2013

Curricular Treatments of Length Measurement in the United States: Do They Address Known Learning Challenges?

John P. Smith III
Michigan State University, jsmith@msu.edu

Lorraine Males
University of Nebraska-Lincoln, lmales2@unl.edu

Leslie C. Dietiker
Boston University

KoSze Lee
University of North Florida

Aaron Mosier
Michigan State University

Curricular Treatments of Length Measurement in the United States: Do They Address Known Learning Challenges?

John P. Smith III, Lorraine M. Males, Leslie C. Dietiker, KoSze Lee, and Aaron Mosier

Michigan State University

Lorraine M. Males is now at the Department of Teaching, Learning & Teacher Education, University of Nebraska–Lincoln.
Leslie C. Dietiker is now at the School of Education, Boston University.
KoSze Lee is now at the Department of Foundations and Secondary Education, University of North Florida.

Corresponding author — John P. Smith III, Department of Counseling Educational Psychology and Special Education, 509C Erickson Hall, 620 Farm Lane, Michigan State University, East Lansing, MI 48824; email jsmith@msu.edu

Abstract
Extensive research has shown that elementary students struggle to learn the basic principles of length measurement. However, where patterns of errors have been documented, the origins of students' difficulties have not been identified. This study investigated the hypothesis that written elementary mathematics curricula contribute to the problem of learning length measurement. We analyzed all instances of length measurement in three mathematics curricula (grades K–3) and found a shared focus on procedures. Attention to conceptual principles was limited overall and particularly for central ideas; conceptual principles were often presented after students were asked to use procedures that depended on them; and students often did not have direct access to conceptual principles. We also report five groupings of procedures that appeared sequentially in all three curricula, the conceptual principles that underlie those procedures, and the conventional knowledge that receives substantial attention by grade 3.

Published in *Cognition and Instruction,* 31:4 (2013), pp 388–433.
doi 10.1080/07370008.2013.828728

From the primary grades forward, many students in the United States, as well as in other countries (Bragg & Outhred, 2001; Hart, 1981; Nunes & Bryant, 1996), struggle to learn measurement. More than 30 years of empirical research, both large-scale studies and smaller more focused studies targeting student reasoning, have shown substantial weaknesses in students' understanding of measurement. In the United States and other countries, children work in the elementary grades to measure many quantities, including time, weight/mass, capacity/volume, and temperature. But more attention is devoted to measuring *space*—length, area, and volume—than other quantities, starting in the first year of formal schooling.[1] Despite frequent everyday experiences with spatial quantities, many students do not understand units of measure, how the iteration of units produces spatial measures, or how commonly used tools (rulers and computational formulas) generate measures (Battista, 2003; Lehrer, 2003**;** Lehrer, Jenkins, & Osana, 1998; National Research Council, 2001, 2007). U.S. elementary students' performance on the National Assessment of Education Progress (NAEP) has been lower in measurement than in most other domains (Thompson & Preston, 2004), even when important aspects of measurement have not been assessed (Blume, Galindo, &Walcott, 2007). The NAEP performance gap between White students and students of color has been greater for measurement than any other content domain (Lubienski, 2003). Though case studies have shown that deeper and more robust learning is possible with the right classroom experiences and teaching practices (Lehrer, Jaslow, & Curtis, 2003; Stephan, Bowers, Cobb, & Gravemeijer, 2003), we know little from empirical research about the factors that contribute to weak learning of spatial measurement in typical classrooms.

This article explores the hypothesis that the nature of current elementary *curriculum materials,* especially their procedural and conceptual content and how they express that content to students, is one factor contributing to students' difficulties to learn length measurement. We will not claim a causal relationship between curricular content and students' documented learning challenges. Indeed, we devote substantial space to presenting a framework that identifies multiple factors that interact to cause the patterns of weak learning. Rather, our central claim is that a strong correlation exists between students' challenges and the content of current U.S. elementary mathematics curricula, particularly with the documented challenges that students face understanding length measurement.

1 In the United States, the first year of formal schooling is Kindergarten; in other countries, it is grade 1.

To introduce and frame our curricular analysis, we first review the problem of learning spatial measurement, then summarize what research has revealed (and not yet explained) about children's understanding of length measurement specifically, and last describe a framework of factors that are jointly responsible for weak measurement learning.

Measurement and Evidence of Weak Learning

Our analysis focuses on length measurement, but the problem of learning measurement extends beyond that specific quantity. That said, the measurement of space begins with length measurement, and more attention is given to length measurement in the primary grades in the United States than any other quantity, as judged by state standards (Kasten & Newton, 2011). At its core, measurement is the assignment of a numerical value to a continuous quantity. Given a suitable unit, all similar continuous quantities (e.g., all lengths) are made discrete by segmenting them in unit parts. So all measures of continuous quantities have two components, the unit of measure and the number of those units that fill out or exhaust the particular quantity measured.

In the United States, NAEP has consistently provided evidence of weak student learning of measurement. One indicator is performance on measurement items relative to other content domains. Although fourth graders have performed comparatively well on measurement, eighth graders' measurement performance has remained low since 1990, along with geometry and spatial sense (Thompson & Preston, 2004). Similarly, on the Third International Mathematics and Science Study (TIMSS), eighth grade U.S. students also scored significantly lower on measurement, both relative to other countries and to their own performance on other content domains (National Center of Educational Statistics, 1997). A second indicator has been performance on items requiring explanation or solution of nonroutine problems. Where U.S. fourth graders have been generally successful reading a ruler when one end of the object is aligned with the zero mark, they performed very poorly (20–25% correct over multiple assessments) when the object was aligned at a nonzero unit mark and the zero mark was not shown—frequently called the *broken ruler* task (Kloosterman, Rutledge, & Kenney, 2009). This result suggests that students do not understand the structure of rulers and may be simply reading off the ruler number at the end of the object whether it is appropriate or not. The performance gap between White and minority students noted above was greatest for nonroutine items like the broken ruler task (Lubienski & Crockett, 2007).

Smaller-scale studies targeting students' reasoning have provided more insight into the nature of their struggles with measurement. Multiple studies have shown that middle school students do not distinguish the perimeter (a length) from the area of simple geometric shapes (Chappell & Thompson, 1999; Woodward & Byrd, 1983). They take equal perimeters as evidence of equal areas, apparently applying "same A, same B" reasoning (Stavy & Tirosh, 2000), and area responses have been attractive distracters on perimeter items (and vice-versa). In studies of length and area measurement, students are frequently drawn to count points rather than intervals of space, whether these are ruler marks for length (Bragg & Outhred, 2004) or geoboard pegs for area (Kamii & Kysh, 2006), suggesting interference from their experience of counting sets of discrete objects. Substantial evidence has also indicated that computational formulas for area and volume are weakly understood (Battista, 2003). Even adults report that "area is length times width," apparently as a general definition, where that formula only applies to rectangles (Schifter & Szymaszek, 2003).

Weak measurement learning carries significant costs for subsequent learning of mathematics and science. Measurement gives elementary students' direct contact with continuous quantities to complement their extensive experience with discrete quantities in learning base-10 number and operations. Even before they encounter the mathematical and scientific content in middle and high school that depends on the measurement of continuous quantities (e.g., density, work, force, torque), students are expected to learn about fractions and rational numbers in the upper elementary grades—a topic that is difficult to teach and learn solely from a basis of discrete quantity (Freudenthal, 1983; Thompson & Saldanha, 2003). Weak understanding of measurement, particularly of how unit iteration transforms continuous quantities into discrete quantities, may also limit students' ability to understand calculus, where the interplay between continuous and discrete quantities is so central (Thompson, 1994a).

Reviewing research on length and area measurement, Stephan and Clements (2003) have commented, "'Something is clearly wrong with [measurement] instruction' (Kamii & Clark, 1997) because it tends to focus on the procedures of measuring rather than the concepts underlying them" (p. 3). Though the character of classroom *instruction* may be strongly shaped by curriculum materials, no study to date has examined the nature and quality of length measurement content (or spatial measurement more generally) in current U.S. elementary *curriculum* materials. This study examined the character of the written curriculum for length measurement to assess the evidence that curricular limitations contribute to the problem of weak learning.

Research on Children's Understanding of Length Measurement

A significant body of empirical research has examined the development of students' understanding of length measurement and the teaching practices that support and limit it. We emphasize four main themes in this literature, each representing a major challenge to student learning: (a) understanding length as a stable and measurable quantity, (b) understanding the properties of manipulable units of length, (c) understanding the structure of rulers, (d) measuring complex paths (not single line segments). In each case, we also consider important issues that this research has left open.

Piaget's Work: Length Conservation and Transitive Inference

Research on children's understanding of length measurement began with Piaget's foundational studies (Piaget, Inhelder, & Szeminska, 1960). He claimed that the conservation of length as a quantity and the understanding of transitive inference were prerequisites to understanding units and "metric" measurement, including the use of rulers. Piaget also asserted that young children experience *length*, an amount of "filled" space occupied by an object or path of travel, and *distance*, an amount of "unfilled" space, as different quantities, and only later come to see them as equivalent. He argued that children's understanding of length measurement proceeds sequentially from comparative judgments of length (by direct comparison and visual inspection), to the use of intermediate objects to compare objects by transitive inference when visual inspection does not suffice, to the use of iterable units as universal intermediate objects. He argued that two basic mental processes were involved in the construction of length units: the conservation of length when subdivided into parts and the coordination of subdivision with the order of spatial position of the iterated part. For Piaget, measurement only became "metric" when the child understood length units conceptually; only then was counting units sensible for measuring and comparing lengths. Understanding length units conceptually was seen as prerequisite to understanding and using rulers.

Where researchers have not questioned the conceptual importance of conservation and transitivity, they have debated their status as developmental prerequisites to length measurement (Clements & Sarama, 2009). Kamii and Clark's (1997) study of grade 1–5 students provided support for Piaget's claim. When presented with two equal-length segments in an inverted *T* shape that appeared unequal due to perceptual illusion, many primary grade students failed to compare the segments because they could not use either a larger intermediate object (and thereby transitive inference) or a smaller one (unit iteration) effectively. Kamii (2006) has argued that transitivity must be mastered before unit iteration becomes conceptually meaningful. In contrast, Hiebert (1981, 1984) reported that some length measurement competencies such as indirect comparison and the iteration of units were not dependent on conservation and transitivity. His first graders who conserved length and used transitive inference only differed from those who failed in those tasks in their understanding of the inverse relation between the size of a length unit and the number of units required—a key measurement concept that Piaget did not address.

Informed by Piaget's work, educators have consistently argued for a sequence of instructional activities for length measurement that begins with qualitative comparison (e.g., which of two objects is longer?); moves to indirect comparison and the repeated use of body parts, everyday objects, and manipulatives as nonstandard units; and finally introduces standard units and rulers (National Council of Teachers of Mathematics [NCTM], 2000, 2006; Van de Walle, 1994; van den Heuvel-Panhuizen & Buys, 2008). The argument for nonstandard units preceding standard units—an issue of sequence that Piaget did not address—is that the diversity of and variation within nonstandard units (e.g., body parts) motivates the need for standard units (NCTM, 2000; Van de Walle, 1994). Although this sequence has deeply influenced curriculum and instruction, it has not gone unchallenged. Boulton-Lewis and colleagues (1996) and Clements (1999) have provided evidence that primary students can effectively work with both nonstandard and standard units simultaneously. Nunes, Light, and Mason's (1993) primary students performed better with rulers than with suitable everyday materials like string in some length measurement contexts.

Piaget's foundational work and the educational practice that has followed are not without their limitations. As with other parts of his research, Piaget's data did not clearly indicate the particular mental operations that he claimed to underlie and explain children's development. The construction of units from subdivision and change/order of position remains covert in children's activity and therefore difficult for educators to see. The prerequisite role of conservation and transitivity has been questioned on empirical grounds. Piaget offered little evidence for the experiential distinction between length and distance, and his work was silent on the use of rulers and whether they introduce any new challenges for children. Subsequent research has taken up some of these issues.

Length Units and Their Properties

Where Piaget saw length units as the synthesis of partitioning and order/change of position,[2] more recent research has identified different conceptual components of length units and examined how students struggle to coordinate them in their measurement activity. These elements are that: (a) all instances of a unit must be identical, (b) a collection of units must exhaust the entire space to produce a length measure (tiling), (c) the count of those units is the length measure, (d) the careful successive placement of a single unit (iteration) can produce the equivalent of an exhaustive tiling, and (e) the size of a length unit is inversely related to the number required to measure any object or path (Lehrer, 2003; Stephan & Clements, 2003). This research has generally shown that students master these ideas gradually; their early use of length units frequently satisfies some elements while violating others. For example, students, especially in the primary grades, (a) can mix different-sized units in measuring simple lengths (Clements & Sarama, 2009), even when enough same-sized units are available (Lehrer et al., 1998); (b) can partition lengths into "equal" segments that are in fact unequal (Clements, Battista, & Sarama, 1998); (c) can align units with the endpoints of objects, leaving spaces between units (Lehrer, et al., 2003; National Research Council, 2007); (d) fail to complete length measurements if their supply of units is insufficient (Bragg & Outhred, 2001; Lehrer, 2003); (e) resist placing a final unit if that unit will overlap that endpoint of the space (Clements & Sarama, 2009); and (f) struggle to coordinate the counts of units (and fractions of units) with their placement or use (Lehrer, 2003; Lehrer et al., 1998; Stephan, et al., 2003). Some of Hiebert's (1984) first graders who conserved length but did not understand the inverse relation between unit size and measure treated different size length units as equivalent. Lehrer (2003) has suggested the opposite is also possible: understanding the inverse relationship without making all units identical. In

2 This explanation paralleled his argument for cardinal numbers as the synthesis of classification and seriation (order relationships).

sum, these observations show that mastering length units is not an easy achievement for many children, because it involves the coordination of many interrelated conceptual elements.

On the other hand, where studies have enriched our understanding of the challenge and understanding of length units, they have not revealed much about the dynamics of learning. For example, cross-sectional studies have shown that older children (e.g., fifth graders) are more successful than younger children in always using identical units (Lehrer et al., 1998), but longitudinal studies of the progressive development of unit have been absent. Recently, Barrett and colleagues (2012) carefully examined the length measurement work of a small group of grade 2 students over 7 months of schooling, frequently assessing their learning using tasks designed to challenge and build from their current understandings. One case provided explicit evidence that some children can tile consistently before they can iterate a single unit and that physical motion (sweeping a finger along the path) appears to support the understanding of unit iteration. This study has shown how the coordination of physical motion, marks locating intervals of space, and counts of those intervals remains challenging for some students, even with carefully chosen tasks and supportive interactions with an interviewer.

Understanding and Using Rulers

Most cultures provide children with physical tools (foot rulers, yard and meter sticks, and tape measurers—collectively *rulers*) specifically designed for measuring length. Understanding rulers as length measurement tools includes knowing that they are composed of identical units (and subunits), that their marks indicate the beginning and end of units (and subunits), and that any unit mark can serve as zero in measuring lengths. But extensive evidence has shown that students' ability to use rulers in the standard way—aligning the object at zero and reading their measures from the ruler mark opposite the other end—does not imply their understanding of how and why they work. Because the standard use of rulers can be learned in a rote fashion, one common test of understanding, in both large-scale and small-scale studies, has been to present objects to be measured *not* aligned at the ruler's zero mark (Barrett et al., 2012; Bragg & Outhred, 2001, 2004; Kamii, 2006; Lehrer et al., 1998; Levine, Kwon, Huttenlocher, Ratliff, & Deitz, 2009; Nunes & Bryant, 1996). In some studies, the part of the ruler with the zero mark has been broken off (e.g., Kamii, 2006).

Elementary students find these tasks difficult and make two common errors: counting the marks on the ruler, not intervals separating them and reading the number adjacent to the end of the object at the "large" end of the ruler without compensating for the object's displacement from zero. Bragg and Outhred (2001) reported this task was challenging even for fifth- and sixth-grade students, who also struggled to identify the length units on their rulers. More generally, Battista (2006, 2012) has shown that the presence of marks segmenting linear space into equal intervals (e.g., dots and tick marks) leads many students to count marks rather than intervals to determine lengths. Another grade 2 student in Barrett and colleagues' longitudinal study (2012), who was able to iterate units vacillated for some time between counting intervals and counting unit marks on rulers when objects were not aligned at zero. Levine and colleagues' training study (2009) showed that grade 2 students whose training involved physically placing units on ruler intervals *and* comparing aligned and misaligned length measurements were more successful on length measurement than other students whose training involved only rulers and objects aligned at zero. Many did not initially see that measuring the same object with rulers and with physical standard units necessarily produced the same measure.

A second task assessing students' understanding of rulers gives children the choice of measuring with either "correct" rulers (all equal intervals) or "incorrect" rulers (some unequal intervals). First grade children have been content to use incorrect rulers, where most third graders have selected only correct rulers (Lehrer et al., 1998; Pettito, 1990).

This research has shown that many elementary students may not understand that the ruler marks divide equal intervals of space (so any unit mark can server as zero), even when they have successfully used rulers in the conventional way for some time. Knowledgeable use of rulers requires children to coordinate intervals of space (units), the marks used to indicate and count those units, the accumulated distance associated by those counts, and the numeral marks on rulers (Barrett et al., 2012). This coordination takes time and repeated experience and may crucially depend on the character of tasks and instructional support. But as yet, research has only begun to explore the instructional conditions that effectively support such growth.

Complex Paths and Perimeter

In Piaget's early work, some children who were successful reasoning about the length of simple straight-line paths failed when paths had corners (Piaget et al., 1960). Some who conserved length for two straight-line paths that were not aligned opposite each other failed to conserve when the one path was bent. They instead attended to the relative position of endpoints. Hiebert (1981, 1984) replicated this result. For the bent path, distance along the path was greater than the distance between the endpoints. In addition, when measuring complex paths (both closed and open) that are drawn on grids, students frequently count squares as units of length. Generally, each square corresponds to one unit length (the side of the square), but corners create problems. Some students count squares that touch the path at the corner but account for no distance or fail to double count "inside" corner squares that have two sides on the path (Battista, 2006, 2012). These responses indicate either loss of attention to the length attribute or failure to distinguish the side (as a length unit) from the more visually salient square. It is another example of how the material conditions of length measurement tasks can influence students' reasoning.

Similarly, Barrett and Clements (2003) showed that elementary students' reasoning about the length of simple segments was not a good predictor of their reasoning about the perimeter of rectangles. Even with supportive instruction, some fourth graders struggled to coordinate the marks partitioning line segments with their counts of the corresponding length units, reasoned inconsistently with length units around corners, and could not coordinate the spatial properties of rectangles (e.g., opposite sides are equal in length) with their judgments of side lengths. In a subsequent cross-sectional analysis of elementary to high school students, the older students focused more on part–whole numerical relationships involving lengths but often lost the coordination of number with space (e.g., making drawings whose parts summed correctly but were disproportionate in length; Barrett, Clements, Klanderman, Pennisi, & Polaki, 2006). Overall, both open and closed paths with corners present new challenges for students that do not arise in the measurement of simple segments and objects.

Finally, once two-dimensional shapes are the focus of length measurement work (e.g., to determine their perimeter), students struggle to distinguish linear measurement from area measurement. This confusion is indicated by the frequency with which perimeter responses are given to area questions (and vice-versa). Attention to area as the quantity of space contained "inside" a two-dimensional shape may lead to the faulty assertion that perimeter is what is "outside" (Clements & Sarama, 2009). Yet it is unclear how frequently the perimeter-area confusion results from vague classroom discourse (where speakers fail to distinguish which quantity in a two-dimensional figure they are referring to) and from a deeper struggle to distinguish linear and area measurement. What is clear is that complex paths, both open and closed, introduce new challenges for students in measuring length.

Conceptualizing the Relationship Between Curriculum and Learning

We now present the main theoretical constructs that shaped our methods for analyzing written measurement curricula, including (a) the meaning of the term *curriculum*, especially the relationship between *written* and *enacted* curriculum; (b) the factors that have been implicated in creating the problem of weak learning of measurement; (c) the core distinction between mathematical concepts and procedures; and (d) the role of a specific tool (rulers) in length measurement. The concepts, procedures, and tools of length measurement directly informed how we structured our scheme for coding the content of written curricula.

Conceptualizing Curriculum

Students' experience with school mathematics is a joint product of curriculum that presents that content and the teaching practices that enact that curriculum. What students learn from that experience is substantially shaped by their prior knowledge used to make sense of the given content (National Research Council, 2001). We use the term *written curriculum* to mean published resources that present mathematics content and have been designed for teachers' use in classrooms (Stein, Remillard, & Smith, 2007). Written curriculum includes traditional print materials (textbooks and Web-based activities and lessons that take similar form) as well as technology that presents or frames specific mathematics work and activity for students. Following Stein et al. (2007), we distinguish the written curriculum from the *intended curriculum* and the *enacted curriculum*.[3] The former references the teachers' plans for how they will present written lessons to their students, while the latter refers to the lessons that actually unfold in their classrooms. Two important aspects of the intended curriculum are the deletions of parts of the written curriculum (e.g., lessons and particular activities or elements of lessons) as well

3 See also the framework developed by the *Center for the Study of Mathematics Curriculum* available at http://www.mathcurriculumcenter.org

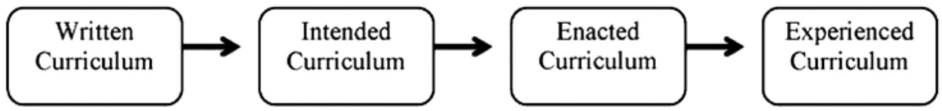

Figure 1. Major steps in curriculum transformation.

the addition of written materials drawn from other sources. To complete this series of transformations, the *experienced curriculum* designates the mathematics content that students actually engage. The enacted and experienced curriculum would coincide were it not for students' potential to selectively attend and filter what is happening in their classroom. Figure 1 summarizes this process of curriculum transformation.

Factors Influencing Weak Learning of Measurement

This view of curriculum (and curriculum transformation) informs a more specific framework for understanding the different factors that are jointly responsible for the problem of weak learning of measurement. Figure 2 identifies seven factors that research has indicated, either explicitly or inferentially, are responsible for shaping the enacted measurement curriculum. Factors indicated with (C) are features of written curriculum; factors indicated with (T) are features of the intended or enacted curriculum or teachers' knowledge that shape it.

Three factors concern mathematics curriculum broadly and measurement curriculum specifically: written curricula, state standards and assessments, and static representations. Written curricula are one likely factor of influence, given the evidence that mathematics teaching in many classrooms is strongly influenced by the content of teachers' assigned textbooks (Grouws, Smith, & Sztajn, 2004; Silver, 2009; Weiss, Pasley, Smith, Banilower, & Heck, 2003). If textbooks do strongly influence instruction and students' struggle with measurement, some influence of written curricula is likely. Two different features of written curricula are likely involved: textbooks may lack important elements of measurement *content* (Smith et al., 2008) and they also typically *place* measurement content late in their materials, decreasing the likelihood that teachers will teach that content before the school year ends. Late placement may also signal lower importance to teachers. *State standards and assessments* have also influenced the elementary mathematics content, especially since the passage of *No Child Left Behind* legislation. In some classrooms, daily instruction has been strongly shaped, if not determined by

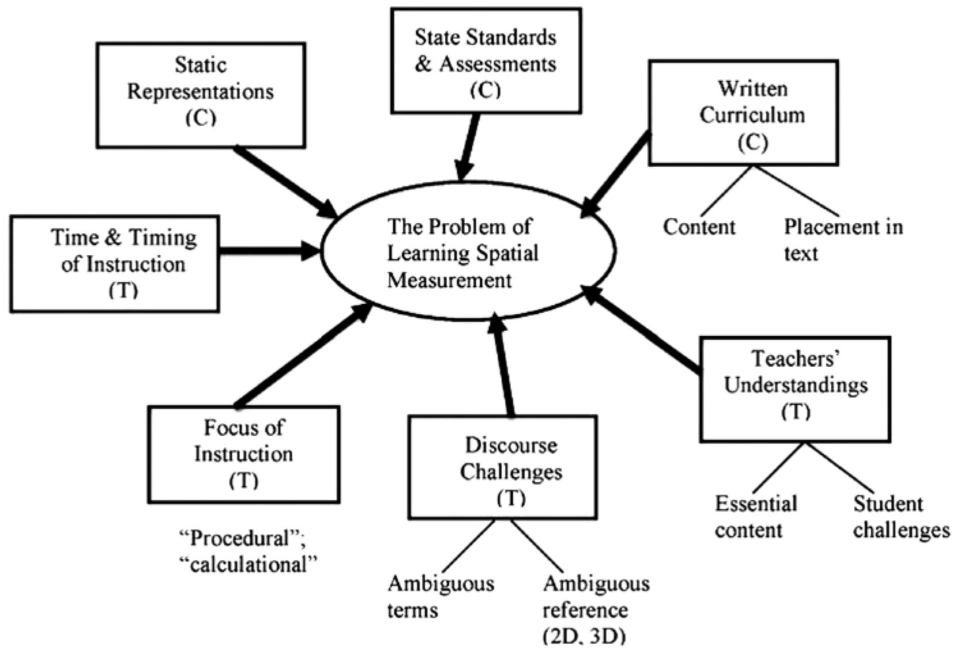

Figure 2. Factors contributing to the problem of weak measurement learning. "C" indicates curricular factors; "T" indicates factors that concern teaching or teachers.

particular standards, so content that does not appear in those standards may not be taught. The new *Common Core State Standards in Mathematics* may come to have an equal or even greater influence (Common Core State Standards Initiative, 2010). Where the content of written curricula and state standards and assessments influence all of school mathematics, a third factor applies specifically to measurement: the dominance of *static representations* of space, units, and space-filling actions. It is difficult to represent the movement (iteration) of units through space within the confines of the printed page, yet bodily motion and the movement of units may be an important, even essential element in students' learning (Barrett et al., 2012; Lehrer, 2003). The development of dynamic geometry software can be seen as a response to this limitation (Clements et al., 1998; Sinclair & Jackiw, 2002), but these tools are not widely used in elementary classrooms.

Four other factors concern the nature of measurement teaching. *Time and timing of instruction* concerns the duration and location of instructional time during the school year. The presence of measurement content in the curriculum does not translate into students' opportunity to learn until teachers allocate significant instructional time to that content. Elementary teachers report heavy emphasis on number and operations far more often than they do for measurement (Gehrke, Knapp, & Sirotnik, 1992; Smith, Arbaugh, & Fi, 2007). The late placement of measurement content in textbooks may also decrease the time devoted to teaching that content. With respect to the *focus of instruction*, many researchers have reported elementary teachers' emphasis on teaching procedures (Gehrke, et al., 1992) and to numerical calculation (Thompson, Philipp, Thompson, & Boyd, 1994) in carrying out those procedures. This characterization, and a related absence of attention to conceptual issues, has been cited for the teaching of measurement specifically (Kamii & Clark, 1997; Kamii & Kysh, 2006; Lehrer, 2003). Where a strong procedural focus in instruction may not be inherently problematic for students' learning, conceptual issues can be productively addressed in work on procedures (Thompson et al., 1994), but many elementary teachers lack the deep understanding that would support thoughtful attention to connections between measurement procedures and concepts (see below).

Teachers' work to lead classroom discussions that deeply engage students in mathematics can be difficult in any topic area, but measurement poses specific *discourse challenges*. Some basic spatial measurement terms lack clear meaning and patterns of use. *Length* is difficult to define accurately and accessibly for elementary students, and length-related terms have multiple and intersecting meanings (Battista, 2006). For example, the *width* and *length* of rectangles are both lengths, and the adjective *long* can refer either to linear paths or temporal durations. *Base* and *height* can refer either to geometric features of rectangles and parallelograms or to the algebraic expressions (*B* and *H*) that stand for their lengths (Herbel-Eisenmann & Otten, 2011). Complex and shifting patterns of reference likely add to students' challenges in learning from classroom discussions of measurement. Last, because written curricula must be enacted by teachers, and these enactments can change the nature and demand of mathematical tasks and activities for learning (Stein, Grover, & Henningsen, 1996), *teachers' understandings* of measurement shape and constrain students' opportunities to learn. Substantial evidence indicates that both practicing and preservice elementary teachers' understanding of measurement is limited, often in ways that reflect students' limitations (Baturo & Nason, 1996; Chappell & Thompson, 1999; Simon & Blume, 1994; Woodward & Byrd, 1983). Teachers' knowledge can be limited in their mastery of measurement content and in their understandings of how and why students struggle with that content.

Although interactions among factors are not explicitly represented in Figure 2 and have not explicitly explored in research, they are likely. For example, late placement of measurement content in textbooks likely influences “downward” pressure on the overall time that teachers devote to teaching measurement. Similarly, teachers’ understandings are likely influenced, in the past and present, by the content of written curricula and state standards and assessments that structure their teaching environments. Where the nature and force of these interactions remain to be explored; we have not presented Figure 2 to indicate a set of disjoint, noninteracting factors.

Students’ Opportunity to Learn

Figure 2 represents the main factors implicated by prior research in the weak learning of measurement. These factors can also be seen as sources of limitation in *students’ opportunities* to develop deep understandings of measurement. Prior work on students’ opportunity to learn (OTL) has focused on the written curriculum and examined and analyzed broad topics or problem types, such as “measurement units” or “measurement estimation and errors” (Floden, 2004; Schmidt, McKnight, & Raizen, 1997). But given recent theoretical work in mathematics and science learning (diSessa, 1993; Siegler, 1996; Smith, diSessa, & Roschelle, 1993/94), topics and problem types may be too coarse a grain size to understand OTL in detail. If learning depends on acquiring and coordinating many different elements of content knowledge, as some measurement studies have directly suggested (e.g., Barrett et al., 2012), then focusing the analysis of OTL on more specific ideas and processes in measurement may be more productive. Independent of the grain size of target content, explicit statements of the ideas and processes in the written curriculum provide students some access to those ideas, and their absence restricts if not eliminates their opportunities. Generally speaking, when particular ideas and processes are mentioned more often in written curricula, students’ OTL increases—particularly when that content appears in materials that students see and work with (i.e., in student editions of textbooks). Although very low OTL in written curriculum for any particular idea or process of measurement may not radically undermine students’ ability to develop deep understandings, such limits for many elements would be problematic. That pattern would put more pressure on students’ independent constructive activity, and most elementary teachers lack the resources to make up for those deficits.

Procedural and Conceptual Knowledge

The distinction between mathematical procedures (and procedural knowledge) and mathematical concepts (and conceptual knowledge) is implicit in numerous factors in Figure 2 above. Typical measurement instruction has been characterized as procedurally focused, with infrequent attention to basic measurement concepts (e.g., Stephan & Clements, 2003). Teachers' understandings are more centered on procedures, and the content of written curricula may be as well. But how do these terms apply to measurement, especially length measurement? Hiebert and Lefvre (1986), whose work is frequently cited, defined conceptual knowledge as "knowledge that is rich in relationships . . . as a connected web of knowledge" (p. 3) and procedural knowledge as consisting of two parts—"the formal language, or symbol representation system, of mathematics" and "the algorithms, or rules, for completing mathematical tasks" (p. 6). Where we agree that conceptual knowledge must be richly interconnected to constitute deep understanding, such connections may not singularly define conceptual knowledge but may apply to procedural knowledge as well (Star, 2005). Moreover, the central characteristic proposed for conceptual knowledge, "rich in relationships," is vague and apparently applicable only to clusters of related elements, not individual conceptual elements. Similarly, procedural knowledge may include a broader range of mathematical processes than well-defined algorithms alone.

As we will describe in more detail below, we use the term *concepts* (and *conceptual knowledge*) to designate the general principles that underlie and justify procedures. In measurement, two important sets of principles concern the nature of spatial quantities (length, area, and volume) and the properties of units applied to measure those quantities. Where it is important that students learn interrelationships among these ideas, these and other conceptual principles can be stated individually. For example, a key conceptual principle in measurement is the inverse relationship between the size of the unit and size of the resulting measure (in our coding scheme, *Unit- Measure Compensation*): When measuring the same quantity, larger units will produce smaller measures (and vice-versa). We use the term *procedures* (and *procedural knowledge*) to refer both to well-specified algorithms (e.g., the formula for computing the perimeter of rectangles) and to other more weakly specified measurement actions that students are asked to carry out. For example, the *Visual Estimation of Length* is a measurement procedure in our analysis, even though its constituent steps—choosing a unit, mentally iterating that unit through the object (or some part of it), and composing the resulting estimate—are often not made explicit in written curricula (Chang, Males, Mosier, & Gonulates, 2011).We also include very simple measurement actions (e.g., the visual comparison of two lengths) that may be unitary acts, rather than ordered sequences of steps.

Physical Tools

Length measurement is relatively unique in spatial measurement in that the culture provides a class of related physical tools, *rulers* in the broadest sense, that directly support the production of length measures. Rulers are "given" in the practices of the adult culture (Vygotsky, 1978), practices carried out both in school and outside of school. Where it is important to recognize that rulers of various sizes and shapes (foot rulers, yard and meter sticks, tape measurers) are widely used, their structure, as we have seen, often remains opaque. Tools are not knowledge *per se*, either procedural or conceptual, though concepts structure them and procedures engage them. They introduce the contingent and culturally defined dimension of measurement that we incorporated in our coding scheme as *Conventional* knowledge, along with other conventions of length measurement.

Curriculum Analysis Methods

Oriented by this perspective, our study assessed elementary students' opportunity to learn length measurement based on a detailed analysis of the content of elementary written curricula (textbooks). We focused on describing, with substantial precision, what length measurement content appeared in these materials and how that content was expressed in text. We carefully distinguished different length measurement concepts, procedures, and conventions and searched for all instances of them in the curricula. This analysis allowed us to address the following research questions, which each address students' opportunity to learn length measurement from written curricula at successively finer levels of detail.

First, how much attention do written elementary mathematics curricula give to length measurement? Where the vast majority of our analysis focused on particular sentences in the text and knowledge elements, we also carried out a coarser analysis at the lesson level as a rough measure of how much space was devoted to length measurement in each curriculum and grade. It also allowed us to characterize where, in general terms, length measurement content appeared in each curricula.

Second, how do elementary curricula present length measurement content in grades K–3? That is, in what sequence do the curricula present length measurement topics and what specific concepts and procedures structure these topics? In answering this question we examined the similarity in the topic sequences of the different curricula and the degree to which they either match or deviate from the epistemological sequence proposed by Piaget.

Third, we examined how directly the curricula addressed four specific challenges that prior research has shown many students face in learning to measure length: How do curricula address the challenge of helping students understand length as a stable and measurable attribute? What opportunities do they provide for learning the conceptual properties of units? What opportunities do they provide for understanding the structure of rulers? And what opportunities do they provide for learning to measure complex paths (curvilinear and "jagged") in comparison to simple paths? In each case, we consider the frequency of specific concepts and procedures that provide opportunities to address and deal with those challenges. We also consider issues of order among concepts and the procedures they underlie and constrain.

There were five major steps in our method that addressed these questions: (a) our choice of mathematics curricula, (b) our process for locating length measurement content in those curricula, (c) our conceptualization of students' opportunity to learn, (d) our coding scheme for characterizing the content and expression of length measurement on textbook pages, and (e) our process for applying that scheme to the textbook data. These methods have also been described in prior work (Lee & Smith, 2011; Smith et al., 2008).

Selecting Written Curricula

Many different elementary mathematics curricula presently are used in U.S. classrooms (Dossey, Halvorsen, & McCrone, 2008), some with nontrivial differences in content (Stein et al., 2007). Given the fine-grained nature of our analysis, sampling from this population of written materials was necessary; we could not analyze all elementary curricula currently in use. After lengthy consideration, we selected three elementary textbook series: *Everyday Mathematics*, third edition, published by the University of Chicago School Mathematics Project (2007)—henceforth *EM*; *Mathematics*, *Michigan edition* (Charles, Crown, & Fennell, 2008), published by Scott-Foresman/Addison-Wesley—*SFAW*; and *Saxon Math*

(Larson, 2004)—*Saxon*. Curricula written in the spirit of NCTM's (1989) *Curriculum and Evaluation Standards for School Mathematics* (like *EM*) have been characterized as significantly different from more traditional, publisher-developed materials (like *SFAW*; Stein et al., 2007; Trafton, Reys, & Wasman, 2001). We chose *EM* and *SFAW* as examples of standards-based and publisher-developed curricula respectively, because they ranked first and second in national market share for elementary mathematics textbooks when we began our analysis (Dossey et al., 2008). *Saxon* materials were not as widely used, but their direct instruction approach and curriculum structure made them quite different from both *EM* and *SFAW*.[4]

We also analyzed one Singapore elementary mathematics curriculum that is also widely used (60% of that nation's elementary schools): *My Pals are Here! Math—MPAH* (Fong, Ramakrishnan,& Lau, 2007).We selected a Singapore curriculum to widen our focus from U.S. curricula and because that country's generally strong performance on cross-national comparisons of achievement in mathematics (Mullis et al., 1997; Mullis, Martin, & Foy, 2008). That said, Singapore's grade 4 students performed significantly less successfully in the geometric shape and measure domain than they did in number and data display in a recent TIMSS study—while outperforming U.S. students in all three domains (Mullis et al., 2008). We used the same methods in the Singapore analysis that we applied to the U.S. curricula (see below).We report some of the results from the *MPAH* analysis below (see also Lee & Smith, 2011), but primarily for purposes of top-level comparison to U.S. curricula.

Last, we examined more cursorily the more recent editions of two of the U.S. curricula, *EM*'s *Common Core edition* (University of Chicago School Mathematics Project, 2012) and *SFAW's EnVision Math* curriculum (Charles et al., 2011) to assess whether these newer materials had substantially changed students' opportunity to learn measurement while our analysis of their prior materials was underway. In both cases, these editions were written by the same author team that produced the prior versions. We were unable to procure a more recent edition of the *Saxon* curriculum from the publisher. Our more limited analysis of these materials explored the question: Were there significant differences between more recent and prior editions in terms of overall content and placement of length measurement content?

4 Stein and her colleagues (2007, pp. 325–326) characterized *EM* as "standards-based" and *SFAW* and *Saxon* as "conventional."

Locating the Length Measurement Content in the Texts

We first located all the curricular content in the grades K through 3 materials of the three main U.S. curricula that could reasonably be taken to provide opportunities to learn length measurement. We found significant content in all four grades in each textbook series. We stopped at grade 3, not because length measurement was absent from subsequent grades, but because the central processes, concepts, and tools for length measurement have been presented by grade 3 (Kasten & Newton, 2011; NCTM, 2000). It did not appear that adding additional grades would fundamentally influence our answers to the questions posed above. If there were problems with opportunity to learn in grades K through 3, it seemed unlikely that those problems would be corrected in later grades. Since kindergarten is not mandatory in Singapore, the *MPAH* curriculum did not produce materials for that grade, so we analyzed the length measurement content in primary 1, 2, and 3, equivalent to grades 1 through 3 in the United States.

Our criterion—"reasonably be taken to provide OTL for length measurement"—meant that locating length measurement content involved much more than simply finding the units and lessons that explicitly targeted that topic. U.S. curricula systematically distribute topic-specific tasks and problems throughout their materials, so length measurement tasks could be (and were) present in lessons focusing on other topics. Our criterion also led us to include content that some mathematics educators might not recognize as concerning length measurement. For example, we included the partitioning of one-dimensional objects and segments into equal-size parts and the construction and interpretation of bar graphs—when the bars were not "unitized" to support counting to determine their height—because these topics required students to reason about length. We adopted this broader and more inclusive criterion to decrease the likelihood that we would miss opportunities for learning about length measurement; the costs of excluding relevant content seemed to outweigh by far the costs for coding more data.

We included all teacher lesson guide pages and all student pages if these pages contained at least one instance of length measurement content. We also included all pages from other printed elements of the curriculum (e.g., student workbooks) if they were explicitly referenced on teacher lesson guide pages and included at least one such instance. *EM* and *SFAW* teacher lesson guide pages explicitly included particular student pages in a "wrap-around" format. In *Saxon*, which did not follow this format, teacher lesson guide pages were followed by student problem pages in most all lessons, first “guided practice” and then “homework.” Two coders examined every textbook page from each curriculum and grade, applied the criterion above, and resolved any disagreements about the presence of length measurement content as a pair. Activities and problem types that seemed questionably related to length measurement were brought to the larger research team for discussion and resolution. One particular recurring challenge was deciding where the counting of discrete quantity was more likely than reasoning about length as a continuous quantity. This challenge was addressed by carefully examining and judging particular problem types where the discrete/continuous challenge arose. For example, we excluded bar graphs with unitized bars because we judged that counting the collection of units as a set was more likely than reasoning about length, but we included bar graph content when bars were not segmented and bar heights had to be determined by qualitative comparison or estimation relative to a numerical scale. The outcome of this phase of the analysis was a large collection of textbook pages that contained some length measurement content, one set of pages for each curriculum and grade (K–3).

Specifying Dimensions of Opportunity to Learn

We tracked OTL in two complementary ways—as access to particular elements of measurement knowledge and via the different textual forms that express or call for that knowledge. First, OTL involves access to the mathematical ideas and processes involved in length measurement. We identified length measurement knowledge at the level of individual ideas, creating long lists of different elements of length measurement knowledge that could be expressed in single sentences or problems. But the measurement ideas and processes must be expressed in some textual form, and forms of textual expression also influence students’ OTL. First and most basically, some parts of the written curriculum are presented directly to students in *student* materials; other parts are presented to teachers and reach students only through *teachers’* speech and/or action. Equally important, whether students meet ideas and processes directly on the written page or indirectly through their teachers’ activity, forms of expression vary, and these variations also have implications for students’ OTL. Ideas and processes may be stated to students, they may be demonstrated or modeled for them, or students may be assigned tasks that engage an idea or require a process. Substantial evidence indicates that listening and viewing alone may provide insufficient support for many students to learn mathematics and that more direct activity and engagement is productive, if not necessary (Carpenter, Franke, & Levi, 2003; Hiebert et al., 1997; Lehrer, 2003; National Research Council, 2001; Stevenson & Stigler, 1992). If so, our analysis of OTL should address both knowledge content and expression.

Coding the Knowledge Content of Textbook Pages

Coding the textbook pages required carefully identifying different elements of length measurement knowledge. We knew of no such frame at the start of our study, so we developed one. Our scheme consisted of two independent dimensions: a list of individual elements of length measurement knowledge (e.g., all length measurements involve error) and a set of textual forms that textbooks use to express that knowledge, (e.g., statements, questions, demonstrations). We developed our list of knowledge elements from three sources: the mathematics itself, research on students' learning of measurement (e.g., Lehrer, 2003; National Research Council, 2001; Stephan & Clements; 2003), and careful inspection of the textbook materials, especially for procedural knowledge. Since researchers had argued that U.S. instruction in measurement has often been conceptually deficient (Kamii & Kysh, 2006; Lehrer, 2003; Stephan & Clements, 2003), we focused first on identifying elements of conceptual knowledge, then procedural knowledge, and finally a separate category of conventional knowledge. Our coding scheme was developed iteratively, as we sought to account for all instances of length content in our data.

Types of Length Measurement Knowledge and of Textual Expression

We viewed *conceptual* knowledge as expressions of basic principles that underlie and justify measurement procedures, systems, notations, and tools. Such knowledge provides the rationale for measurement procedures and practices. We understood *procedural* knowledge to refer to the actual methods for producing length measures, including measuring with tools, estimating, and computing. This category included (a) qualitative procedures (e.g., comparing which of two objects is longer by placing them side by side), (b) nonstandard and standard measurement procedures that produce length measures for simple segments or objects (such as aligning and moving physical units and using rulers, respectively), and (c) procedures for reasoning with lengths (e.g., determining the perimeter of a polygon from its side lengths). But we found conceptual and procedural knowledge insufficient to account for all length measurement content and developed a third category to code our data completely. *Conventional* knowledge concerns the systems, notations, and tools that cultures have invented to carry out measurement (e.g., the actual size of individual units like inches and centimeters, the abbreviation of units, and the placement of numerical scales on rulers). Since these conventions are mathematically arbitrary, they are epistemologically distinct from conceptual knowledge. In its final form our coding scheme contained 103 different knowledge elements (43 *conceptual*, 52 *procedural*, and 8 *conventional*). Due to its length, we have not included the complete scheme but exemplify it repeatedly, below and in Results.

We identified five basic textual forms that expressed length measurement content: *Statements*, *Demonstrations*, *Worked Examples, Questions*, and *Problems*. *Statements* are particular assertions about the nature of length, length units, and length measurement. They may express elements of *Conceptual*, *Procedural*, or *Conventional* knowledge. For example, *Conceptual* statements include explicit definitions of concepts such as the following for perimeter: "The distance around a figure is its perimeter" (*SFAW*, grade 3, p. 484A). *Demonstrations* are displays of measurement knowledge, almost always procedures, by teachers or students designated by teachers. In *Saxon*, grade 2, the text directed the teacher to demonstrate how to measure a four-inch line segment using a centimeter ruler:

> Let's measure our line segment using centimeters. There is usually a line near the beginning of the ruler that shows where to begin measuring. We will put this line on the first endpoint. Look along your ruler until you come to the other endpoint. About how many centimeters long is this line segment? Write "10 centimeters" below the line segment. (Lesson 102, page 5)[5]

Worked Examples present the solutions to measurement problems and therefore, like *Demonstrations*, generally expressed procedural knowledge. *Questions* are queries posed to students that require little reasoning, may be answered by one student, or are under teacher direction. For example, we coded the embedded query from the teacher in the *Demonstration* cited above as a *Question*. By contrast, *Problems* are queries to students that require a greater amount of reasoning and/or activity. Most students, if not all, are expected to respond, and the immediate context suggests that students are given time to respond. We coded the following

5 The *Demonstration* by the *Teacher* code was applied to the entire text, but here and in other cases, we also applied additional codes to particular constituent sentences.

task in *SFAW*'s grade 3 cumulative review of chapter 7 as a *Problem*: "Angie used all of her string to make jewelry. She used 8 inches of string to make a bracelet. She used 14 inches to make a necklace. How much string did Angie start with?" (p. 416).

Generally, we report our results for these five types only—with two exceptions. One concerns *Statements*. We coded instances that included all or most of the content of knowledge elements as articulated in our coding scheme as *Full Statements*. When omissions were quite significant, we coded those as *Partial Statements*. For example, we stated the key conceptual principle of *Unit Iteration* as, "Measures of length are produced by iterating a length unit (repeatedly adjoining) from one end of an object, segment, or distance to the other and then enumerating the number of iterations (e.g., by counting). Iterated units may not overlap or leave gaps." We coded the following text in *EM* Grade 1 as a *Partial Statement*: "There may have been gaps between units or accidental overlapping of units when measuring" (p. 283). This reference stated one component in *Unit Iteration* explicitly—the need to avoid gaps or overlaps between units—but not others (e.g., end-to-end placement). Second, *Statements*, *Questions*, and *Problems* were located either in the teacher lesson guides that students do not see or in the student materials that they do see. For those three textual types, we also coded each instance for *location*—in the *Teacher* (*T*) or *Student* (*S*) materials—because location could influence students' OTL. By definition, *Demonstrations* were given in *Teacher* materials, and *Worked Examples* in *Student* materials.

Making reliable distinctions between *Questions* and *Problems* required considerable discussion, operationalization, and in some cases, recoding. We used three equally weighted criteria: (a) Is the cognitive demand of the query more than simple recall or observation? (b) How many students does the curricular context suggest are expected to respond? and (c) Does the curricular context suggest that students can work autonomously on the task or under teacher's guidance? Queries were coded as *Questions* if (a) only simple recall or observation was required, (b) one student could answer, and (c) the context suggested that the query was embedded in a sequence of activity that the teacher should control. By contrast, queries were coded as *Problems* if (a) cognitive demand surpassed simple recall or observation, (b) response from most, if not all, students was expected, and (c) the context suggested that teachers should provide time and space for students to produce a response. In most cases, the last criteria was inferred when no evidence of teacher control (the third criterion for *Questions* above) was found. These criteria did not always align perfectly; in mixed cases, outcomes on two of three of the criteria decided the issue. But this operationalization of the distinction was certainly imperfect and involved significant coder judgment. A second limitation is that our definition of *Problem* set a low threshold for cognitive demand and, as a result, combined simple "exercises" (Schoenfeld, 1992) and more nonroutine and potentially demanding "problems." On the other hand, we felt that some effort to distinguish among textbook queries was necessary and worthy of a principled effort. We return to this thorny issue in the Discussion.

Applying the CCS to the Textbook Data

Typically, the coded unit of curricular content was a single sentence or clause; less frequently, it was two or more consecutive sentences. Occasionally (e.g., for *Worked Examples*) it was a short paragraph. Each such content unit was assigned a knowledge element code and a textual element code. Since most textbook pages contained numerous units of text that expressed length measurement content, the result of coding a textbook page was a list of ordered trios (knowledge element, textual element, frequency), where frequency was the number of times a particular knowledge and textual element pair were identified on that page. The number of coded content units on a given textbook page ranged widely, from one to more than 40. Pages with one or two codes were typically single problem pages; pages with large numbers of codes were usually pages in length measurement lessons.

The coding proceeded sequentially through the four grades (K–3) beginning with Kindergarten, and at each grade, the corpus of pages and codes increased. Two members of the research team coded each textbook page. The total number of pages for any grade was divided among these two-person teams (three teams for grade K, four teams for all other grades); each team received an approximately equal-size fraction of the grade corpus and an equal share of the pages from each curriculum. In dividing the textbook pages into equal-sized parts, we avoided dividing length measurement lessons between different teams. All coders either held a bachelor's degree in mathematics (or a higher degree) or were within a year of earning a bachelor's degree. The majority had experience teaching precollege or collegiate mathematics.[6] Most had multiple years of teaching experience in K–12 classrooms. Coders first coded their

6 The two undergraduate students in the final year of their teacher education program and two of the seven graduate students lacked classroom teaching experience; the other six coders had that experience.

assigned pages separately, then compared their preliminary work with their partner and resolved any disagreements. Coding disagreements that could not be resolved within the pair were presented, discussed, and resolved in discussion among the entire research team (all coders). To minimize bias due to particular pairings, the pairing of coders was systematically permuted across grades, so at each grade, coders worked with a different partner.

One analyst completed the more cursory analyses of the length measurement lessons in grades K through 3 in *EM*'s *Common Core edition* and SFAW's *enVision Math* materials. The analysis of *EM*'s *Common Core edition* (2012) involved a lesson-by-lesson comparison with their third edition (2007), carried out by an experienced undergraduate research assistant whose work was reviewed by the first author. The first author completed the analysis of the *enVision Math* materials.

Results

Our research questions address students' opportunity to learn from written curriculum at successively finer levels of detail, and our results follow that order. We first examine the proportion of lessons devoted to length measurement in grades K–3—as a crude measure of opportunity to learn. Then we draw on our more detailed analysis of knowledge and textual elements to examine the broad sequence of length measurement content presented in the three main U.S. curricula, making comparisons with the other curricula as appropriate. We do that in two steps, first with an overview by knowledge type and textual type and then by characterizing in greater detail the most common procedures and related concepts. Finally, we address the general question of how much and how directly the curricula address five challenges identified in prior research on students' learning. This analysis considers the frequency and location of specific conceptual elements and procedures that are directly related to those challenges.

Relative Attention to Length Measurement

A simple measure of students' opportunity to learn any specific content is the number of daily lessons that focus on that content. **Table 1** presents this measure of opportunity to learn length measurement, for the three main U.S. curricula and for *MPAH* and *enVision*. *EM*'s *Common Core edition* (2012) is not included because its length and total lesson counts

Table 1. Percentage of Length-Focused Lessons to Total Lessons by Curriculum and Grade

Curriculum	*Grade K*	*Grade 1*	*Grade 2*	*Grade 3*
EM	6%; 8 of 134	8%; 10 of 120	7%; 8 of 123	12%; 14 of 121
SFAW	3%; 4 of 127	6%; 9 of 157	6%; 9 of 159	10%; 16 of 162
Saxon	6%; 8 of 135	4%; 7 of 160	7%; 11 of 160	11%; 18 of 160
MPAH	N/A	8%; 5 of 66	10%; 7 of 71	13%; 8 of 72
enVision	6%; 7 of 120	4%; 5 of 130	7%; 9 of 130	11%; 12 of 110

Note. EM= *Everyday Mathematics*, SFAW= Scott-Foresman/Addison-Wesley's *Mathematics* (Michigan edition), Saxon = *Saxon Math*, MPAH= *My Pals Are Here! Math*, envision = Scott-Foresman/Addison-Wesley's *enVision Math.*

were unchanged from the third edition (2007). Each cell entry presents the percentage of length lessons of total lessons for that grade, as well as the frequencies of each.

All curricula, not only the three main U.S. curricula, were quite consistent on this measure. Length measurement lessons accounted for less than 10% of the yearly content prior to grade 3, and slightly more than 10% at grade 3. The grade 3 increase was partly due to lessons that focused on making and interpreting bar graphs (with nonunitized bars), plotting points on coordinate grids and reviewed prior length content. Some curricula also introduced fractions via the partitioning of one-dimensional quantities in grade 3. *MPAH* allotted slightly more attention to length measurement than the U.S. curricula, via a smaller number of lessons, but relative to a much smaller yearly total—roughly half of U.S. totals. Neither recent revision of two U.S. curricula (*EM* and *SFAW*) increased attention to length relative to their prior versions. In all curricula and grades, much more attention was given to base-10 number and operations.

This simple analysis also showed *where* length measurement content was placed in each year's list of lessons. In three curricula (*SFAW*, *Saxon*, and *enVision*), no length measurement lesson appeared in the first half of the year in grades K to 2. *EM* placed six of its 10 length lessons in the first half of its grade 1 content, and *MPAH* placed six of its seven length lessons just before the midpoint of its grade 2 content. But in the other primary grades, *EM* placed most, and *MPAH* placed all of their length measurement content in the second half of the year's content. In grade 3, all curricula but *MPAH* deviated from this pattern and presented length measurement content in the first third of the year.

The lesson-level analysis shows that length measurement has generally received modest curricular attention, especially in grades K through

2 when the basic concepts and procedures are presented and that length measurement lessons have been generally positioned in the second half of the year's content, after extensive work on base-10 number and operations.

Overview by Knowledge and Textual Type

To frame our results on content sequence, we address four basic questions: (a) How do curricula distribute attention over *Conceptual*, *Procedural*, and *Conventional* knowledge? (b) How do they use *Statements*, *Demonstrations*, *Worked Examples*, *Questions*, and *Problems* to express that knowledge? (c) How does textual expression vary by knowledge type? and (d) How is length content distributed between *Teacher* versus *Student* materials?

Knowledge Type. **Table 2** presents the percentage of *Conceptual*, *Procedural*, and *Conventional* codes, of total codes for the three main U.S. curricula and the Singapore curriculum in grades K–3. Recall that *MPAH* did not produce materials for grade K.

For the most part, curricular attention to length measurement content increased in each successive grade. Length measurement content of each knowledge type appeared in all curricula and grades. But the clearest

Table 2. Percentage of Length Measurement Codes by Knowledge Type, Curriculum, and Grade

		Conceptual	*Procedural*	*Conventional*	*Total*
EM	K	16.5	79.1	4.4	206
	1	11.1	79.7	9.2	404
	2	6.4	79.6	14.0	769
	3	6.8	74.6	18.6	1,425
Saxon	K	1.1	97.1	1.7	174
	1	3.5	93.0	3.5	263
	2	6.4	89.0	4.6	392
	3	1.4	80.4	18.2	858
SFAW	K	1.7	98.0	0.3	300
	1	14.8	80.2	5.0	1,013
	2	10.0	84.6	5.4	877
	3	5.6	87.2	7.1	2,002
MPAH	1	8.9	85.8	5.3	302
	2	9.7	89.0	1.3	390
	3	8.6	88.7	2.7	627

Note. Entries in the Total column are the total length measurement codes for each curriculum and grade. The *MPAH* curriculum did not include materials for grade K.

pattern in Table 2 is the dominance of *Procedural* knowledge across curricula: At least 75% of all length measurement content in all curricula and grades (and generally much more) was *Procedural*. *Conceptual* knowledge accounted for no more than 17% of that content in any curriculum or grade. It was generally less frequent in *Saxon* than the other curricula. *Conventional* knowledge appeared less frequently in the early grades, but rose significantly by Grade 3 to surpass *Conceptual* knowledge in all three U.S. curricula. *EM* gave more attention to *Conventional* knowledge in all grades, but particularly in grades 1 and 2.

Textual Expression

With one exception (*Saxon*, grade K), *Problems* were used most frequently to express length measurement content (see Table 3 for details). The relative frequency of *Problems* rarely fell below 50%; for *EM* and *SFAW* in grades 2 and 3 it was 60% or more. *Saxon* used *Questions* more frequently in all grades, especially in grades K and 1. Taken together, *Problems* and *Questions* accounted for at least 70% of all codes in all curricula and grades. Each curriculum used *Demonstrations*, especially in grade K, but *Saxon* used them more frequently than the other curricula in later grades. *Saxon* was the only curriculum not to use *Worked Examples*. All curricula included *Statements* at each grade, but *EM* did so more frequently at each grade. In grades 1 and 3 *EM* included twice as many *Statements* as the other curricula. In comparison, *MPAH* generally used even greater proportions of *Problems* (between 65% and 82% of all codes in all three grades) and more *Worked Examples* than U.S. curricula.

Knowledge Type by Textual Expression. How curricula present length measurement knowledge in their materials matters to students' opportunities to learn. **Table 3** presents the cross-tabulation of knowledge type by textual expression: Within each knowledge type, how often were the five textual forms used to express length measurement knowledge? Textual types are ordered in columns by frequency (e.g., P for *Problems*, Q for *Questions*); the entries in each row sum to 100%. Because of the density of information in Table 3, we have left the essentially zero frequency cells (≤0.5% of total codes) blank for ease of viewing. Since the percentages are relative frequencies computed from different numbers of total codes, care must be exercised in comparing corresponding values across curricula or grade.

Despite the density of information in Table 3, some important patterns are evident. The presentation of *Procedural* knowledge was quite

Table 3. Percentage of Length Measurement Codes by Knowledge Type, Textual Type, Curriculum, and Grade

		Conceptual					*Procedural*					*Conventional*				
		P	*Q*	*S*	*D*	*W*	*P*	*Q*	*S*	*D*	*W*	*P*	*Q*	*S*	*D*	*W*
EM	K	1	10	4	1		51	10	1	16				4		
	1	1	5	5			51	16	3	8	2	1	1	8		
	2	3	2	3			64	9	1	2	1	1	2	11	1	
	3	1	2	3			57	8	2	2	3	1	2	17		
Saxon	K			1			33	44	4	16			1	1		
	1		1	2			43	39	1	11			2	1		
	2	1	2	3			62	18	3	8				4		
	3						51	24	2	5		1	7	7	3	
SFAW	K	1	1				47	28	3	19	1					
	1	8	6	2			49	16	1	7	4		1	4		
	2	6	2	4			69	5	1	4	3			5		
	3	4	2	1			69	10	1	1	4	1	2	6		
MPAH	1	6	1	2			59	14	3	7	4			5		
	2	6	2	1		1	68	7	4	4	5			1		
	3	7		1			75	5	2	1	5			3		

Note. P = *Problems*, Q = *Questions*, S = *Statements*, D = *Demonstrations*, W = *Worked Examples.* The *MPAH* curriculum did not include materials for grade K.

uniform across curricula; all curricula most frequently expressed *Procedural* knowledge in *Problems* and *Questions*. In most curricula and grades *Problems* were more frequent (especially so for *MPAH*); in *Saxon* grade K, *Questions* were the dominant form. *Conceptual* knowledge, by contrast, was most frequently expressed in *Statements*, *Questions*, or *Problems*. If *Questions* and *Problems* are combined as "queries," *EM* and *SFAW* asked about *Conceptual* knowledge more often than they stated it. But for both curricula, *Conceptual Questions* and *Problems* repeatedly referenced a small number of knowledge elements. For example, in *EM,* grade K, 18 of 21 *Conceptual Questions* were posed about only three knowledge elements—*Unit-Measure Compensation*, *Additive Composition of Lengths*, and *Greater Means Longer*. In *SFAW*, grade 1, 67 of 70 *Conceptual Problems* and 49 of 53 *Conceptual Questions* addressed *Greater Means Longer*, *Unit-Measure Compensation*, and *Rulers Measure Length*. Within the trio (P, Q, S), *EM* and *Saxon* presented more *Statements* and *Questions* than *Problems*, and *SFAW* presented more *Problems* and *Questions* than *Statements*. These differences become more significant given the location of *Conceptual* knowledge in *Student* vs. *Teacher* materials (see below). The presentation of *Conventional* knowledge was also relatively uniform across curricula and grades. In most curricula/grades *Statements* dominated the expression of length measurement conventions.

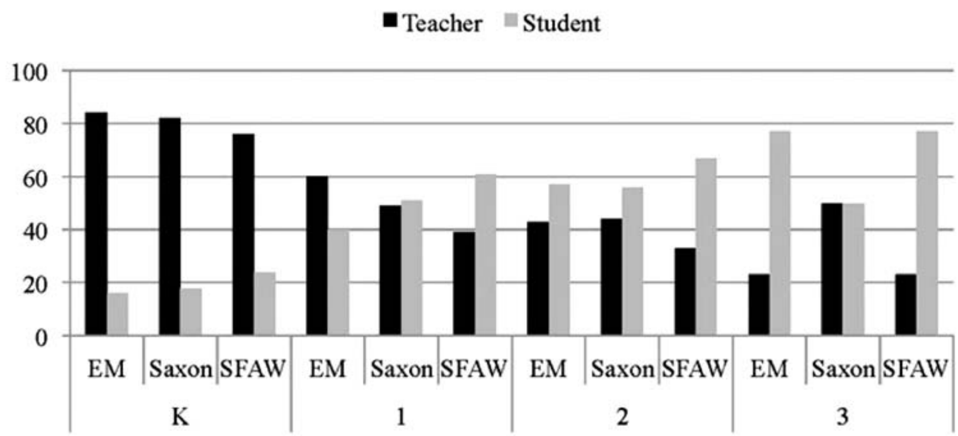

Figure 3. Percentage of length measurement codes in *Teacher* and *Student* materials by curriculum and grade.

Location in Student Versus Teacher Materials

Figure 3 shows how frequently *Problems*, *Questions*, and *Statements* were presented in *Teacher* and *Student* materials. It excludes *Worked Examples* and *Demonstrations* because, by definition, the former appears only on student pages and the latter only on teacher pages.

Generally speaking, the proportion of length measurement content in *Teacher* materials in U.S. curricula decreased across the grades. In later grades, more content appeared in *Student* materials, indicating that older students have more direct access to that content, without the mediation of teachers. The pattern in *Saxon* was again different; in grades 1 and 3 about the same amount of content appeared in *Teacher* and *Student* materials. In contrast, over 80% of *MPAH*'s length content in all three grades appeared in *Student* materials, though the proportion was greatest in grade 3. So students had consistently greater direct access to length content in *MPAH* than in the U.S. curricula.

But Figure 3 does not show how the three most frequent textual types, *Statements*, *Questions*, and *Problems*, separately appeared in *Student* vs. *Teacher* materials. This distinction proved important for *Statements*. In all but two curricula/grades (*EM* and *SFAW*, Grade 3), *Statements* appeared more often in *Teacher* materials than *Student* materials. In particular, *Statements* of *Conceptual* knowledge appeared more frequently in *Teacher* materials in most grades, with the exception of grade 3. In *Saxon*, all *Conceptual Statements* in all grades appeared in *Teacher* materials. This result is important because it shows that students' direct access to *Statements* of *Conceptual Knowledge* was limited. For students to have

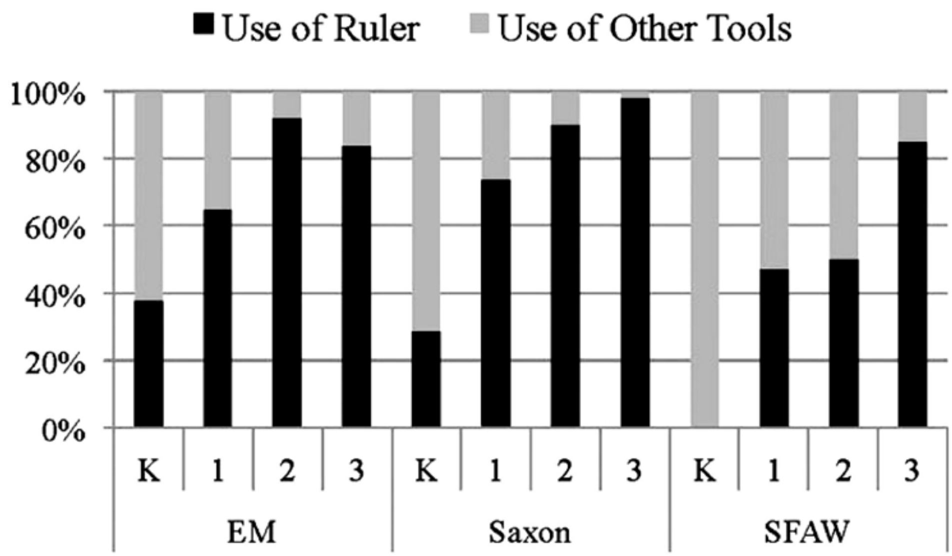

Figure 4. Percentages of codes for ruler procedures and procedures using other tools for measuring simple paths.

access to explicit statements of conceptual principles for length measurement in grades K–2, teachers needed to voice those statements. *MPAH* was quite similar to U.S. curricula: *Conceptual Statements* appeared more often or entirely in *Teacher* materials in grades 1 to 3.

The Sequence of Procedures and Associated Concepts

Table 2 showed the general dominance of procedural content in all grades and curricula, but not surprisingly, the curricula focused on different types of length measurement procedures at different points across the four grades. To understand how procedures and related concepts were sequenced and therefore how length measurement was presented across the grades, we aggregated procedures into five groups: *Qualitative Judgments*, *Measurement of Simple Paths* (using nonstandard units or rulers), *Visual Estimation*, *Sums and Differences* (including perimeter and word problems), and *Multiplicative Relationships* (including unit conversion and distance-speed-time relationships). Procedures in each of these groups engaged similar reasoning. *Qualitative Judgments* involved only comparative judgments of longer, shorter, and same lengths. *Measurement of Simple Paths* included procedures that used nonstandard or standard units to measure or draw straight-line paths along segments or objects, and all procedures for using rulers. *Visual Estimation* involved the visual projection of units to estimate the length of objects without using physical units or tools.

Table 4. Percentage of Length Measurement Procedural Codes by Procedural Group, Curriculum, and Grade

		Procedural Group					
Curr	*Gr*	*Qualitative Judgments*	*Simple Paths*	*Visual Estimation*	*Sums & Differences*	*Multiplicative Relations*	*Other*
EM	K	29	63	2	2		4
	1	27	33	16	7	5	12
	2	13	43	5	15	13	11
	3	10	43	7	13	19	8
Saxon	K	63	34	2			1
	1	52	45	2			1
	2	8	70	5	7		10
	3	13	63	5	11	4	4
SFAW	K	68	18	10	1		3
	1	18	45	18	13		6
	2	10	46	17	21	1	5
	3	14	29	5	16	16	20

Table 4 shows the percentage of *Procedural* content by group across curricula and grades. The *Other* group represents all residual procedural content. We have left zero cell entries blank for ease of viewing.

These five groupings collectively accounted for most of the procedural content in all curricula and grades—at least 80% of all procedural codes, and generally more. Therefore they effectively "covered" most all of the dominant procedural content reported in Table 2.

We now examine these entries from the perspective of sequence, here with a broad brush and below in more detail: How did attention to these procedures flow across grades? *Qualitative Judgment* procedures were most frequent in grades K and 1. *EM* gave attention less to these procedures than the other curricula; *SFAW*'s attention was strongest in grade K, but it dropped dramatically in subsequent grades. Procedures for *Measuring Simple Paths* appeared frequently in all four grades, but particularly so in grades 1 and 2. The relatively high frequency of these procedures in grade 3 was due to measuring one or more segments as the first step in computing the perimeter of polygons. Taken together, *Qualitative Judgment* and *Measurement of Simple Path* procedures made up a very large part of the length measurement content; in only one curriculum-grade (*SFAW*, grade 3) did the combined percentage fall below 50%. Attention to the task of *Visual Estimation* was less consistent across curricula. *EM* focused on estimation in grade 1 only. Saxon gave it much less attention overall, where *SFAW* gave it significant attention in all but grade 3. Procedures for finding *Sums and Differences* and for reasoning with *Multiplicative Relationships* were more frequent in grades 2 and 3, with the latter primarily a focus in grade 3.

The content of Table 4 shows very substantial, but not complete, agreement in the sequence in which length measurement procedures are presented to students in the U.S. curricula. All three focused initially on qualitative comparisons before introducing and using length units, both nonstandard and standard. Using those units to measure simple paths became a major focus in all grades and curricula, but especially in grades 1 and 2. In part at least, the results cohere with Piaget's proposed development sequence: Curricula have accepted the premise that comparative judgments should precede metric determinations of length. We now examine these procedures and the concepts closely related to them in greater detail.

Qualitative Judgment

We found seven procedures that involved reasoning about length in purely qualitative terms. In *Direct Comparison*, two objects are placed side by side to determine if one is longer by visual inspection; *Visual Comparison* involves the same judgment when objects are not physically aligned. *Indirect Comparison* uses a third intermediate object and direct comparison of two objects to determine if one of the pair is longer. We also included procedures for partitioning lengths into equal parts. Finally, we included procedures for drawing segments equal to, longer, or shorter than another.

Direct Comparison was the dominant procedure in this group. Where *Direct* and *Visual Comparison* together accounted for most of the *Qualitative Judgment* procedures in all curricula and grades, in most grades *Direct Comparison* appeared far more often. It was the most frequent procedure in all curricula in grade K—and in *EM* and *Saxon* in grade 1 as well. In grades K and 1, these two procedures were directed at pairs of objects present in the classroom or drawn in the text. But beginning in grade 1, *Direct Comparison* was also used in early work on bar graphs when students were asked to compare the heights of bars. Instances of *Indirect Comparison* and *Draw Segment Shorter/Longer/Equal* were infrequent. The lone exception was *Saxon* grade K, where *Indirect Comparison* accounted for about 9% of procedural content at that grade, and *Draw Segment Longer/Shorter/Equal* for about 8%.

Qualitative judgment procedures make minimal conceptual demands beyond an intuitive understanding of the length attribute. Three conceptual elements in our coding scheme were directly related to qualitative length judgments: *Definition of Length*, *Conservation Under Motion, and*

Conservation Under Partitioning. We discuss their frequency and variation (for definitions) below when we consider how the curricula faced the challenge of defining length as a measurable quantity.

Measurement of Simple Paths

Procedures for measuring simple paths—the length of single segments, objects, and distances—were more common than any other group in most all curricula and grades (Table 4). These procedures represent the heart of length measurement—choosing, placing, and enumerating a number of length units in a given linear space. This group combined two subgroups: Procedures for placing and interpreting nonstandard length units and procedures for applying standard units, typically measurement with rulers. Both include procedures for measuring simple paths and drawing segments whose length is a given number of units, though the former were much more frequent than the latter. *Unit Iteration* is the primary conceptual foundation for procedures for measuring simple paths. As indicated earlier, we stated *Unit Iteration* as, "Measures of length are produced by iterating a length unit (repeatedly adjoining) from one end of an object, segment, or distance to the other and then enumerating the number of iterations (e.g., by counting). Iterated units may not overlap or leave gaps."

Although this statement implicates a procedure for placing and enumerating physical units to produce length measures, it is not stated in procedural terms, and we carefully distinguished references to the procedures for placing units from references to constraints on the placement of units. We include specific examples that illustrate this distinction below.

Measuring Simple Paths With Nonstandard Units. All three U.S. curricula first addressed length measurement with nonstandard units (e.g., paper clips, square tiles, and linking cubes) before introducing standard units and rulers. Measuring simple paths with nonstandard units involves placing units along a path adjacent to the object or distance to be measured from one end to the other so that the linear space is filled without leaving gaps or overlaps between units—that is, it involves the physical enactment of *Unit Iteration*. We distinguished two such procedures, one where sufficient numbers of units are available and one where the supply is insufficient so some units must be reused (moved). The first procedure produces *tilings* of paths where the measured space is completely filled by the end of the procedure (Lehrer, 2003). The second involves *iterating* a unit one or more times to achieve a *tiling*. With iteration, the measurer must reuse some physical units one or more times and keep track of the total number of units placed (e.g., by counting each placement or marking the location of prior placements) because the whole space is never completely filled at any point in the procedure.

All three curricula included situations where sufficient numbers of nonstandard units were available for measuring simple paths and those where they were not, but with different order and emphasis. *EM* provided more situations involving insufficient units in grades K and 1 before shifting toward sufficient situations in the later grades. The pattern for *SFAW* and *Saxon* was opposite: More situations supplied sufficient numbers of units in grades K, 1, and 2, where situations requiring iteration were slightly more common in grade 3. Across grades, *EM* roughly balanced the number of situations of each type, where sufficient situations were more common in *Saxon* and *SFAW*.

Both procedures depend on *Unit Iteration*, though our statement does not explicitly distinguish tiling from iterating. Overall, *Unit Iteration* was infrequently expressed; we found only 21 instances in all three curricula and grades. Of these, 11 were located in grade 1 materials. *EM* materials contained half of all instances (n = 11), and the other two curricula about a quarter each (*Saxon*, n = 6; *SFAW*, n = 4). Placement in grade 1 was generally consistent with the appearance of procedures for using nonstandard units, though *EM* presented situations with insufficient units in grade K. More problematic was the fact that half of all instances of *Unit Iteration* (n=10) were *Partial Statements*—most emphasizing the need to avoid gaps and overlaps between units. All *Statements* (10 *Partial* and 2 *Full*) were located in *Teacher* materials. Only three instances of *Unit Iteration* were located in *Student* materials; these were *Problems* that drew students' attention to the conceptual dimensions of unit placement. In sum, the opportunity to learn the principle of *Unit Iteration*, especially in the *Student* materials, was very limited.

Measuring Simple Paths With Rulers. All curricula moved quickly from qualitative comparisons and simple path measurement with nonstandard units to introduce rulers to measure the same simple paths in whole numbers of inch or centimeter units. *EM* and *Saxon* introduced ruler measurement in grade K; *SFAW* did so in grade 1. Throughout the grades the curricula generally interleafed the use of rulers and nonstandard units in measuring simple paths. As Figure 4 shows, ruler use generally increased across the grades and dominated the measurement of simple paths by grade 3. "Other tools" included nonstandard units and materials that supported the drawing of segments (e.g., dot paper and grid paper).

In measuring simple paths with rulers, all three curricula presented paths that were shorter than the ruler with some that were longer. Paths that were longer require students to iterate the ruler at least once to complete their measurement work. Most instances of measuring longer paths were located in grades 1 and 2; their frequencies were about one quarter of the corresponding frequencies of measuring shorter paths. *EM* addressed both tasks in grade K with about equal frequency (n = 14 [longer paths] to n = 18). Also, *EM* and *SFAW* provided a small number of opportunities for measuring objects that were not aligned with the zero mark on the ruler (n = 15 and n = 3, respectively). All were *Problems* or *Questions*. No U.S. curricula stated a procedure for measuring with a ruler when the object was not aligned at the zero mark, and nine of *EM*'s 15 instances were *Problems* on a single grade 3 page. *Saxon* materials did not address this situation.

Although ruler use was widespread, support for understanding how rulers represent a sequence of length units was not. We coded three *conceptual* knowledge elements that provide meaning for rulers and their use: (a) The simple idea that rulers measure lengths (*Rulers Measure Length*), (b) the notion that marks on rulers represent units of length (*Rulers Represent Iterated Units*), and (c) the notion that any point on the ruler can serve as the zero point (*Zero/Scale on Rulers*). The latter concept supports use of the broken ruler procedure, where counting starts at a nonzero ruler mark adjacent to one end of the path. All curricula addressed each of these ideas at least once in the four grades. But there were only six instances of *Rulers Represent Iterated Units*, all in *Teacher* materials, half of which appeared in grade 3, long after all curricula had asked students to use rulers to measure lengths. Likewise, all six instances of *Zero/Scale on Rulers* were found in grade 3; three were *Partial Statements*. More frequent attention was given to *Rulers Measure Length*, and its placement—most frequently in grade 1—was consistent with the appearance of ruler procedures. That said, both *EM* and *Saxon* introduced rulers in Grade K. In all three cases, the appearance of relevant *Conceptual* knowledge lagged behind calls to use procedures justified by that knowledge.

We found a similar trend with two *Conventional* knowledge elements related to rulers. The first concerned the construction of most rulers used in schools where the customary (inch) and metric (centimeter) scales appear on opposite sides or opposite edges of the ruler. The second

Table 5. Frequency of Conventional Knowledge Codes Related to Ruler Construction by Curriculum and Grade

Curriculum	*EM*				*Saxon*				*SFAW*			
Knowledge Element/Grade	K	1	2	3	K	1	2	3	K	1	2	3
Basic Ruler Construction	2	3	1	12	2		6	15	1		2	5
Units/Subunits on Rulers		3	12	22			3	7			1	3

concerned the convention that larger marks indicate whole units of length and smaller marks indicate subunits.[7] **Table 5** presents the frequency of these *Conventional* knowledge elements across curricula and grades. As before, we have left zero cells entries blank.

Although all curricula addressed *Basic Ruler Construction* at least once in grades K and 1 when students are first asked to use these tools, most instances (64%) appeared late, in grade 3. Similarly, the vast majority of instances of *Units/Subunits on Rulers* appeared in grade 2 and 3. This placement matches the curricula's attention to greater precision in ruler measurement in the later grades (i.e., to the nearest 1/4 and 1/8 inch). However, since the same rulers were presumably used in grades K and 1, any questions students might have about the nature of these marks were left unaddressed in the curricula for a long period.

Visual Estimation

In contrast to the other groups, *Visual Estimation* was a single procedure that called for selecting a length unit, nonstandard or standard, and iterating it visually to produce an approximate measure of the length of an object, segment, or distance. As shown in Table 4, *Visual Estimation* appeared in grade K in all curricula and continued at each grade. The procedure addressed two types of situations: calls for approximate length measures for given objects, segments, or distances and the generation or location of objects, segments, or distances approximately equal to a given length. In typical instances of the first type, the text depicted the object (e.g., a pencil or ribbon) or described a distance such as "the distance

7 This conventional issue is different than the conceptual issue of how smaller subunits nest within larger units. The *Conventional Knowledge* element refers to the marks on rulers as indicators of whole length units and subunits.

from your classroom to the library." Sometimes the target attribute and/or length unit was left unspecified (Chang et al., 2011). A typical instance of the latter case was "Find an object that is 1 meter long."

Generally, *Visual Estimation* appeared more frequently in *EM* and *SFAW* than in *Saxon*. *SFAW* addressed estimation more frequently in all grades, at least doubling the number of instances in *EM* grades K through 2 and in *Saxon* in all grades. It was the most frequent single procedure in *SFAW* grades 1 and 2. Frequently, *SFAW*'s calls to estimate followed a fixed pattern: An object was depicted in the text and students were asked first to "estimate" and then to "measure" the object, using the same unit. In all curricula, *Visual Estimation* was predominantly presented in *Problems* or *Questions*. Despite the frequent calls to estimate lengths, little attention was given in any curriculum to specifying the estimation process, either in *Student* or *Teacher* materials. *EM* grade 1 included one noted exception: Estimation was introduced and the process of using a reference object to obtain an estimate was discussed in the *Teacher* materials. However, that description appeared after students had been asked to estimate lengths numerous times. A similar concern applied to *SFAW*, where students were frequently asked to estimate with nonstandard units in grade K, prior to the definition of nonstandard units.

Sums and Differences

All three U.S. curricula also presented opportunities for students to work with additive combinations of two or more lengths. Additive relationships appeared in two main forms: word problems that asked students to determine sums or differences of lengths and computations of (or from) problems involving lengths. These were frequent in grades 2 a the perimeter of polygons. All three curricula presented additive word nd 3, especially in *EM* and *SFAW* (Table 4). *Saxon*'s frequencies were lower—5% of procedural codes at grade 3.We found four different forms of word problems, those presented in (a) words only; (b) words with the objects represented (e.g., two pencils); (c) words with lengths represented in units; and (d) words and numerical representations (e.g., tables of values). Types (a) and (b) were most frequent in all curricula. Their location in the text and their constituent numerical values suggested that their role was "applied" practice for whole number addition as much as for learning about length.

Three additive procedures involved perimeter, the most frequent by far was *Find Perimeter by Adding Lengths*. It first appeared in grade 1 in

SFAW, and by grades 2 and 3 it was relatively frequent in all curricula (≥5% of procedural content in that grade). *Draw a Figure with a Given Perimeter* was infrequent, appearing only in *EM* and *SFAW*, grade 3, and in most of those cases (88%), some structured space (e.g., dot paper) facilitated students' construction of the appropriate polygon. *EM* and *SFAW* provided occasions for *Determining the Side Length of a Regular Polygon*, but very infrequently and generally in grade 3. *Saxon* did not present this procedure.

The conceptual principles most directly related to additive relationships were again sparse in the curricula. Of four *Conceptual* elements, only one (*Definition of Perimeter*) appeared in all three curricula. *EM* and *SFAW* repeatedly defined perimeter, starting in grade 1 and with increasing frequency at each successive grade. *Saxon* defined perimeter only at grade 3 but did so repeatedly ($n = 6$ times). All three curricula asked students to determine perimeter before the term was explicitly defined. Only *EM* addressed the principle of *Additive Composition*—that sums and differences of lengths must themselves be lengths—and did so early; seven of eight instances appeared in grade K. No curriculum addressed the numerical extension of that idea (*Numerical Sums/Differences*) that sums and differences can be computed only when the lengths are measured in the same units. Finally, *EM* and *SFAW* addressed the issue that *Perimeter is Not Area*, primarily in grades 2 and 3 when perimeter computations were frequent. But they did so infrequently ($n = 14$ overall; $n = 7$ in grade 3); Saxon never did. However, when *EM* and *SFAW* did address the issue in grades 1 and 2, the discussion was in close proximity to work on perimeter computations.

Multiplicative Relationships

Relative to additive relationships, multiplicative relationships are conceptually diverse (Schwartz, 1988; Thompson, 1994b; Vergnaud, 1994), and this diversity extends to relationships involving lengths. One type involves scalar multiplication and multiplicative comparisons of lengths, which produce ratios. There are also products of a scalar and a length (e.g., "three times as long as this length") that produce a second length that is longer or shorter. Two lengths can also be compared multiplicatively (e.g., "how long is this length relative to this length?"). As long as the two lengths are measured in the same unit, the resulting ratio is a scalar (e.g., "three times as long"). More generally, multiplicative relationships exchange or equate length measures for other quantities, including other length measures; these relationships are often called rates (Thompson, 1994b). In length measurement, these exchanges include unit conversions in both customary and metric systems (e.g., “two feet is the same as 24 inches”). They also include exchanges of lengths for time in distance-speed-time relationships. Finally, lengths can be composed multiplicatively with other quantities to produce new spatial quantities. Most commonly in elementary mathematics, the product of two (or three) length measures produces area (or volume) measures.

By grade 3, all three curricula asked students to convert lengths measured in one unit into another unit (*Unit Conversion*). *EM* and *SFAW* presented this procedure in grade 2, and by grade 3 these frequencies were significant (13% of all procedural codes in *EM* and 12% in *SFAW*). *Saxon* gave limited attention to this procedure (only 1.4% in grade 3). All curricula also presented the procedure for producing a length from a given length and a ratio relating the two (between 2% and 3% of grade 3 procedural content in all three curricula). Beyond those two procedures, *EM* and *SFAW* gave some attention to determining a ratio from two lengths and to finding a missing quantity when two of three terms in a distance–speed–time trio was given, primarily at grade 3. *Saxon* presented neither procedure. In sum, all curricula gave some attention to the procedures involving multiplicative relationships among lengths, mostly at grade 3 and with greatest attention to *Unit Conversion*.

Conceptual support for understanding multiplicative relationships was once again sparse. Of seven *Conceptual* elements in our coding scheme, we found instances of only two. *Units Can Be Converted* appeared in *EM* and *Saxon* at grade 3.[8] *EM* also addressed the issue that the circumference is a ratio of the circle’s radius at grade 3; no other curriculum did so. But the fact that multiplying lengths by any other quantity (length or not) produces quantities that are not lengths (*Multiplicative Composition*) was never mentioned, nor was the fact that a length compared multiplicatively with another length produces a ratio (or scalar), not a length. More generally, no attention was given to unpacking how ratios as scalars and lengths interact multiplicatively. In sum, there were few direct opportunities to explore conceptual diversity of multiplicative relationships involving length in grade K through 3.

8 Our coding scheme distinguished the concept that units can be converted from the procedure for carrying out such conversions. The concept that all length units can be converted to any other length unit was stated in length terms, but this principle applies generally to all quantities.

Specific Conceptual Challenges

We now turn to the specific question of how frequently and directly the curricula addressed particular challenges that students face in learning length measurement. In each case, we consider the attention given to particular concepts and procedures that address those challenges.

Understanding Length as a Stable and Measurable Quantity

Piaget's early work indicated that, for young children at least, length may not be a stable attribute of objects, that displacing one of two equal length objects can make them unequal (nonconservation), and that bent and straight paths with aligned endpoints can be seen as equal in length (Piaget et al., 1960). More generally, transitivity and length conservation were proposed as conceptual prerequisites to understanding units of length (and measurement via units). To assess how the curricula addressed these issues, we aggregated all instances of 11 *Conceptual* knowledge elements that addressed either explicit attempts to define length or conceptual properties of length as a quantity that did not require the identification or understanding of units. Their frequencies are presented in **Table 6** by grade, not by curriculum (for reasons of space). As above, we have left grade cells with zero frequency blank. Significant differences between curricula, where they exist, are noted below. We included knowledge elements for *Varieties of Paths* and *Relation to Distance* to code any attention the curricula gave to distinguishing different types of paths (e.g., simple, curved, bent) and distinguishing or relating length and distance.

Table 6. Frequency of Conceptual Knowledge Codes for Definition and Basic Properties of Length

		Grade				
Group	*Element*	*K*	*1*	*2*	*3*	*Total*
Definition	Definition of Length		2	13	3	18
	Varieties of Paths					0
	Relation to Distance			1	1	2
Basic Properties	Positive Values		1			1
	Order/Equality	2				2
	Conservation under Partitioning		2	1	8	11
	Conservation under Motion	2			8	10
	Transitivity		1	5		6
	Trichotomy				1	1
	Additive Composition	7	1			8
	Measurement involves Error	3	2	1	4	10

All three curricula explicitly defined length. *EM* and *SFAW* did so repeatedly (n = 9 and n = 8 times, respectively); *Saxon* did so only once. But these definitions appeared well after students were asked to compare and measure lengths. *EM* and *Saxon* defined length once in grade 1; *SFAW* did not do so until grade 2, where all its instances appeared. Moreover, the vast majority of *Definition* codes (16 of 18) appeared in grades 2 and 3. No curriculum addressed the issue of different types of linear paths; and the relation between length and distance was essentially not addressed (one instance each in *EM* and *SFAW*). This pattern held generally for basic properties as well; they were infrequently addressed overall. When they were addressed (e.g., *Conservation Under Partitioning* and *Conservation Under Motion*), the instances appeared in later grades. All instances of *Additive Composition*, an exception to this pattern, appeared in *EM*; seven of the 10 instances of *Measurement Involves Error*, also appeared in *EM* materials. *Saxon* never addressed either property.

While the frequency and location of *Definitions* are relevant to our question, so is the content of those definitions. **Table 7** presents five representative examples quoted from the curricula, including the single *Saxon* instance and two instances each from *EM* and *SFAW*. We coded the *Saxon* definition as a *Partial Statement* because it did not directly address the nature of the length quantity. All five instances appeared in *Teacher* materials. Italics indicate emphasis in the text.

These examples suggest that authors struggled with the task of defining length. Four instances made reference to objects but did not clarify how to locate the length attribute (or "the ends") of objects—a nontrivial issue for some objects (e.g., shoes or feet). The first *SFAW* grade 2 definition could be interpreted as suggesting that length and height are

Table 7. Examples of Length Definitions

Curriculum, Grade	*Statement in the Text*	*Textual Code*
EM, 1	*Length* is the size of something from one end to the other.	*Full Statement*, *Teacher*
EM, 3	Remind children that the measure of a distance between two points is called *length*.	*Full Statement*, *Teacher*
Saxon, 1	Sometimes we need to tell someone how long or wide something is.	*Partial Statement*, *Teacher*
SFAW, 2	*Length* tells how long an object is. *Height* tells how tall an object is.	*Full Statement*, *Teacher*
SFAW, 2	*length* How long something is from one end to the other. (Statement in vocabulary review)	*Full Statement*, *Teacher*

different spatial quantities, without clarifying that heights are lengths. Only the *EM* grade 3 statement was relatively well formed mathematically and resembled definitions we have found elsewhere (e.g., in dictionaries of mathematics and mathematic texts written for elementary teachers). But it appeared late in students' work on length. All definitions are consistent with length along a simple path (connecting two points); extension to more complex paths (e.g., jagged, polygonal, and curvilinear) was not explicitly addressed.

Understanding the Conceptual Properties of Units

Prior research has shown that the appropriate use of length units requires understanding numerous properties and constraints on their use as well as the ability to coordinate them in measurement activity. In learning to measure length, students may satisfy some properties and constraints in a particular effort while violating others. Some will successfully tile with units before they can iterate them. Eight *Conceptual* elements in our coding scheme addressed the properties of length units; **Table 8** presents their frequencies across grades.

Compared to definitions and basic properties, curricular attention to unit properties was markedly greater but very uneven. Instances of *Greater Means Longer*, the basic idea that greater numerical measures (that is, counts of units) indicate longer lengths, accounted for 37% of all *Conceptual* content. *EM* and *SFAW* addressed this idea in all grades, but *SFAW* did so much more, accounting for 73% of all instances. *Saxon* only addressed it in grade 2. *EM* and *SFAW* also addressed *Unit-Measure Compensation* in all grades; *SFAW* did so most frequently (70% of all instances). *Saxon* did so less frequently and only in grades 2 and 3. Other concepts received markedly less attention.

Table 8. Frequency of Conceptual Knowledge Codes for Length Units by Grade

		Grade				
Group	*Element*	*K*	*1*	*2*	*3*	*Total*
Units	Greater Means Longer	5	74	59	94	232
	Unit-Measure Compensation	11	56	26	11	104
	Numerical Sums/Differences					0
	Meaning of Length Measure	1	1	1	4	7
	Standard vs. Nonstandard Units	2	13	4	4	23
	Unit Iteration	4	11	4	2	21
	Length Measure Requires Length Units		1	8	7	16
	Units Can Be Converted				5	5

Our statement of *Unit Iteration* expressed multiple conceptual properties of units (the reuse of units, exhaustion, and avoidance of gaps). As noted above, half of the observed instances were *Partial Statements* that primarily focused on avoiding gaps and overlaps between units. Only 4 of 21 instances explicitly referenced the need for units to be identical; only three addressed the need for unit placement along a path parallel to the measured object, although 10 did address the need to exhaust the entire linear space via end-to-end placement of units. All curricula also presented some situations where student had to iterate their rulers to measure paths that were longer than their rulers. These could have been contexts for students to learn the *Unit Iteration* principle by interpreting their rulers as length units. But longer paths were presented far less often than shorter paths, and none of those occasions merited a *Unit Iteration* code in addition to the procedural code.

One instance of *Unit Iteration* merits specific mention (see also Dietiker, Smith, & Gonulates, 2011). A grade 1 *EM* assessment task presented four different "measurements" of the width of a sheet of paper using buttons—that is, pictures of four different placements of buttons across the sheet. One was the correct tiling that filled space along a path straight across the paper; the other three each violated some constraint in *Unit Iteration*, either by overlapping units, leaving gaps between units, or placing units along a diagonal path. Students were asked to choose and justify the "best measurement" (though they were not asked to explain why the other measurements were deficient). This was the only coded instance of *Unit Iteration* that asked students to choose among different placements of units and therefore to consider the conceptual properties and constraints of units on correct placement.

Understanding the Structure of Rulers

Prior research has shown that students can learn to use rulers in a rote manner, without understanding how marks indicate equal intervals of linear space. When the zero mark/end of the ruler has been removed and/or the objects to be measured are aligned at nonzero unit marks, students may read the length off the ruler incorrectly. As stated above, attention to the relationship between ruler marks and length units (*Rulers Represented Iterated Units*) was infrequent and late in appearance (half of the $n = 6$ instances appeared in grade 3). All $n = 6$ instances of *Zero/Scale on Rulers* appeared in grade 3. Few situations were provided to measure objects not located at the zero point. The infrequent attention to these conceptual and procedural issues and the early introduction of rulers in grades K and 1 may indicate that the curriculum authors did not see major challenges for students in understanding the structure of rulers.

Measuring Complex Paths

Research has also shown that children can face difficulty in discriminating between the lengths of simple and bent paths when their endpoints align and in reasoning about the length of complex paths that combine multiple line segments, especially those drawn on grids. Some count squares, including those on the outside corners of complex paths, rather than the sides of squares that compose the linear path. Our coding scheme distinguished perimeter measurement for nonpolygonal shapes (*Measure Perimeter With Flexible Tools)* from perimeter measurement of polygons and coded the measurement of nonclosed paths (*Measure Complex or Curved Paths*). Neither procedure appeared frequently, especially compared to the very extensive work on simple paths and the perimeter of polygons. *SFAW* (Grade K) and *EM* (Grade 1) began work with perimeter with tools like string before shifting attention to polygons. Complex nonclosed paths rarely appeared. *EM* and *SFAW* presented these situations only six times across all grades; *Saxon* never did. Students certainly had many occasions to reason about the length of nonsimple paths in computing the perimeter of polygons and often those polygons were presented on grids. Where those situations could help students sort out how to handle the corners of bent paths (closed and open), the geometric properties of squares and rectangle also provide some support for avoiding or correcting the corner error. These supports are not available for complex open paths.

Discussion

We have analyzed students' opportunity to learn length measurement from written curricula, targeting knowledge and its expression in a fine-grained way. We have characterized the sequence/ flow of that content from Kindergarten through grade 3 and have specifically examined how the curricula addressed known challenges for students in learning length measurement. Because of the length and detail of our results, we first summarize them. Then we return to the central focus that framed the study: What lessons has the analysis produced for understanding the roots of the well-documented problem of weak learning of measurement? We close with suggestions for curriculum development (especially the revision of current materials) and consider both the study's limitations and future steps for research.

In all curricula we examined, thoroughly and more cursorily, we found the broad topic of length measurement received modest attention, especially in grades K through 2 when the most fundamental concepts and procedures are presented. We also found no change in attention to the topic in the two more recent editions of U.S. curricula relative to previous editions. Attention to length measurement was overshadowed in all curricula by work on base-10 number and operations, both in the number of lessons and placement in the text. Length content was generally placed in the second half of elementary curricula after extensive number and operations content in the first half. This basic result may or may not be problematic for student learning, but it raises at least one major cause for concern. Given the importance of learning fractions—for their own importance and for subsequent mathematics and science—the emphasis on work with discrete quantities (counting sets of objects, base-10 numeration, place value and arithmetic operations) may not serve as sufficient foundation for learning fractions and rational numbers. Rather, the partition of continuous quantities, like length and other spatial quantities, seems a more productive foundation for understanding fractions and operations on fractions (Freudenthal, 1983; Thompson & Saldanha, 2003). If understanding fractions is as important as mathematics educators have argued, then some rethinking of the curricular attention and location of spatial measurement content (beginning with length) may be in order.

With respect to length measurement content presented in the four grades, we found strong commonalities among the curricula. All four curricula examined in detail showed a strong procedural focus. This characterization of written curricula matches those made by numerous researchers of typical classroom instruction (Bragg & Outhred, 2001; Hiebert, 1984; Lehrer, 2003; Schifter & Szymaszek, 2003; Stephan & Clements, 2003; Van de Walle, 1994). *Procedural* knowledge dominated all curricula in all grades (at least 75% of all length content); attention to *Conceptual* knowledge was modest—especially when the frequencies of two concepts (*Greater Means Longer* and *Unit-Measure Compensation*) were set aside; *Conceptual* knowledge very often appeared after work on related procedures began; and *Conventional* knowledge was similarly modest before grade 3 with work on unit conversion. *Conceptual* knowledge, especially in the early grades, more frequently appeared in the curricula's *Teacher* materials, thus limiting students' direct access to that content. In sum, students' access to the conceptual principles of length measurement was limited in three main ways—in the low frequency of many central elements, in their location in the text relative to associated procedures, and in the need for the teacher to voice that content to students. The pattern of placing *Conceptual* knowledge after repeated references to related *Procedural* knowledge suggests that curriculum authors may believe that conceptual knowledge is of limited importance or that students can usefully carry out measurement procedures before they learn the principles that underlie them. These results for the frequency and placement of conceptual and procedural knowledge in written curricula raise the issue of what level of attention to each may be most productive for student learning. We address this difficult issue below.

Our fine-grained analysis also supported a detailed description of the sequence that written curricula followed in presenting length measurement content across the grades. Here we also found substantial commonalities and some significant differences in the order of specific procedures and concepts. All curricula began with the qualitative comparison of lengths before introducing units of length. All began metric measurement with nonstandard units (e.g., paper clips, linking cubes, centimeter cubes) before standard units, but quickly introduced rulers (only *SFAW* waited until grade 1). All gave extensive attention to measuring simple paths, primarily with rulers, and later to the perimeter of polygons, where comparatively little attention was given to complex and curved paths. In measuring simple paths, all curricula but *EM* gave greater attention to procedures for *tiling* of units than iterating units. All curricula asked students to estimate the length of objects by visually applying standard or nonstandard units. By grade 2, all curricula frequently included word problems involving lengths, although it was not clear whether these were designed as opportunities to reason about length or to practice base-10 arithmetic. All curricula began to explore multiplicative relationships involving lengths by grade 3, principally through conversions between units.

These commonalities in curricular sequence reflect some aspects of the epistemological sequence, initially proposed by Piaget and explored by others (Lehrer, 2003; NCTM, 2000; National Research Council, 2001; Piaget et al., 1960; Van deWalle, 1994). The observed sequence reflects the initial presumption that qualitative comparison should precede metric measurement (using any units). While Piaget drew no distinction between standard and nonstandard units, all curricula have followed current mathematics education policy documents (e.g., NCTM, 2000, 2006) and guidelines for teachers (Van de Walle, 1994) in introducing nonstandard units before standard units of length. One major departure from the Piagetian sequence is the near absence of *Indirect Comparison* in the curricula as an intermediate step between qualitative comparison and the use of units to determine “how much longer?” Similarly, Sarama and colleagues (2011) have reported that their elementary students did not use indirect comparison across a range of length measurement tasks. The curricular sequence has also included metric elements (visual estimation, sums and differences, and multiplicative relationships) that were not addressed in early developmental research.

Attention to conceptual principles was also sparse in all curricula. Most notably, references to *Unit Iteration* were few, and those that did appear primarily focused on one constraint (avoiding gaps and overlaps between units) without addressing the need for identical units and filling the space to be measured (exhaustion). Two conceptual elements, *Greater Means Longer* and *Unit-Measure Compensation* accounted for more than half (54%) of all conceptual content. Many other elements, including *Conservation Under Motion*, *Conservation Under Partitioning*, *Transitivity*, *Additive Composition*, *Rulers Represent Iterated Units*, *Zero/Scale on Rulers*, were scarcely mentioned. All curricula attempted to define length, although these statements raise issues about clarity, and they generally appeared after qualitative comparison and measurement with nonstandard units and rulers had already begun. It was notable and surprising that references to the *Definition of Perimeter* were three times more common than references to the *Definition of Length*—the more fundamental quantity.

None of the four specific learning challenges identified in prior empirical work were strongly addressed in the U.S. curricula. With respect to understanding length as a stable and measurable attribute, we found issues of concern in how the curricula defined length explicitly—concerns that, in fairness, we have found in other sources. Few definitions were mathematically well formed. Basic properties of length (prior to the introduction of units) were rarely expressed. With respect to understanding the properties of units, we found much greater attention to the inverse relationship between the size of units and the resulting length measure and almost no attention to the requirements that length units be identical in size, exhaust the whole, and be placed along a path parallel to the measured object. Little attention was given to unpacking the structure of rulers, beyond their conventional features (e.g., the placement of inch and centimeter scales), and the scant attention to seeing ruler marks as the endpoints of length intervals appeared well after these tools were used. Very few occasions asked students to use rulers to measure objects that were not aligned with the zero mark, despite the wide reporting of students’ struggle with this situation (e.g., Kamii, 2006; Lehrer et al., 1998; Nunes & Bryant, 1996). Few opportunities were provided for measuring complex or curved paths. Given this pattern, it is not clear that curriculum authors are preparing their materials with full understanding of the documented challenges that students face in learning length measurement.

Despite these strong commonalities in overall content and sequence of particular procedures and concepts, we also found some nontrivial differences between curricula. Overall, *Saxon* focused even more attention on procedures and less on concepts than the other curricula (Table 2); addressed a narrower range of conceptual content and revisited that content less often; posed more *Questions*, especially from *Teachers*; and included more *Demonstrations*. *SFAW*'s treatment of conceptual content focused repeatedly on a few elements (e.g., *Greater Means Longer* and *Unit- Measure Compensation*) and scarcely mentioned many other principles. *SFAW* was the most consistent curriculum of the three in presenting conceptual principles after students began work with related procedures. By contrast, *EM* addressed a much wider range of conceptual content (see also Lee & Smith, 2011) and gave the greatest attention to moving/reusing units, both conceptually and procedurally.

Implications of the Procedural Focus of Written Curriculum

Given the strong procedural focus in written curricula, the assertions that typical classroom instruction is also procedurally focused, and the evidence that some of students' struggles with length measurement align with the character of written curricula, it is tempting to "connect the dots" and directly link the character of written curricula to students' learning problems. Our results may suggest that linkage and generally support the conjecture that the content of written curricula is one cause of those problems, but our data are insufficient to assert that causal claim directly. As we have argued from the outset, the content of written curricula is only one factor shaping students' learning of any topic, including length measurement (Figure 2). All curricula must be enacted, one day and lesson at a time. The planned and actual enactment of written lessons includes teachers' choices to selectively include and delete written lesson elements and to supplement those lessons with additional content. They also include the myriad ways in which teachers frame and shape tasks and activities, manage classroom discussion, and assess student progress. Careful analyses of written curricula, no matter how revealing, can orient and inform research on the actual lessons that students experience, but they cannot replace such research. On the other hand, it is also true that written curricula exert a substantial influence on teachers' enacted lessons (Grouws et al., 2004); that teachers read and learn from curriculum materials (e.g., Choppin, 2008; Remillard, 2000); and that curricula are written to align with high-stakes assessments that increasingly influence what teachers teach. Those considerations are reasons not to expect strong and widespread separation between the written and enacted elementary mathematics curricula.

In this context, how has this analysis deepened our understanding of the sources of the problem of learning measurement that framed the study? How do the observed patterns in procedural and conceptual content relate to this problem? First, given the complex and uncertain relationship between conceptual and procedural knowledge in mathematics (Baroody, Feil, & Johnson, 2007; Star, 2005, 2007), we do not see the basis for claiming that the observed procedural focus is itself problematic for students' learning. Measurement involves doing—acting on the physical world—as much if not more than for other mathematical domains. Instead, we see one important implication to concern the impact of limited access to conceptual knowledge. Even the present range of 10–15% of all length measurement content in each grade could well be deployed to greater effect. Only a few key conceptual elements received repeated attention across grades; many others scarcely received any attention, making important principles effectively invisible for students.

Second, even when key conceptual ideas appeared in the curriculum, they were frequently not fully articulated. The *Partial Statements* of *Unit Iteration* were one of the clearest examples, but there were others. So, clarity and completeness in the expression of conceptual content is a second concern. Third, much conceptual content was only explicit in *Teacher* materials, especially in the early grades, so students did not have direct access. Even by grade 1 we believe it is possible to express key conceptual principles in appropriate ways for and directly to children. Fourth, addressing conceptual content well after calls for students to apply procedures that depend on those principles seems problematic. There is broad agreement that we seek to raise students who expect to understand mathematical procedures and can explain why those procedures are appropriate (National Research Council, 2001). If so, then addressing conceptual principles in closer proximity and more explicit relationship to relevant procedures seems the most promising approach (Baroody et al., 2007; Lehrer, 2003). Delay runs the risk of suggesting that mathematics is not to be understood—a message that too many students learn in school (Schoenfeld, 1988; Skemp, 1978).

One potential response is that teachers may repair the conceptual limitations of written curricula by supplementing with content from other sources. While we do not question the fact of curriculum supplementation or that some exemplary teachers are capable of supplementing in conceptually significant ways, it seems unlikely that such repairs will effectively address the conceptual limitations we have reported on a wide scale—for two reasons. First, it is not clear that print-based or Web-based materials that would address the concerns we have reported are widely available. Second and more important, the research literature on teachers' knowledge of length measurement, as well as our own experience in professional development in measurement, indicates that elementary teachers are often unsure about the conceptual foundations of measurement and/or how to engage measurement concepts in their teaching. Some struggle with the same challenges that their students face (e.g., Baturo & Nason, 1996; Simon & Blume, 1994; Woodward & Byrd, 1983). Teachers whose orientation to measurement is procedural may not see conceptual limitations in their curriculum's treatment of the topic.

Implications for Curriculum Development and Revision

Some simple suggestions for curriculum development and revision flow from this argument. First, authors should consider how students could get more direct and early access to conceptual principles in student materials and how to link those principles to key procedures. Second, they should explore ways of representing the iterative movement of units more dynamically and explicitly. New forms of digital curriculum materials provide new opportunities for overcoming the reliance on static representations on paper textbook pages. Third, greater attention should be given to establishing the connections between physical units, such as tiles and paperclips, and the marks on rulers in order to support students' attention to intervals of space. When physical units are standard (e.g., inch tiles and centimeter cubes), tiling a length and measuring the same length with a ruler can be seen as equivalent. To assess students' understanding of rulers, more measurement tasks should involve objects that are nonaligned with the zero mark or use broken ruler tasks. And more attention should be given to measuring complex paths, including those with corners, as these also reveal more about students' understanding of length than do simple paths.

But enriching the conceptual content of written curricula may have little effect if teachers do not appreciate the role played by conceptual knowledge in explaining and justifying measurement procedures. In addition to arguing for more attention to measurement in preservice education, we also advocate enriching the educative qualities of written curricula to support teacher learning (Ball & Cohen, 1996; Davis & Krajcik, 2005; Males, 2011). This means speaking directly to teachers (e.g., in "professional development notes") about the logic of students' misconceptions, how conceptual principles shape and constrain measurement procedures, and how they apply generally to measurement of many quantities.

Limitations and Next Research Steps

Given the diversity of elementary mathematics curriculum materials in use in the United States and the evidence that they differ in nontrivial ways (Dossey et al., 2008; Stein et al., 2007), the most obvious limitation of this study concerns the curriculum sample. How well do the three U.S. curricula represent the national population of curricula that shape the enacted curriculum around the country? Although we selected our small sample carefully, we cannot completely answer this challenge, as an empirically grounded answer would involve applying our method to all such curricula. But we make three claims about the relevance of our work for addressing the problem of weak student learning: (a) we have analyzed elementary curricula that have shaped the measurement lessons experienced by many U.S. students, (b) we have found no evidence that two of the three curricula have significantly changed their treatment of length measurement in more recent editions, and (c) the overall procedural focus observed in the U.S. curricula held for the Singapore curriculum as well. In addition, our more cursory examinations of other curricula have produced more similarity than difference to what we have reported. While we cannot claim these curricular patterns are completely uniform, even in the United States, we have also found no evidence that we are overstating our case.

Second, since we have not studied the planned or enacted measurement lessons from these curricula (beyond a few cases of informal observation), we do not know what typical enactments look like or what the range of variation may be. Our careful analysis of the written curricula is, however, strong preparation for such work, and we hope to address this limitation in the future. But framing studies of the enacted curricula also poses substantial challenges—principally, the problem of diversity. For each lesson in written curricula, there may be thousands of different enactments, and uncertainty about the structure of the population of enactments makes it difficult to design appropriate samples. That said, sampling classrooms in a variety of communities with teachers of different levels of experience and mathematical background makes sense, as does coupling observation with interviews to begin to assess the reasons why teachers select, omit, and supplement elements of written lessons. Such studies may well reveal more about the effect of other factors depicted in Figure 2.

Finally, our analysis has convinced us of the importance of fine-grained curriculum analysis, but it has also generated challenges and problems. One challenge that we have addressed well has been to demarcate the boundaries of the target content domain on mathematical, rather than curricular grounds. Rather than simply analyzing length measurement lessons, we have analyzed all content that sensibly called for length reasoning. This step led us to include content (e.g., bar graphs, partitioning, and plotting points) that others may not have. Assessing content that calls for length reasoning certainly involves subjective interpretation, but we believe that our more inclusive approach is superior to solely analyzing designated length lessons. We have also allowed higher-level distinctions in measurement knowledge to emerge in the early phases of our analysis. *Conventional* knowledge emerged as a knowledge type to complement *Conceptual* and *Procedural* knowledge in length measurement, and it may apply to all domains of elementary mathematics as well. But we have struggled with other issues. We have found the distinction between *Question* and *Problem* difficult to operationalize in a manner that generates reliable results. We have also struggled to identify the optimum level of detail for particular knowledge elements, sometimes crafting some elements to include too much content (e.g., *Unit Iteration*) and other times perhaps framing elements too narrowly (e.g., *Meaning of Length Measure* and *Length Measure Requires Length Units*). It is not clear that there is a principled way to resolve these questions.

Conclusion

Current treatments of length measurement in elementary written curricula may be enriched and improved in a number of related ways. Making these changes could strengthen students' (and teachers) learning of length measurement and thereby provide a stronger foundation for understanding measurement more generally (especially, area and volume) and other core elementary mathematics content (e.g., fractions). Curricula may profitably identify and clearly present more central conceptual principles; locate that content more closely to work with related measurement procedures; present concepts, especially definitions, directly in student materials; include tasks and activities that specifically address known challenges for students; and communicate conceptual principles and students' learning challenges directly to the teachers who will use those materials.

Acknowledgments — The research reported in this article was carried out by the *Strengthening Tomorrow's Education in Measurement* (STEM) project team at Michigan State University with support from the National Science Foundation (REC-0634043 and DRL-0909745). The opinions expressed here are those of the authors and do not necessarily reflect the views of the Foundation. We acknowledge the significant contributions of Kuo-Liang Chang, Hanna Figueras, Funda Gonulates, Patrick Greeley, Matthew Pahl, and Alexandria Musselman to the coding scheme development and the coding of textbook content. Some parts of this analysis were presented at the 2008 Annual Meeting of the American Educational Research Association in New York.

References

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*, 6–14.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*, 115–131.

Barrett, J. E., & Clements, D. H. (2003). Quantifying length: Children's developing abstractions for measures of linear quantity in one-dimensional and two-dimensional contexts. *Cognition & Instruction*, *4*, 475–520.

Barrett, J. E., Clements, D. H., Klanderman, D., Pennisi, S., & Polaki, M. K. (2006). Students' coordination of geometric reasoning and measuring strategies on a fixed perimeter task: Developing mathematical understanding of linear measurement. *Journal for Research in Mathematics Education*, *37*, 187–221.

Barrett, J. E., Sarama, J., Clements, D. H., Cullen, C., McCool, J., Witkowski-Rumsey, C., & Klandeman, D. (2012). Evaluating and improving a learning trajectory for linear measurement in elementary grades 2 and 3: A longitudinal analysis. *Mathematical Thinking and Learning*, *14*, 28–54.

Battista, M. T. (2003). Understanding students' thinking about area and volume measurement. In D. H. Clements & G.

Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 122–142). Reston, VA: National Council of Teachers of Mathematics.

Battista, M. T. (2012). *Cognition-based assessment and teaching of geometric measurement*. Portsmouth, NH: Heinemann.

Baturo, A., & Nason, R. (1996). Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics*, *31*, 235–268.

Blume, G. W., Galindo, E., & Walcott, C. (2007). Performance in measurement and geometry from the viewpoint of *Principles and Standards for School Mathematics*. In P. Kloosterman & F. K. Lester (Eds.), *Results and interpretations of the 2003 Mathematics Assessment of the National Assessment of Educational Progress* (pp. 95–138). Reston, VA: National Council of Teachers of Mathematics.

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.

Bragg, P., & Outhred, L. (2001, April). *Procedural and conceptual knowledge: The case of measurement*. Paper presented to the 2001 Annual Meeting of the American Educational Research Association, Seattle, WA.

Bragg, P., & Outhred, L. (2004). A measure of rulers: The importance of units in a measure. *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, vol. 2 (pp. 159–166). Bergen, Norway: Program Committee.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth, NH: Heinemann.

Chang, K., Males, L. M., Mosier, A., & Gonulates, F. (2011). Exploring US textbooks' treatments of the estimation of linear measurements. *ZDM*, *43*, 697–708.

Chappell, M. F., & Thompson, D. R. (1999). Perimeter or area?: Which measure is it? *Mathematics Teaching in the Middle School*, *5*, 20–23.

Charles, R., Chappell, J.H., Cavanaugh, M., Chancellor, D., Copley, J. V., Crown, W. D., . . . Schielack, J. F. (2011). *enVision Math.* Glenview, IL: Pearson Education Inc.

Charles, R., Crown, W., & Fennell, F. (2008). *Mathematics, Michigan edition*. Glenview, IL: Pearson Education Inc.

Choppin, J. (2008). Curriculum-context knowledge: Teacher learning from successive enactment of a Standards-based mathematics curriculum. *Curriculum Inquiry*, *39*, 287–320.

Clements, D. H. (1999). Teaching length measurement: Research challenges. *School Science and Mathematics*, *99*, 5–11.

Clements, D. H., Battista, M. T., & Sarama, J. (1998). Development of geometric and measurement ideas. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 201–225). Mahwah, NJ: Lawrence Erlbaum.

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

Common Core State Standards Initiative (CCSSI). (2010). *Common Core State Standards for Mathematics*. Washington, DC: The National Governors Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.coreStandards.org

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*, 3–14.

Dietiker, L., Gonulates, F., & Smith, J.P. (2011). Enhancing opportunities for student understanding of length measure. *Teaching Children Mathematics*, *18*, 252–259. diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition & Instruction*, *10*, 105–225.

Dossey, J., Halvorsen, K., & McCrone, S. (2008). *Mathematics education in the United States 2008: A capsule summary fact book*. Reston, VA: National Council of Teachers of Mathematics.

Floden, R. E. (2004). The measurement of opportunity to learn. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational assessment* (pp. 231–266). Washington, DC: National Academy Press.

Fong, H. K., Ramakrishnan, C., & Lau, P. W. (2007). *My pals are here! Maths* (2nd ed.). Singapore: Marshall Cavendish.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: Reidel.

Gehrke, N. J., Knapp, M. S., & Sirotnik, K. A. (1992). In search of the school curriculum. *Review of Educational Research*, *18*, 51–110.

Grouws, D. A., Smith, M. S., & Sztajn, P. (2004). The preparation and teaching practices of United States mathematics teachers: Grades 4 and 8. In P. Kloosterman & F.K. Lester (Eds.), *Results and interpretations of the 1990 through 2000 Mathematics Assessments of the National Assessment of Educational Progress* (pp. 221–267). Reston, VA: National Council of Teachers of Mathematics.

Hart, K. M. (1981). *Children's understanding of mathematics*: 11–16. London, UK: John Murray.

Herbel-Eisenmann, B. A., & Otten, S. (2011). Mapping mathematics in classroom discourse. *Journal for Research in Mathematics Education*, *42*, 451–485.

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

Hiebert, J. (1984). Why do some children have trouble with learning measurement concepts? *Arithmetic Teacher*, *31*, 19–24.

Hiebert, J. & Lefvre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J.

Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearner, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Kamii, C. (2006). Measurement of length: How can we teach it better? *Teaching Children Mathematics*, *13*, 154–158.

Kamii, C., & Clark, F. (1997). Measurement of length: The need for a better approach to teaching. *School Science and Mathematics*, *97*, 116–121.

Kamii, C., & Kysh, J. (2006). The difficulty of "length x width": Is a square a unit of measurement? *Journal of Mathematical Behavior*, *25*, 105–115.

Kasten, S. E., & Newton, J. (2011). An analysis of K–8 measurement grade level expectations. In J. P. Smith (Ed.), *Variability is the rule: A companion analysis of K–8 state mathematics standards* (pp. 13–40). Charlotte, NC: Information Age Publishing, Inc.

Kloosterman, P., Rutledge, Z., & Kenney, P. A. (2009). Exploring results of the NAEP: 1980s to the present. *Mathematics Teaching in the Middle School*, *14*, 357–365.

Larson, N. (2004). *Saxon math*. Austin, TX: Saxon Publishers, Inc.

Lee, K., & Smith, J. P. (2011). What's different across an ocean?: How Singapore and U.S. elementary mathematics curricula introduce and develop length measurement. *ZDM*, *43*, 681–696.

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

Lehrer, R., Jaslow, L., & Curtis, C. (2003). Developing an understanding of measurement in the elementary grades. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 100–121). Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., Jenkins, M., & Osana, H. (1998). Longitudinal study of children's reasoning about space and geometry: In R.

Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 137–167). Mahwah, NJ. Lawrence Erlbaum.

Levine, S. C., Kwon, M., Huttenlocher, J., Ratliff, K., & Deitz, K. (2009, July). Children's understanding of ruler measurement and units of measure: A training study. Paper presented at the Annual Meeting of the Cognitive Science Society, Amsterdam, The Netherlands.

Lubienski, S. T. (2003). Is our teaching measuring up: Race-, SES-, and gender-related gaps in measurement achievement. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 282–292). Reston, VA: National Council of Teachers of Mathematics.

Lubienski, S. T., & Crockett, M. D. (2007). NAEP findings regarding race and ethnicity: Mathematics achievement, student affect, and school-home experiences. In P. Kloosterman & F. K. Lester (Eds.), *Results and interpretations of the 2003 Mathematics Assessment of the National Assessment of Educational Progress* (pp. 227–260). Reston, VA: National Council of Teachers of Mathematics.

Males, L.M. (2011). *Educative supports for teachers in middle school mathematics curriculum materials: What is offered and how is it expressed?* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (Accession Order No. AAT 3489750).

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's third international mathematics and science study (TIMSS)*. Boston, MA: TIMSS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grade*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Center of Education Statistics. (1997). *Pursuing excellence: A study of U.S. eighth-grade mathematics and science achievement in international context*. Washington, DC: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

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

Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Oxford, UK: Blackwell Publishers.

Nunes, T., Light, P. & Mason, J. (1993). Tools for thought: The measurement of length and area. *Learning and Instruction*, *3*, 39–54.

Pettito, A. L. (1990). Development of number line and measurement concepts. *Cognition & Instruction*, *7*, 55–78.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry*. New York, NY: Basic Books.

Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *The Elementary School Journal*, *100*, 331–350.

Sarama, J., Clements, D. H., Barrett, J., Van Dine, D. W., & McDonel, J. S. (2011). Evaluation of a learning trajectory for length in the early years. *ZDM*, *43*, 667–680.

Schifter, D., & Szymaszek, J. (2003). Structuring a rectangle: Teachers write to learn about their students' thinking. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 143–156). Reston, VA: National Council of Teachers of Mathematics.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, The Netherlands: Kluwer Publishing.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist*, *23*, 145–166.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making. In D. A. Grouws (Ed.), *Handbook of research in teaching and learning mathematics* (pp. 334–370). New York, NY: Macmillan.

Schwartz, J. L. (1988). Intensive quantity and referent transforming arithmetic operations. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 41–52). Reston, VA: NCTM.

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

Silver, E. A. (2009). Cross-national comparisons of mathematics curriculum materials: What might we learn. *ZDM*, *41*, 827–832.

Simon, M. A., & Blume, G. W. (1994). Building and understanding multiplicative relationships: A study of prospective elementary teachers. *Journal for Research in Mathematics Education*, *25*, 472–494.

Sinclair, N. & Jackiw, N. (2002). Microworld design in a whole class context. *Journal for Educational Research in Computers*, *27*, 111–145.

Skemp, R. R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher*, *26*(3), 9–15.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993/94).Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*, 115–163.

Smith, J. P., Dietiker, L., Lee, K., Males, L. M., Figueras, H., Mosier, A., Chang, K., & Sisman, G. (2008, March). *Framing the analysis of written measurement curriculum*. Paper presented to the Annual Meeting of the American Educational Research Association, New York, NY.

Smith, M. S., Arbaugh, F., &Fi, C. (2007). Teachers, the school environment, and students: Influences on students' opportunities to learn mathematics in grades 4 and 8. In P. Kloosterman & F. K. Lester (Eds.), *Results and interpretations of the 2003 Mathematics Assessment of the National Assessment of Educational Progress* (pp. 191–226). Reston, VA: National Council of Teachers of Mathematics.

Star, J.R. (2005).Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*, 404–411.

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education*, *38*, 132–135.

Stavy, R., & Tirosh, D. (2000). *How students (mis-) understand science and mathematics: Intuitive rules*. New York, NY: Teachers College Press.

Stein, M.K., Grover, B., & Henningsen, M. (1996). Building students' capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*, 455–488.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of mathematics teaching and learning* (pp. 319–369). Reston, VA: National Council of Teachers of Mathematics.

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

Stephan, M., Bowers, J., Cobb, P., & Gravemeijer, K. (2003). Supporting students' development of measuring conceptions: Analyzing students' learning in social context. *Journal for Research in Mathematics Education*, monograph #12. Reston, VA: National Council of Teachers of Mathematics.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York, NY: Simon & Schuster.

The University of Chicago School Mathematics Project. (2007). *Everyday Mathematics* (3rd ed.). Chicago, IL: Wright Group/McGraw Hill.

The University of Chicago School Mathematics Project. (2012). *Everyday Mathematics (*Common Core ed.*)*. Chicago, IL: Wright Group/McGraw Hill.

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. (1994). Calculational and conceptual orientations in teaching mathematics. In D. B. Aichele (Ed.), *1994 yearbook of the National Council of Teachers of Mathematics* (pp. 79–92). Reston, VA: National Council of Teachers of Mathematics.

Thompson, P. W. (1994a). Images of rate and operational understanding of the Fundamental Theorem of Calculus. *Educational Studies in Mathematics*, *26*, 229–276.

Thompson, P. W. (1994b). The development of speed and its relationship to concepts of rate. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 179–234). Albany, NY: SUNY Press.

Thompson, P. W., & Saldanha, L. A. (2003). Fractions and multiplicative reasoning. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 95–113). Reston, VA: National Council of Teachers of Mathematics.

Thompson, T. D., & Preston, R. V. (2004). Measurement in the middle grades: Insights from NAEP and TIMSS. *Mathematics Teaching in the Middle School*, *9*, 514–519.

Trafton, P. R., Reys, B. J., & Wasman, D. G. (2001). Standards-based mathematics curriculum materials: A phrase in search of a definition. *Phi Delta Kappan*, *83*, 259–264.

Van de Walle, J. A. (1994). *Elementary school mathematics: Teaching developmentally* (2nd ed.). New York, NY: Longman.

van den Heuvel-Panhuizen, M., & Buys, K. (2008). *Young children learn measurement and geometry*. Rotterdam, The Netherlands: Sense Publishers.

Vergnaud, G. (1994). Multiplicative conceptual field: What and why? In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 41–59). Albany, NY: SUNY Press.

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

Weiss, I., Pasley, J., Smith, P., Banilower, E., & Heck, H. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.

Woodward, E., & Byrd, F. (1983).Area: Included topic, neglected concept. *School Science and Mathematics*, 83, 343-347.