure. Children who solved the item Broken Ruler were able to see the blue strip as made up of the same units as the ruler.

A transitional strategy used by some children for the item Broken Ruler provided insight into the mental objects and actions available to children at this level of thinking. These children put the ruler to the strip so, instead of their ends aligned, the ruler was shifted to the right. They explained that they imagined the 0 and the 1, the 0 being at the end of the strip and the 1 between the imagined 0 and the visible number 2 on the ruler. Because the imaginary units were not the same size as the units represented on the ruler, this strategy did not result in a correct response. These children's use of the ruler was still under development in the sense that they did not picture the ruler as an object composed of equal units, but their focus was on accounting for all the numbers on the ruler.

Length Measurer. Although the large overlapping confidence intervals (see Figure 3) did not indicate the existence of a distinct level of thinking, the nature of mental objects and actions on them available to the children substantiated the separate discussion of this sublevel of thought. Qualitative findings indicated that all the skills that were built at the Path Measurer sublevel become complete, generalizable, and interrelated at the Length Measurer sublevel.

Children's mental representations of length at this sublevel become flexible and integrated, especially with respect to the part–whole relationship. Their mental representations of length not only include lengths composed of units and units of units, but also relationships between those entities. Therefore, these children are able to relate these representations and flexibly move among them based on the needs of measurement. For example, they understand that the measure of an object of length 3 meters is also 300 centimeters, but also 2 meters and 100 centimeters, and so forth. When asked to measure the length of a 43-inch ribbon using a 1-foot ruler, to solve the item Long Ribbon, these children were able to express length as a combination of units and parts of those units; for example, 3 feet and 7 inches.

Before attaining the Length Measurer sublevel of thinking, children may understand the relationship between small units and larger units composed of them, but their understanding is not yet flexible. Research shows that even first graders can use the technique of iterating units of units with understanding (McClain et al., 1999; Stephan et al., 2001). Although those children may be able to iterate units composed of smaller units, and therefore consider a group of units as a new unit, they cannot move between the representations flexibly.

Conceptual Ruler Measurer

Qualitative analyses revealed that the two items that were designed to measure children's abilities related to this level, Add Strips and Length of Sign (see Figure 1), measured competencies lower in the developmental progression than the ones thought to be needed for this level. Because none of the tasks included in the assessment instrument proved to be useful for measuring this level of thinking, new items need to be developed that are targeted at this level. We can therefore say nothing about the existence or usefulness of this level from our results.

Summary: Modifications to the Original Learning Trajectory

Figure 4 provides descriptions of the modifications to the levels of the original learning trajectory's developmental progression and to the mental actions on objects that define each level. The tasks that corresponded to each level are also listed. These descriptions are addendums to those in Figure 1, not replacements.

Implications

The findings of this study suggest that children with a wide array of abilities and from two different countries follow a similar progression in their development of concepts and strategies related to length measurement, at least within the limits of the model and assessment items used here. Although there was some variation in the actual progressions followed by the children and our sample was a convenience sample, knowledge of a shared progression can be a useful tool in the hands of teachers, curriculum developers, teacher educators, and policy makers. For example, they may help synchronize the production of standards, textbooks, and assessments for length measurement in the early years.

We emphasize that the developmental progression and the related tasks described in this study represent one possible route based on patterns of development shared by the majority of the participants. Due to the nature of learning trajectories, there is a possible factor of task effect, which has to do with the transferability of

Levels of the developmental progression	**Actions on objects**	**Tasks**
Length Quantity Recognizer Initially identifies the longer object when the objects are in alignment, but does not know to align objects to determine which is longer. May indicate that a straight object (e.g., chenille stick) is longer than a round object, even when the round object is actually longer.	Initial schemes relate length vocabulary terms to the aspect ratio of the shape of objects, such that, for example, "long" is applied to objects that have a large aspect ratio. This tendency can coexist, and initially override, nascent length comparison operators. As the action schemes (described in Figure 1) develop, they take precedence over these initial schemes.	Length Puzzle
Length Direct Comparer Initially may not physically move two objects in alignment to determine which one of them is longer, but does so to provide evidence. "If you turn it (—) that way (|) it'll be longer." May not be able to infer a comparative length relationship between the pictures of two objects based on how they each relate to the length of a third.	Gross visual comparison is preferred for a considerable period, with direct comparison (physical superposition) used only as a check, and often only under external pressure for such verification. With tasks that defy such visual comparison or demand precise comparisons, or with developing metacognition allowing for recognition of the limits of visually based operations such as mental rotation, the action scheme of physically moving and aligning objects gradually achieves ascendence.	Two Pencils

Figure 4. Modifications to the learning trajectory.

Figure 4 (*continued*)

Levels of the developmental progression	**Actions on objects**	**Tasks**
Indirect Length Comparer–Nascent (previously Indirect Length Comparer)	An abstract concept of length, one that can be extracted from an object and mentally represented (versus a need for being physically connected), must first be created. In addition, the explicit use of transitive inference (whether or not it can be verbalized) develops slowly, often overlapping with progress to the subsequent levels.	Picture of Two Pencils
Serial Orderer to 6+		**Order Towers**
End-to-end Length Measurer May lay different-size units end to end along an object, and therefore may assign different numbers (measures) to the same length. When multiple copies of a unit are not available, may sequentially, although arbitrarily, slide the unit along the object and count, similar to the strategies of "point-and-count" or "trace-and-count" (described in Figure 1).	Although the length of the object to be measured is represented and maintained, the scheme has not necessarily applied the same mental operations to the parts imposed on that length as "units."	Blue Strips Ladybugs

Figure 4 (*continued*)

Levels of the developmental progression	**Actions on objects**	**Tasks**
Length Unit Relater and Iterater (previously separate levels: Length Unit Iterater and Length Unit Relater) May be able to iterate using a longer unit, such as a ruler, before using a smaller, inch-long unit in that fashion. Initially may not be able to apply the idea that unit size has an impact on the number of units needed. Therefore, may incorrectly conclude in comparative situations that the length of an object is greater than that of another simply based on the number of units used, not taking into consideration the size of those units (smaller units used end to end along a shorter object may therefore result in the shorter object being longer).	Action schemes include the ability to iterate a mental unit along a perceptually available object. The image of each placement can be maintained while the physical unit is moved to the next iterative position (initially with weaker constraints on this placement). With the support of a perceptual context, scheme can predict that fewer larger units will be required to measure an object's length. These action schemes allow the application of counting-all addition schemes to be applied to measures.	Add Strips Different Units Different Feet Spaghetti

Figure 4 (*continued*)

Levels of the developmental progression	**Actions on objects**	**Tasks**
Path Measurer (new level; includes previous Length Measurer level)		
Uses a flexible object (e.g., a chenille stick) to represent the length of a nonstraight path in some indirect comparison situations.	An anticipatory scheme conceptually predicts that the object to be measured can be partitioned into equal-length segments, constraining the objects used to measure the object to be units in the mathematical sense.	Two Roads Length of Sign
Straightens out zigzag paths in some direct comparison situations.	Lengths and paths composed of connected lengths can be flexibly composed and decomposed, maintaining (conserving) the total length. This scheme is available for conscious reflection, allowing it to be applied to comparisons of lengths.	Broken Ruler Snakes 1
Understands the ruler as an object composed of equal units, therefore uses a broken ruler, or a ruler without numbers.	Previously developed schemes are interrelated and integrated, allowing them to be applied to a wider range of situations (e.g., when movement along an object or path is not involved).	Long Ribbon

knowledge to other conceptually related tasks (Baroody, Cibulskis, Lai, & Li, 2004). The tasks involved have a crucial role in giving meaning to the developmental progression and in shaping the hypothesis of the learning process. Therefore, it is important to emphasize that the task sequence is neither the only nor the best route; it is simply one possible plausible path out of many (Clements & Sarama, 2004b). Thus, because the researchers designed the assessment instrument used in this study, the findings are likely to reflect their interpretation of the developmental progression. Different tasks might have resulted in different findings—especially because many factors contribute to the tasks' difficulty. Thus, segmentation into levels can only be tentative, and absolute segmentation is not required by our theoretical framework (Sarama & Clements, 2009) but rather is another indication of distinct levels.

Children's ways of knowing are known to be variable. Task-based differences in performance and inconsistencies may be due to the contextual nature of knowledge (Vygotsky, 1978), or due to a lag between the first use of a correct strategy and its consistent use. According to Siegler's Adaptive Strategy Choice model (Siegler, 1995, 1996), the acquisition of a new correct strategy involves the process of learning to use a new strategy while learning to stop using earlier ones. We used the standard psychometric method of having at least two items for every level we wished to assess. We also checked the validity of the test items/levels of thinking connections with qualitative analyses. Nevertheless, the potential circularity of these relationships is another caveat. This is an initial validation; the combination of quantitative and qualitative methods restricted the number of items and the number of students involved. Future research building on this research could increase both. In summary, this study is an initial one, which must be replicated and extended to test more rigorously the (revised) learning trajectory.

Despite these caveats, the findings contribute to the knowledge base on length measurement by providing additional information concerning the different strategies children use to measure length, and how those strategies change as children get older. A valid developmental progression "provides a practical way of initially planning and organizing instruction" (Baroody et al., 2004, p. 248) whether or not all children progress along the predicted progression.

As documented in this study, development occurred consistent with Newcombe and Huttenlocher's (2000) descriptions, that is, involving the coordination of perceptual and conceptual judgments and the ability to verbally express knowledge. This is less consistent with Piaget's theory, which views development as a stepwise advancement through qualitatively distinct stages represented by the emergence of certain general reasoning abilities. Although children may not be able to verbalize and explain their use of the logical operations involved in a strategy, they can have an intuitive knowledge of them. Experience with appropriate activities can help the children stabilize and verbalize such knowledge. The findings regarding the contextual nature of children's knowledge is also in support of this view of development. The application of knowledge in a contextually complex problem-solving situation is harder in general than in a contextually simple, concrete situation (Bryant & Kopytynska, 1976; Carpenter & Lewis, 1976; Miller, 1989).

The main implication of the study for the teaching of measurement is that the teacher, starting from the child's level of development, should make available ample opportunities for the application and coordination of ideas and concepts. Well-targeted experiences help children achieve "integrated-concrete" knowledge, so "physical objects, actions performed on them, and abstractions are all interrelated in a strong mental structure" (Clements, 1999a, p. 48). Understanding the developmental progression described here should help teachers create or appropriate research-based learning trajectories that underlie and support effective instructional strategies such as formative assessment (Clements & Sarama, 2009; National Mathematics Advisory Panel, 2008; Sarama & Clements, 2009).

Further research is needed to continue the study of the developmental progression for length measurement. The partial credit Rasch model (Bond & Fox, 2001; Masters, 1982) or other alternatives might be used on new data. Such models could be useful in finding quantitative support for the development of strategies found in this research based on qualitative analyses that involved examining the performance of children with similar abilities for every task. The two extremities of the logit scale need to be examined in more detail. Considerations for sample selection include a better targeted sample with more students at the high-ability levels, and additional items of greater difficulty to better estimate the ability of children at the top end of the progression. More tasks for each level of the developmental progression could be added to make the error of child ability estimates smaller and to provide more detail regarding the evolvement of the strategies. That also would make the assessment more reliable and accurate in estimating children's abilities based on their performance on the assessment and, therefore, in determining the children's developmental levels. Finally, case studies of children representative of each level of the developmental progression, as well as of children in transition between two levels, would help to explore and elaborate the nature of development and the levels of the learning trajectory for length measurement. Teaching experiments should be done to explore how children move from one level of the developmental progression to the next.

References

Baroody, A. J. (2004). The developmental bases for early childhood number and operations standards. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 173–219). Mahwah, NJ: Erlbaum.

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. doi:10.1207/s15327833mtl0602_8

Battista, M. T. (2004). Applying cognition-based assessment to elementary school students' development of understanding of area and volume measurement. *Mathematical Thinking and Learning, 6*, 185–204. doi:10.1207/s15327833mtl0602_6

Bearison, D. J. (1969). Role of measurement operations in the acquisition of conservation. *Developmental Psychology, 1*, 653–660. doi:10.1037/h0028260

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum. doi:10.1111/j.1745-3984.2003.tb01103.x

Boulton-Lewis, G. M. (1987). Recent cognitive theories applied to sequential length measuring knowledge in young children. *British Journal of Educational Psychology, 57*, 330–342. doi:10.1111/j.2044-8279.1987.tb00861.x

Boulton-Lewis, G. M., Wilss, L. A., & Mutch, S. L. (1996). An analysis of young children's strategies and use of devices for length measurement. *Journal of Mathematical Behavior, 15*, 329–347. doi:10.1016/S0732-3123(96)90009-7

Braine, M. D. S. (1959). The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychological Monographs: General and Applied, 73*(5), 1–43.

Bredekamp, S. (2004). Standards for preschool and kindergarten mathematics education. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 77–82). Mahwah, NJ: Erlbaum.

Bryant, P. E., & Kopytynska, H. (1976). Spontaneous measurement by young children. *Nature, 260,* 773. doi:10.1038/260773a0

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499–531. doi:10.3102/00028312026004499

Carpenter, T. P., & Lewis, R. (1976). The development of the concept of a standard unit of measure in young children. J*ournal for Research in Mathematics Education, 7*, 53–58. doi:10.2307/748765

Clarke, D., Cheeseman, J., McDonough, A., & Clark, B. (2003). Assessing and developing measurement with young children. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement*, 2003 Yearbook of the National Council of Teachers of Mathematics (NCTM) (pp. 68–80). Reston, VA: NCTM.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*, 408–411. doi:10.1111/1467-9280.00177

Clements, D. H. (1999a). "Concrete" manipulatives, concrete ideas. *Contemporary Issues in Early Childhood, 1*, 45–60. http://www.wwwords.co.uk/ciec/

Clements, D. H. (1999b). Teaching length measurement: Research challenges. *School Science and Mathematics, 99*, 5–11. doi:10.1111/j.1949-8594.1999.tb17440.x

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*, 35–70. http://www.nctm.org/publications/jrme.aspx

Clements, D. H. (2008). Linking research and curriculum development. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed.; pp. 589–625). New York, NY: Taylor & Francis.

Clements, D. H., Battista, M. T., & Sarama, J. (2001). Logo and geometry. *Journal for Research in Mathematics Education Monograph Series: Vol. 10*. Reston, VA: National Council of Teachers of Mathematics. doi:10.2307/749924

Clements, D. H., & Sarama, J. (2004a). *Building Blocks* for early childhood mathematics. *Early Childhood Research Quarterly, 19*, 181–189. doi:10.1016/j.ecresq.2004.01.014

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

Clements, D. H., & Sarama, J. (2007a). Early childhood mathematics learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 461–555). Charlotte, NC: Information Age.

Clements, D. H., & Sarama, J. (2007b). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*, 136–163. http://www.nctm.org/publications/jrme.aspx

Clements, D. H., & Sarama, J. (2007c). *SRA Real Math Building Blocks PreK*. Columbus, OH: SRA/McGraw-Hill.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.

Clements, D. H., Sarama, J., & DiBiase, A.-M. (2002). Preschool and kindergarten mathematics: A national conference. *Teaching Children Mathematics, 8*, 510–514.

Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 299–317). Mahwah, NJ: Erlbaum.

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_5

Confrey, J. (1990). What constructivism implies for teaching. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics. Journal for Research in Mathematics Education Monograph Series: Vol. 4* (pp. 107–122). Reston, VA: National Council of Teachers of Mathematics.

Confrey, J., & Kazak, S. (2006). A thirty-year reflection on constructivism in mathematics education in PME. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present, and future* (pp. 305–345). Rotterdam, the Netherlands: Sense.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33–66. doi:10.1006/cogp.2001.0760

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403–434. doi:10.2307/749875

Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. New York, NY: Cambridge University Press.

Goldin, G. A. (1997). Observing mathematical problem solving through task-based interviews. *Qualitative research methods in mathematics education. Journal for Research in Mathematics Education Monograph Series: Vol. 9* (pp. 40–62). Reston, VA: National Council of Teachers of Mathematics. doi:10.2307/749946

Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517–546). Mahwah, NJ: Erlbaum.

Hawkins, K. W. (1987). Use of the Rasch model in communication education: An explanation and example application. *Communication Education, 36*, 107–118. doi:10.1080/03634528709378651

Hiebert, J. (1981). Cognitive development and learning linear measurement. *Journal for Research in Mathematics Education, 12*, 197–211. doi:10.2307/748928

Huttenlocher, J., Newcombe, N. S., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology, 27*, 115–147. doi:10.1006/cogp.1994.1014

Johnson, M. L. (1974). The effects of instruction on length relations on the classification, seriation, and transitivity performances of first- and second-grade children. *Journal for Research in Mathematics Education, 5*, 115–125. doi:10.2307/748948

Kamii, C., & Clark, F. B. (1997). Measurement of length: The need for a better approach to teaching. *School Science and Mathematics, 97*, 116–121. doi:10.1111/j.1949-8594.1997.tb17354.x

Kidder, F. R., & Lamb, C. E. (1981). Conservation of length: An invariant—A study and a follow-up. *Journal for Research in Mathematics Education, 12*, 225–230. doi:10.2307/748931

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

Linacre, J. M. (2006). WINSTEPS Rasch Measurement [Computer software]. Chicago, IL: Winsteps.com.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174. doi:10.1007/BF02296272

McClain, K., Cobb, P., Gravemeijer, K., & Estes, B. (1999). Developing mathematical reasoning within the context of measurement. In L. V. Stiff & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K–12*, 1999 Yearbook of the National Council of the Teachers of Mathematics (NCTM) (pp. 93–105). Reston, VA: NCTM.

Miller, K. F. (1989). Measurement as a tool for thought: The role of measuring procedures in children's understanding of quantitative invariance. *Developmental Psychology, 25*, 589–600. doi:10.1037/0012-1649.25.4.589

Miller, K. F., & Baillargeon, R. (1990). Length and distance: Do preschoolers think that occlusion brings things together? *Developmental Psychology, 26*, 103–114. doi:10.1037/0012-1649.26.1.103

Murata, A., & Fuson, K. C. (2006). Teaching as assisting individual constructive paths within an interdependent class learning zone: Japanese first graders learning to add using 10. *Journal for Research in Mathematics Education, 37*, 421–456. http://www.nctm.org/publications/jrme

National Association for the Education of Young Children & National Council of Teachers of Mathematics. (2002). *Early childhood mathematics: Promoting good beginnings—A joint position statement of the National Association for the Education of Young Children (NAEYC) and the National Council of Teachers of Mathematics* (*NCTM*). Washington, DC: NAEYC/NCTM.

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

Newcombe, N. S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.

Nunes, T., Light, P., & Mason, J. (1993). Tools for thought: The measurement of length and area. *Learning and Instruction, 3*, 39–54. doi: 10.1016/S0959-4752(09)80004-2

Outhred, L., Mitchelmore, M., McPhail, D., & Gould, P. (2003). Count Me Into Measurement: A program for the early elementary school. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement*, 2003 Yearbook of the National Council of Teachers of Mathematics (NCTM) (pp. 81–99). Reston, VA: NCTM.

Petitto, A. L. (1990). Development of numberline and measure concepts. *Cognition and Instruction, 7*, 55–78. doi:10.1207/s1532690xci0701_3

Piaget, J., & Inhelder, B. (1967). *The child's conception of space* (F. J. Langdon & J. L. Lunzer, Trans.). New York, NY: W. W. Norton.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry* (E. A. Lunzer, Trans.). London, England: Routledge & Kegan Paul.

Piaget, J., & Szeminska, A. (1952). *The child's conception of number*. London, England: Routledge & Kegan Paul.

Raudenbush, S. W. (2009). The *Brown* legacy and the O'Connor challenge: Transforming schools in the images of children's potential. *Educational Researcher, 38*, 169–180. doi:10.3102/0013189X09334840

Sarama, J., & Clements, D. H. (2002). Mathematics curricula in early childhood. *Teaching Children Mathematics, 9*, 163–166. http://www.nctm.org/publications/tcm.aspx

Sarama, J., & Clements, D. H. (2003). Building Blocks of early childhood mathematics. *Teaching Children Mathematics, 9*, 480–484. http://www.nctm.org/publications/tcm.aspx

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

Sarama, J., & DiBiase, A.-M. (with Clements, D., & Spitler, M. E.). (2004). The professional development challenge in preschool mathematics. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 415–446). Mahwah, NJ: Erlbaum.

Sawada, D., & Nelson, L. D. (1967). Conservation of length and the teaching of linear measurement: A methodological critique. *The Arithmetic Teacher, 14*, 345–348. doi:10.2307/1129812

Schiff, W. (1983). Conservation of length redux: A perceptual-linguistic phenomenon. *Child Development, 54,* 1497–1506. doi:10.2307/1129812

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225–273. doi:10.1006/cogp.1995.1006

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*, 114–145. doi:10.2307/749205

Snyder, S., & Sheehan, R. (1992). The Rasch measurement model: An introduction. *Journal of Early Intervention, 16*, 87–95. doi:10.1177/105381519201600108

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

Stephan, M., & Clements, D. H. (2003). Linear, area, and time measurement in prekindergarten to grade 2. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement*, 2003 Yearbook of the National Council of Teachers of Mathematics (NCTM) (pp. 3–16). Reston, VA: NCTM.

Stephan, M., Cobb, P., Gravemeijer, K., & Estes, B. (2001). The role of tools in supporting students' development of measuring conceptions. In A. A. Cuoco & F. R. Curcio (Eds.), *The roles of representation in school mathematics*, 2001 Yearbook of the National Council of Teachers of Mathematics (NCTM) (pp. 63–76). Reston, VA: NCTM.

Szilágyi, J. (2007). *Young children's understandings of length measurement: A developmental progression* (Doctoral dissertation). The State University of New York, University at Buffalo.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Wilson, M. (1990). Measuring a van Hiele geometry sequence: A reanalysis. *Journal for Research in Mathematics Education, 21*, 230–237. doi:10.2307/749376

Wright, B. D., & Mok, M. (2000). Rasch models overview. *Journal of Applied Measurement, 1*, 83–106.

Authors

Janka Szilágyi, Assistant Professor, Department of Education and Human Development, The College at Brockport, State University of New York, 274 Brown, 350 New Campus Drive, Brockport, NY 14420; jszilagy@brockport.edu

Douglas H. Clements, Kennedy Endowed Chair and Professor in Early Childhood Learning, University of Denver, Educational Research, Practice and Policy, Morgridge College of Education, Katherine A. Ruffatto Hall 514, 1999 East Evans Avenue, Denver, CO 80208-1700; Douglas.Clements@du.edu

Julie Sarama, Kennedy Endowed Chair in Innovative Learning Technologies and Professor, University of Denver, Educational Research, Practice and Policy, Morgridge College of Education, Katherine A. Ruffatto Hall 514, 1999 East Evans Avenue, Denver, CO 80208-1700; Julie.Sarama@du.edu

Accepted August 14, 2012