Many types of studies contribute to the field of education. But too few, in our opinion, go to the heart of the educational enterprise—developing scientifically based practices, pedagogies, programs, and policies. Whether developed for children or their teachers, these are the main malleable factors that affect the quality of children’s educational experiences. In this chapter, we describe why we believe that this type of research-and-development program should take precedence in early childhood education and then describe a framework for such a program.
WHY DO WE NEED RESEARCH-BASED DEVELOPMENT
OF PRACTICES, PEDAGOGIES, PROGRAMS, AND POLICIES?

What directly affects the quality and effectiveness of young children's experiences in the classroom? Teachers do, including their practices and pedagogical strategies (Darling-Hammond, 1997; Ferguson, 1991; National Research Council, 2001; Schoen, Cebula, Finn, & Fi, 2003). In addition, programs or curricula for children have a substantial impact on teachers and their practices and on what children experience and learn (Goodlad, 1984; Grant, Peterson, & Shojgreen-Downer, 1996; National Research Council, 2009; Whitehurst, 2009). Similarly, professional development practices (e.g., workshops) and programs (e.g., certifications, degrees) affect teachers (Darling-Hammond, 1997; Ferguson, 1991; National Research Council, 2001; Sarama & DiBiase, 2004; Schoen, et al., 2003). (Indeed, a combination may be best. Top-down imposition of a new curriculum with limited professional development and support, for example, may have limited influence on teachers’ beliefs and practices, Stein & Kim, 2009.) However, the quality of all these varies widely and does not show steady improvement year to year (Early, et al., 2005; Goodlad, 1984; National Research Council, 2009). A major reason is that practices, programs, and the policies that should support them are rarely developed or evaluated and revised following systematic, much less scientific, research methods (Clements & Battista, 2000; Davidson, Fields, & Yang, 2009). We begin by defining what we mean by scientific research.

Science includes the observation, description, analysis, hypothesizing, experimental investigation, and theoretical explanation of phenomena. Scientific knowledge is accepted as more reliable than everyday knowledge because the way in which it is developed is explicit and repeatable. “Our faith [in it] rests entirely on the certainty of reproducing or seeing again a certain phenomenon by means of certain well defined acts” (Valéry, 1957, p.1253, as quoted in Glasersfeld, 1995). Scientific method, or research, is disciplined inquiry (Cronbach & Suppes, 1969). The term “inquiry” suggests that the investigation’s goal is answering a specific question. The term “disciplined” suggests that the investigation should be guided by concepts and methods from disciplines and connected to relevant theory in those disciplines, and also that it should be in the public view so that the inquiry can be inspected and criticized. The use of research methods, and the conscientious documentation and full reporting of these processes—data collection, argumentation, reasoning, and checking for counterhypotheses—distinguishes disciplined inquiry from other sources of opinion and belief (Cronbach & Suppes, 1969; National Research Council, 2002; Shulman, 1997).
Science does not, however, produce the "truth" or a single correct view. It provides reliable ways of dealing with experiences and pursuing and achieving goals (Glasersfeld, 1995). It involves the process of progressive problem solving (Scardamalia & Bereiter, 1994). Thus, the goal for scientific methods of research and development cannot be to develop a single "ideal" (practice, pedagogy, program, or policy—hence referred to as products), but rather dynamic problem solving, progress, and advancement beyond present limits of competence (Dewey, 1929; Scardamalia & Bereiter, 1994; Tyler, 1949). Ironically, another implication is that educational products should be based on research—as defined here. Given that traditions, social interactions, and politics have strong effects on education, the checks and balances of scientific research are essential to progress.

Still, does that not limit the creativity of researchers, developers, and teachers? Somewhat ironically, we believe the opposite. Scientific knowledge is necessary but not sufficient, for the continued development of high-quality educational products. More than 120 years ago, William James made this same argument, speaking of the young science of his own time, psychology.

You make a great, a very great mistake, if you think that psychology, being the science of the mind's laws, is something from which you can deduce definite programmes and schemes and methods of instruction for immediate classroom use. Psychology is a science, and teaching is an art; and sciences never generate arts directly out of themselves. An intermediary inventive mind must make the application, by using its originality. (James, 1892/1958, pp. 23-24)

James argues that scientific knowledge is applied artfully to create teaching products. Such research-to-practice methods are included in our framework. However, this method used alone is incorrect in its presumptions (that extant research is a sufficient source for development of products), insensitive to changing goals in the content area (new standards are created at a fast pace that research cannot comprehensively address), and unable to contribute to a revision of the theory and knowledge on which it is built (i.e., it is inherently conservative, evaluating "what is")—the second critical goal of scientific research and development. In contrast, research should be present in all phases of the creative or development process.

In this way, the framework is mainly about researcher-developers (which includes some teachers, of course), but before we leave this section, let's address whether such a scientific approach denies professionalism and creativity to classroom teachers. Again, we claim it promotes them—in scientists, developers, and teachers. One reason is that professionals such as doctors and teachers share a scientific knowledge base; that is, as all professionals, they share scientific guidelines of systematic, rather than idiosyncratic, practice. Such systematic practice is more effective and amenable
to scientifically based improvement than private, idiosyncratic practice (Raudenbush, 2009). This is not to say that teachers should deliver "scripted" lessons with little or no interpretation. Rather, it is to argue that their creativity should be in building upon the research foundation, using the resources of science (and that wisdom of expert practice not yet studied), to create environments and interactions that promote their children's development and learning. It does mean that many of us hold notions of teacher creativity that may benefit from a revision. As a personal example, when one of us (Clements) taught kindergarten, all the early childhood educators around him believed that "creative teachers" made up all their ideas and made all their materials. I too believed, then, that if I copied an idea or game that another successful teacher used, I was "less creative." Such thinking erects an unfortunate barrier to the spread of the most effective practices and programs. Instead, we believe, teachers' creativity is best used to use and imaginatively apply the best of the resources of science and the wisdom of expert practice.

All this is not to say that scientific programs cannot be outperformed (e.g., by a talented, idiosyncratic teacher). James had more to say on this matter.

The science of logic never made a man reason rightly, and the science of ethics (if there be such a thing) never made a man behave rightly. The most such sciences can do is to help us catch ourselves up and check ourselves, if we start to reason or to behave wrongly; and to criticize [sic] ourselves more articulately after we have made mistakes. A science only lays down lines within which the rules of the art must fall, laws which the follower of the art must not transgress; but what particular thing he shall positively do within those lines is left exclusively to his own genius. One genius will do his work well and succeed in one way, while another succeeds as well quite differently; yet neither will transgress the lines.... And so everywhere the teacher must agree with the psychology, but need not necessarily be the only kind of teaching that would so agree; for many diverse methods of teaching may equally well agree with psychological laws. (James, 1892/1988, p. 24)

Thus, there are several approaches, but each should be consistent with what is known about teaching and learning. Those that appear successful, such as our talented teacher's, should be documented, investigated as to why the approach is successful, and added to the research literature. Without such research methodologies, the talented teacher's practices and materials will be limited in their contribution to the myriad teachers and researcher-developers who come after.

This is also not to say that all research should be scientific. Other types of research may make serious contributions, such as narrative (Bruner, 1986), or humanistic (Schwandt, 2002) perspectives, historical research (Darling-Hammond & Snyder, 1992), aesthetic approaches (Eisner, 1998), or literary
criticism (Papert, 1987), just to name a few. Such approaches complement the scientific research methods described here. Of course, no single scientific finding or set of findings should dictate pedagogy. Consistent with James, John Dewey stated the following.

No conclusion of scientific research can be converted into an immediate rule of educational art. For there is no educational practice whatever which is not highly complex; that is to say, which does not contain many other conditions and factors than are included in the scientific finding. Nevertheless, scientific findings are of practical utility, and the situation is wrongly interpreted when it is used to disparage the value of science in the art of education. What it militates against is the transformation of scientific findings into rules of action. (Dewey, 1929, p. 19)

Consistent with Dewey's formulation, our framework for research-and-development rejects strict "rules" but values scientific research for its practical, and political, utility.

In summary, scientific knowledge is valued because it offers reliable, self-correcting, documented, shared knowledge based on research methodology (Mayer, 2000; National Research Council, 2002). Education is a design science (Brown, 1992; H. A. Simon, 1969; R. Walker, 2011; Wittmann, 1995) and knowledge created during research-and-development should be both generalized and placed within a scientific research corpus, peer reviewed, and published. However, this is not a deterministic science and certainly not one limited to (although it includes) quantitative experiments (Dewey, 1929). As the framework presented here will make clear, many research methodologies, mostly qualitative, are used to produce research-based education.

**EARLY ATTEMPTS TO BASE EDUCATION ON RESEARCH**

Research, especially psychological research from the time of William James on, has played a substantial role in education, especially in early childhood (Clements, 2008b). However, its role has been less to produce practical materials for teaching, than to interpret the phenomena of early education (Ginsburg, Klein, & Starkey, 1998). As stated previously, this is important, but indirect. We believe that most of the ways that development might be based on research should be employed. We next describe a small number of early attempts to base product development on research.

Early efforts to write research-based teaching approaches and materials often were grounded in the broad philosophies, theories, and empirical results on learning and teaching. For example, in early childhood, early applications of Piaget's theories often led to suggestions that children be to perform accurately on Piagetian clinical tasks. Other incorporated materials...
directly adapted from those tasks (Forman & Fosnot, 1982; Kamii, 1973). These were not particularly successful. Even detailed analyses of Piagetian research failed to guide the development of programs or curricula in directly useful ways (Duckworth, 1979).

Others based their educational programs on Piaget's constructivist foundation. For example, Duckworth encouraged teachers to create environments in which children would “have wonderful ideas” (Duckworth, 1973). Such programs have been arguably more successful, although the interpretations varied widely (Forman, 1993). Indeed, the programs were distinct. The broad philosophy and theory, unsurprisingly, leaves much room for interpretation and provides little specific guidance for teaching or the development of materials.

In summary, the research-to-practice model has a less than stellar historical record (Clements & Battista, 2000; Gravemeijer, 1994b). As stated previously, it also is limited in its contribution to either theory or practice (Clements, 2007, 2008b). The alternative we propose here is less about new research approaches or practices, and more about a specific framework for synthesizing those that have been used successfully into a complete scientific research-and-development system for designing and evaluating educational products.

A COMPREHENSIVE FRAMEWORK FOR RESEARCH-BASED EDUCATION

We developed a comprehensive framework detailing the methods used to create and evaluate research-based practices, pedagogies, and programs (with implications for policies) so we could contribute to both theory and practice. First, we established goals based on the belief that any valid scientific development product should address two basic issues—effect and conditions—in three domains, practice, policy, and theory, as diagrammed in Figure 21.1.

To achieve these goals, researcher-developers must build on previous research, structure and revise the nature and content of components in accordance with models of children’s thinking and learning in a domain, and conduct formative and summative evaluations in a series of progressively expanding social contexts. These form the categories of research-and-development activity that define our framework. These categories include ten phases of such activity that warrant claiming that a product is based on research (Clements, 2007; Sarama & Clements, 2008). The categories and phases involve a combination of research methods; no single method would be adequate. For example, design experiments (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Ruthven, Laborde, Leach, & Tiberghien, 2009; The Design-Based Research Collective, 2003; R. Walker, 2011), developed as a way to conduct formative research to test and refine educational designs
### Methods for Developing Scientific Education

#### Practice

<table>
<thead>
<tr>
<th>Effects</th>
<th>Policy</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the intervention effective in helping children achieve specific learning goals? Are the intended and unintended consequences positive? (6–10)*</td>
<td>Are the goals important (e.g., to meeting standards)? (1, 5, 10) What are the effect sizes? (9, 10) What effects does it have on teachers? (10)</td>
<td>Why is the intervention effective? (all) What were the theoretical bases? (1, 2, 3) What cognitive changes occurred and what processes were responsible? That is, what specific components and features account for its impact and why? (4, 6, 7)</td>
</tr>
<tr>
<td>Conditions</td>
<td>When and where? Under what conditions is the intervention effective? (Do findings generalize?) (8, 10)</td>
<td>What are the support requirements (7) for various contexts? (8–10)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses refer to the phases of the Framework described in the following sections.

**Figure 21.1** Goals of research and development (adapted from Clements, 2007).

(Collins, Joseph, & Bielaczyc, 2004) are central, but are usually limited to pilot testing (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004; National Research Council, 2004, p. 75), put too little focus on the development of curricula, and do not address the full range of questions (Clements, 2008a; R. Walker, 2011). Our work is based on the assumption that all appropriate methods should be synthesized into a coherent, complete framework for research and development, as described in Figure 21.2 (see Clements, 2007, for a full description). (Space prohibits describing the work of many researcher-developers from which this framework was abstracted, but see citations and descriptions of their work in Clements, 2002, 2007, 2008b; Clements & Battista, 2000; Sarama & Clements, 2008.)

### DESCRIPTION AND APPLICATION OF THE FRAMEWORK

In this section, we describe the categories and phrases of the Framework in more detail. We also briefly illustrate its application using our instantiation of it developing and evaluating the *Building Blocks* early childhood mathematics curriculum and the TRIAD model of intervention at scale.
Summative Assessment

Empirical evidence is collected on the effects of the intervention. Summative phases 9 and 10 both use randomized field trials and differ on scale. Phase 9 checks the efficacy. Phase 10 examines the fidelity of enactment, and sustainability of the curriculum when implemented on a large scale, and the critical contextual and implementation variables that influence its effectiveness.

Formative Assessment

Empirical evidence is collected to evaluate appeal, usability, and effectiveness of an instantiation of the intervention, which is reviewed after each phase. What meaning do teachers and children give to the intervention in expanding social contexts?

Learning Trajectories

Activities are structured in accordance with empirically-based models of children's thinking and learning in the targeted subject-matter domain.

A Priori Foundations

In variants of the research-to-practice model, extant research is reviewed and implications for the nascent development effort drawn in 3 domains:

- Phase 1: Subject matter content, including the role it would play in children's development.
- Phase 2: General issues concerning psychology, education, and systemic change.
- Phase 3: Pedagogy, including the effectiveness of certain types of activities.

Figure 21.2 The Framework for Comprehensive Research and Development.

A Priori Foundations

In this category, established research review procedures (e.g., Galvin, 2009; Light & Pillemer, 1984) and content analyses (National Research Council, 2004) are employed to garner knowledge concerning the specific
subject matter content, including the role it would play in children’s development (phase 1); general issues concerning psychology, education, and systemic change (phase 2); and pedagogy, including the effectiveness of certain types of environments and activities (phase 3).

**Phase 1: Subject Matter A Priori Foundation**

Developing goals is a complex process, not all of which, perhaps most of which, is not amenable to scientific investigation. Societal-determined values and goals are substantive components of any program (Hiebert, 1999; National Research Council, 2002; Schmandt, 2002; Tyler, 1949). Creating goals requires a cooperative process among the many legitimate direct and indirect stakeholders (van Oers, 2003). The array of advice from a wide variety of such stakeholders involved in such large-scale projects as those in the domain of mathematics, the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000) and *Common Core State Standards* (CCSSO/NGA, 2010) illustrate this point. For example, subject matter experts and their organizations evaluated whether these goals included those concepts and procedures that play a central role in the domain (cf. content analyses in National Research Council, 2004). Nevertheless, scientific procedures help identify subject-matter content that is valid within the discipline and makes a substantive contribution to the development of children in the target population. That is, concepts and procedures of the domain should build from the children’s past and present experiences (Dewey, 1902/1976) and be generative in children’s development of future understanding (Clements, Sarama, & DiBiase, 2004).

In our *Building Blocks* project, (Clements & Sarama, 2007a), we built our goals upon the results of a large conference we organized (funded by NSF and the ExxonMobil Education Foundation) that involved representatives from state departments of education and from the U.S. Department of Education, mathematicians, mathematics and early childhood educators (pre-K to university) and researchers, childhood policy makers, and developers. This was preceded and followed by extensive research reviews (for a full report, which also influenced the Curriculum Focal Points and Common Core, see Clements, et al., 2004). We vetted the specific goals for our project to an advisory board consisting of members of these same groups. As an example, we determined that the competence of subitizing is crucial to young’s children’s mathematical development (Clements, 1999; Clements & Conference Working Group, 2004). Subitizing, the ability to recognize and name the numerosity of a group quickly (from the Latin “to arrive suddenly”), is the earliest developing quantitative, or numerical, ability (Sarama & Clements, 2009). It also contributes to many other competencies, such as counting and arithmetic (Baroody, 1987; Clements, 1999; Fuson, 1992; Sarama & Clements, 2009).
Phase 2: General A Priori Foundation

In this phase, philosophies, theories, and empirical results on teaching and general educational issues are reviewed for their applicability to the product. Research-developers might start from an Ausubelian, Piagetian, or general constructivist perspective and proceed in any of several directions (Forman, 1993; Lawton, 1993). In addition, theory and research offer perspectives on children's and teachers' experiences with similar products. For our own part, we used theory and research on early childhood learning and teaching (Clements, 2001; National Research Council, 2001), to decide that the basic approach of Building Blocks would be finding the mathematics in and developing mathematics from children's activity. The materials were designed to facilitate children's extending and mathematizing their everyday activities, such as building blocks, art projects, stories, songs, and puzzles.

Phase 3: Pedagogical A Priori Foundation

In this phase, research relevant to creating specific types of educational environments and activities is reviewed. Intuition of practitioners—the art of teaching—is also garnered as much as possible by viewing patterns of promising practice (Dewey, 1929; Hiebert, 1999).

A science only lays down lines within which the rules of the art must fall, laws which the follower of the art must not transgress; but what particular thing he shall positively do within those lines is left exclusively to his own genius... many diverse methods of teaching may equally well agree with psychological laws. (James, 1892/1958, p. 24)

Note that James treats research only as providing a priori foundations. Our framework uses research in this way, but considers this just a beginning.

Building Blocks pedagogical foundations were based on the same body of research (e.g., National Research Council, 2001), including a wide range of grouping (whole group, small group, individual) and teaching strategies (from the design of the entire environment to explicit instruction to centers and "teachable moments" during play). As just one example, for a minor, but important, component of the curriculum, we consulted empirical data on features that appeared to make computer programs motivating (Escoedo & Evans, 1997; Lahm, 1996; Shade, 1994) and effective (Childers, 1989; Clements & Sarama, 1998; Lavin & Sanders, 1983; Murphy & Appel, 1984; Sarama, Clements, & Vukelic, 1996).

Learning Model, Learning Trajectories

This phase differs from phase 3 in the focus on the children's thinking and learning, rather than teaching strategies alone, in the greater degree
of specificity, and in the iterative nature of its application. That is, in practice, models are usually created or refined along with the development of instructional tasks, using, clinical interviews, teaching experiments, and design experiments.

**Phase 4: Structure According to Specific Learning Trajectories**

Learning trajectories are found or developed to form the core of the product, especially for a curriculum or teaching sequence (M. A. Simon, 1995). Learning trajectories are based on the idea that children follow natural *developmental progressions* in learning and development. As they learn to crawl, then walk, then run, then run, skip, and jump with increasing speed and dexterity, they follow natural developmental progressions in learning in other domains. Learning trajectories built upon natural developmental progressions and empirically based models of children’s thinking and learning are more mature in some areas, such as literacy (e.g., progressions within phonemic awareness and alphabet recognition, as well as movement from these to early graphophonemic analysis, Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Brice & Brice, 2009; Justice, Pence, Bowles, & Wiggins, 2006; Levy, Gong, Hessels, Evans, & Jared, 2006) and mathematics (Carpenter & Moser, 1984; Case, 1982; Griffin & Case, 1997), but are also developed in science (Hmelo-Silver & Duncan, 2009; National Research Council, 2007), albeit less for the earliest years (Brenneman & Gelman, 2009), and in social-emotional development (e.g., Bredekamp, 2014). Sometimes different names are used (e.g., “learning progressions”) and some describe developmental progressions, but not instructional suggestions. However, they share a family resemblance and each can be used to serve the purposes of research and development as proposed here.

We believe that much of the educational potential of learning trajectories lies in their ability to connect developmental progressions to the educational environment and to teaching. We define learning trajectories as “descriptions of children’s thinking and learning in a specific domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking, created with the intent of supporting children’s achievement of specific goals in that domain” (Clements & Sarama, 2004b, p. 83). Thus, in our view, complete learning trajectories have three parts: a goal, a developmental progression or path along which children develop to reach that goal, and a set of recommendations for educational environments and activities, matched to each of the levels of thinking in that path that help children develop ever higher levels of thinking.

The product of this stage is a well-developed cognitive model of children’s learning as expressed in learning trajectories. Ideally, such models
specify knowledge structures, the development of these structures, mechanisms or processes of development, and developmental progressions of nascent learning trajectories that specify hypothetical routes that children might take in achieving the goal (Sarama & Clements, 2009, presents detailed cognitive models not included here).

As an example, our synthesis of research for the Building Blocks project created a first draft of a developmental progression for subitizing. We chose to illustrate the subitizing learning trajectory in this chapter due to its simplicity (i.e., learning trajectories for other topics are longer and more complex). Although there are several features, the basic characteristics are the number of objects and the development of the type of subitizing. First, of course, very young children begin very small numbers—one or two. They slowly develop the ability to subitize larger numbers. Second, the type of subitizing develops from perceptual subitizing to conceptual subitizing. Perceptual subitizing involves recognizing a number of objects without consciously using other mental or mathematical processes and then naming it. This is limited to sets of up to 4 to 6 objects. Conceptual subitizing plays an advanced organizing role, as seeing “10” on a pair of dice by recognizing the two collections (via perceptual subitizing) and consciously composing them. These advancements can be seen in Figure 21.3, which illustrates a portion of this learning trajectory. The left column names and describes each level of thinking in the developmental progression; examples of behaviors are shown in a smaller font (from Sarama & Clements, 2009, which also describes cognitive science descriptions of mental components and processes not included here). We used clinical interviews to check that developmental progression. A simple example of a revision these engendered was a differentiation between the perceptual subitizer to 5 and the conceptual subitizer to 5, which did not exist in the original version.

We then used research reviews and targeted teaching experiments to design the initial instructional activities. The right column of Figure 21.3 provides examples of the types of environments and activities that help children construct that level of thinking (from Clements & Sarama, 2009). These involve informal interactions (e.g., see the first level) and intentional activities. As an example of the latter, a simple, enjoyable “snapshots” game is played at many levels, starting with “Perceptual Subitizer to 4.” The activity is introduced by talking about cameras, and having “our eyes and mind take snapshots” like cameras. The teacher covers, say, three counters with a dark cloth. She reminds children to watch carefully to take a “snapshot,” then uncovers the counters for two seconds only. Children show how many counters they saw with their fingers. Once they have seen all responses, teachers ask children to whisper to each other how many counters they saw. The teacher then uncovers the counters to check answers, and so forth.
<table>
<thead>
<tr>
<th>Developmental Progression</th>
<th>Instructional Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small Collection Namer</strong>&lt;br&gt;Names groups of 1 to 2, sometimes 3.&lt;br&gt;Shown a pair of shoes, says, “Two shoes.”</td>
<td>Gesture to a small group of objects (1 or 2, later 3 when the children are capable). Say, “There are two balls. Two!” When the children are able, ask them how many there are. This should be a natural part of interaction throughout the day. Name collections as “two.” Also include non-examples as well as examples, saying, for instance, “That’s not two. That’s three!” Or, put out three groups of 2 and one group of 3 and have the child find out “the one that is not like the others.” Talk about why.</td>
</tr>
<tr>
<td><strong>Maker of Small Collections</strong>&lt;br&gt;Nonverbally makes a small collection (no more than 4, usually 1–3) with the same number as another collection. Might also be verbal.&lt;br&gt;When shown a collection of 3, makes another collection of 3.</td>
<td>Ask children to get the right number of crackers (etc.) for a small number of children.&lt;br&gt;Lay out a small collection, say 2 blocks. Hide them.&lt;br&gt;Ask children to make a group that has the same number of blocks as your group has. After they have finished, show them your group and ask them if they got the same number. Name the number.&lt;br&gt;In this and every other level, continue to name collections throughout the day. “Would you please put those four books on the shelves?” “Ah, three beautiful flowers.” “Nice design of five squares you made!”</td>
</tr>
<tr>
<td><strong>Perceptual Subitizer to 4</strong>&lt;br&gt;Instantly recognizes collections up to 4 briefly shown and verbally names the number of items.&lt;br&gt;When shown 4 objects briefly, says “four.”</td>
<td>Play “Snapshots” (see the text). At this level, play with collections of 1 to 4 objects, arranged in a line or other simple arrangement, asking children to respond verbally with the number name. Start with the smaller numbers and easier arrangements, (the top row of dots) moving to others as children become competent and confident.</td>
</tr>
</tbody>
</table>
### Developmental Progression

#### Perceptual Subitizer to 5
- Instantly recognizes briefly shown collections up to 5 and verbally names the number of items.
- Shown 5 objects briefly, says "5."

#### Instructional Activities

<table>
<thead>
<tr>
<th><strong>Instructional Activities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Play &quot;Snapshots&quot; on the computer. (a) Children see an arrangement of dots for 2 seconds. (b) They are then asked to click on the corresponding numeral. They can &quot;peek&quot; for 2 more seconds if necessary. (c) They are given feedback verbally and by seeing the dots again.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Instructional Activities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Play &quot;Snapshots&quot; on or off the computer with matching dots to numerals with groups up to and including five.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Instructional Activities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Play &quot;Snapshots&quot; with dot cards, starting with easy arrangements, moving to more difficult arrangements, as children are able.</td>
</tr>
</tbody>
</table>

---

Figure 21.3 (continued) A learning trajectory for subitizing—sample levels (adapted from Clements & Sarama, 2009; Sarama & Clements, 2009).
Developmental Progression

<table>
<thead>
<tr>
<th>Conceptual Subitizer to 5</th>
<th>Instructional Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbally labels all arrangements to about 5, when shown only briefly.</td>
<td></td>
</tr>
<tr>
<td>&quot;5? Why? I saw 3 and 2 and so I said 5.&quot;</td>
<td></td>
</tr>
<tr>
<td>Use different arrangement the various modifications of &quot;Snapshots&quot; to develop conceptual subitizing and ideas about addition and subtraction. The goal is to encouraging students to &quot;see the addends and the sum as in 'two olives and two olives make four olives'&quot; (Fuson, 1992, p. 248).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conceptual Subitizer to 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbally labels most briefly shown arrangements to 6, then up to 10, using groups.</td>
</tr>
<tr>
<td>&quot;In my mind, I made two groups of 3 and one more, so 7.&quot;</td>
</tr>
<tr>
<td>Play &quot;Snapshots&quot; on or off the computer with matching dots to numerals. The computer version's feedback emphasizes that &quot;three and four make seven.&quot;</td>
</tr>
</tbody>
</table>

Figure 21.3 (continued) A learning trajectory for subitizing—sample levels (adapted from Clements & Sarama, 2009; Sarama & Clements, 2009).

The goal, developmental sequence, and instruction make up the complete learning trajectories. Learning trajectories such as these formed the skeleton of the nascent Building Blocks product.

Evaluation

The remaining six phases, in the third category, evaluation, involve collecting specific empirical evidence in marketing, formative, and summative evaluations. The goal is to evaluate the attractiveness, usability, and efficacy of the product, even if it is still in draft form.
Phase 5: Market Research

Market research is consumer-oriented research. Often, it is not done scientifically. For example, publishers may create prototype materials that are presented to “focus groups” in a geographically balanced sample of sites, along with general questions about what they are looking for. Identities and results are hidden—parting with the scientific criterion that methods and results should be in the public view so that the inquiry can be inspected and criticized.

In contrast, collecting useful information about goals, needs, usability and probability of adoption and implementation is important for dissemination and adoption (Tushnet et al., 2000). Our framework includes market research that is scientific; that is, fully grounded in the disciplines, is in the public view, conscientiously documented, and fully reported (Jaeger, 1988). Such market research is conducted at several points in the developmental cycle, from the beginning, as a component of the A Priori Foundations phases, and through the last phase of planning for diffusion (Rogers, 2003). We worked with 35 teachers in developing the Building Blocks products, and reached out to hundreds of others for advice.

Formative Evaluation

The following three phases involve repeated cycles of design, enactment, analysis, and revision (Clements & Battista, 2000), with increasing grain size of the populations and the research variables. The goal is to discover whether the product is usable by, and effective with, various children and teachers.

In formative phases 6 to 8, researchers seek to understand the meaning that both children and teachers give to the product in progressively expanding social contexts. For example, researchers assess the ease of use and efficacy of the parts and attributes of the product as implemented first by a teacher who is familiar with the materials, working with small groups of children (phase 6) and, later, whole classes (phase 7). Later, similar studies are conducted in cooperation with a more diverse group of teachers (phase 8). Methods include interpretive work using a mix of model testing and model generation strategies, including design experiments, microgenetic, microethnographic, and phenomenological approaches (phase 6), classroom-based teaching experiments and ethnographic participant observation (phase 7), and these plus content analyses when appropriate (phase 8). The product is refined based on these studies, especially including issues of support for teachers.

Phase 6: Formative Research: Small Group

This phase involves intensive pilot testing with individuals or small groups of children, often focusing only on one section of the product at a time. Scientific approaches include design experiments, as well as grounded theory, microgenetic, microethnographic, and phenomenological approaches (Siegler & Crowley, 1991; Spradley, 1979; Steffe, Thompson, &
Glasersfeld, 2000; Strauss & Corbin, 1990). The objects of these studies is to gain an understanding of the meaning that children give to the product or the instantiation of the product (e.g., see Lincoln, 1992).

The focus is on the congruity between the actions of the children and the learning model or learning trajectory. If there is a mismatch, then some aspect of the learning trajectory is changed. (This is an advantage of the Framework compared to traditional formative and summative evaluations, which often do not connect to theory and do not typically create new theories, cf. Barab & Squire, 2004.) For example, one asks whether the children use the objects provided (e.g., manipulatives, tables or graphs, software tools or features) to perform the actions they are designed to engender, either spontaneously or with prompting from the teacher (if the later, what type of prompting)? Using the cognitive and learning trajectories as guides, and the tasks as catalysts, the researcher-developer creates more refined models of the thinking and especially the learning of children and particular children or groups of children. At the same time, the researcher-developer describes what elements of the teaching and learning environment, such as teaching strategies or "moves," appear to contribute to learning (D. F. Walker, 1992). The ultimate goal is to connect children's learning processes with specific characteristics of the environment and the teacher's actions, and thus begin to describe the competencies that are expected of the teacher to be effective.

As in all phases, but especially here, equity is a concern (Confrey, 2000). Convenience samples are often inadequate. For example, a product cannot be effectively designed for "all children" or specifically at-risk children if the field testing is done in affluent schools. It is not uncommon to see evaluations in which sites are selected through advertisements, often resulting in samples mostly of white, middle-income, suburban populations.

This may be the most intensive phase of cycling the research and design processes, sometimes as quickly as every twenty-four hours (Burkhardt, Fraser, & Ridgway, 1990; Clements & Sarama, 1995). Refined or newly created activities or approaches might be developed one night and tried the next day. Several classrooms may also be used so that revised lessons can be tested in a different classroom, with one staggered to be from one to five days behind the other in implementing the product (Flagg, 1990). Not only are these activities challenging, but it is easy to fail to perform the necessary documentation that will permit researchers to connect findings to specific revisions of the product. Field notes, audiotapes, and videotapes can help. Computer programs are available that allow researchers to transcribe, code, and analyze such recordings. Technology can also assist in documenting children's ongoing activity, as in solution-path recording (Gerber, Semmel, & Semmel, 1994; Lesh, 1990). Stored solution paths can be re-executed and examined by the teacher, child, or researcher. Such documentation
should be used to evaluate and reflect on those components of the design that were based on intuition, aesthetics, and subconscious beliefs.

Although this phase includes a model-testing approach, there remains significant adaptation to children's actions and their own creative responses to the product. For example, their free exploration of environments and materials may be encouraged and observed before the introduction of any structured activities. As previously stated, one of the beneficial, albeit challenging, features of the proposed research-and-development is that it studies what could be, in contrast to traditional research, which usually investigates what is. The Framework provides an alternative to research that allows, or even encourages, unfortunate confirmation bias and, instead, attempts to invent ways to produce previously unattained results (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Sarama & Clements, 2009, see pp. 363-364 for a description of the problems with confirmation bias in early childhood research).

In summary, research in this phrase has much to offer. Using the learning trajectories as a guide, and the tasks as a catalyst, the researcher-developer creates more refined models of particular children and groups of children. Researchers also learn about the value of various characteristics of the teaching and learning environments, many of which will emerge from interaction of the teacher-developer and the child.

As an example from our Building Blocks project, the “snapshots” activities from our first instructional sequence did not include a second look—another “peek.” We found that young children often needed that repeated exposure to attend, build the image, and generate the quantity. Without such mental activity, some did not progress through the learning trajectory. Therefore, we built it into whole and small group activities, as well as the software (see the blue “Peek” button in the software screens).

As another example, we originally planned an entire separate, related learning trajectory dealing with numerosity estimation. That is, following the subitizing activities in which children named exact quantities, we asked children to estimate the amount in larger sets. We tried a variety of sequences and activities but eventually abandoned this learning trajectory, because children's estimation abilities did not improve. We hypothesized that until exact quantities are well-established, benchmarks for numerosity estimation are too weak to justify the time spent on this learning trajectory in the earliest years. As disappointing as the results seemed at the time (we spent so long developing this learning trajectory!), research at this phase potentially saved many teachers and children from wasting their time on unproductive instructional activities.

One final type of work at this phase is significant. Given the importance yet paucity of child-designed projects, provision for such self-motivated, self-maintained work should not be ignored. Open-ended activities using
the objects and actions should therefore be a part of the design so that the environment can be a setting in which children think creatively. Design activity on the part of children is one for that to happen. In geometry, such activity can be generated, with children producing interesting, relevant, and aesthetically attractive designs (e.g., see Clements & Sarama, 2009). In comparison, design activity with small sets of objects seems difficult or impossible. In our present work, however, we used Donald Crews' book, *Black Dots*, as a starting point, and encouraged children to make their own design with small numbers of black dots. This has been successful.

**Phase 7: Formative Research: Single Classroom**

Teachers are involved in all phases, but this phase includes a special emphasis on the process of enactment (Ball & Cohen, 1996; Dow, 1991; Snyder, Bolin, & Zumwalt, 1992). For example, a goal of the product may be to help teachers interpret children's thinking about the goals or content they are designed to teach; support teachers' learning of the goals and content; and provide support for representing that content (Ball & Cohen, 1996), often in the 100 languages of children (Edwards, Gandini, & Forman, 1993). So, this phase contains two research foci. Classroom-based teaching experiments are used to document and evaluate child development, to understand how children think in learn in a classroom implementing the product (Clements, Battista, Sarama, & Swaminathan, 1996; for examples, see Clements, Battista, Sarama, Swaminathan, & McMillen, 1997; Gravemeijer, 1994a; Pinar, Reynolds, Slautery, & Taubman, 1995). Field notes and often videotapes are used so that children's performances can be examined, often repeatedly, for evidence of their interpretations and learning.

The second focus is on the entire class, as the researchers seeks information about the usability and efficacy of the product. Ethnographic participant observation may used to examine the teacher and children as they interact to build the classroom cultures and learning environments (Spradley, 1980). Observations are on how teachers and children use the materials, how the teacher guides children, what attributes of these interactive environments emerge, and, of course, how these processes are connected to both intended and unintended child outcomes.

During this phase the class may be taught either by a team including one of the researcher developers and the teacher, or by a teacher familiar with and intensively involved in product's development. The goal is to observe learning in the context produced by teachers who can implement the product with high fidelity, consistent with the developers' vision (in contrast to observing how the product is implementing in classrooms in general, which is one focus of the following phase. "High fidelity" does not necessary following a script. Many pedagogical approaches are not implemented with fidelity to the creator's vision without creative and adaptive enactment. In
other words, the philosophy of the product and of the researchers influence the interpretation of fidelity on a continuum from compliance to the creative implementation and adaptation of an individual of particular educational vision.

Whatever the position along this continuum, this phase seeks “super-realization” (Cronbach et al., 1980)—a painstaking assessment of what the product can accomplish “at its best.” This usually implies frequent meetings of teachers and researchers. Video and written records can serve both as research evidence and useful “existence proofs” that are effective complements to other research data for researchers, and especially practitioners and policy makers. The end results of these efforts is a better-developed draft of the product, along with measures of child outcomes and fidelity of implementation (Snyder et al., 1992).

As stated, this pilot test stage involves teachers working closely with the researcher-developers. The class is taught either by a team including one of the researcher-developers and the teacher, or by a teacher familiar with and intensively involved in curricula development. The Building Blocks project included several tests of learning trajectories (e.g., the full subitizing learning trajectory) in this phase (e.g., Clements & Sarama, 2004a).

**Phase 8: Formative Research: Multiple Classrooms**

Building on the previous phase, here several classrooms are observed for information about the efficacy and ease of implementation of the product. The focus turns more to conditions under which the product is more or less effective, and how it might be altered or complemented to better serve any conditions in which it was not as effective. Too often, innovative materials provide less support for teachers relative to their need; that is, because the approach is new, more support is needed. The first of three main research questions for this phase, then, is whether the supporting materials are adequate in supporting multiple contexts, modes of instruction (e.g., whole groups, small groups, centers, incidental and informal interactions), and styles of management and teaching. Addressing this question goes beyond evaluating and increasing a product’s effectiveness—by employing strategies of condition seeking, it extends the research program’s inoculation against the unfortunate phenomenon that we mentioned previously, confirmation bias (Greenwald et al., 1986). That is, by trying to fail (e.g., finding contexts or populations for which the product is less successful), researchers identify the limiting, necessary, and sufficient conditions and may learn how to be successful (often where few were successful previously). In so doing, they extend theory, effectiveness, and guidance to future design and empirical research work. Collaborative work with others, especially those not previously involved in the development (and thus not ego-involved in the product) can also help.
Methods for Developing Scientific Education

A second question is whether the product supports teachers if they desire to learn more about their children's thinking and then teach them differently. A third question asks which contextual factors support productive adaptations and which allow lethal mutations (Brown & Campione, 1996) and why, as well as how, the product might be changed to facilitate the former, which are especially valuable for formative assessment, and eliminate the latter. As learning trajectories in curricula or programs are actually hypothetical learning trajectories (M. A. Simon, 1995) that must be realized or coconstructed in each classroom, so too is a product a hypothetical path to teaching and learning that is sensitive to local contexts (Herbst, 2003). Modification are not expected to make products “fool-proof” but rather support is provided for as wide a variety of contexts as possible.

Ethnographic research (Spradley, 1979, 1980) is again important in this phase, because teachers may agree with the product’s goals and approach but their implementation of these may not be consistent with the researcher developers’ vision (Sarama, Clements, & Henry, 1998). This phase should determine the meaning that the various materials have for both teachers and children. Professional development approaches and materials may be created, or revised, based on this research evidence, and assessment instrumentals for the future summative evaluations may be revised and validated. In addition, qualitative methods may uncover previously ignored factors (variables) that provide a better explanation for a product’s effects and indicate what design features may provide a more efficacious product. Finally, another set of content analyses may inform revisions to the product before summative evaluations begin, ideally conducted by multiple experts from different perspectives using approved procedures (National Research Council, 2004). Our work at these levels is too extensive to summarize here (but see Clements & Sarama, 2004a, 2006; Sarama, 2004).

Summative Evaluation

In these final two phases, researchers determine the effectiveness of the product, now in its complete form, as it is implemented in realistic contexts. Summative phases 9 and 10 both use cluster randomized field trials. They differ in scale. That is, phase 10 involves a few classrooms, whereas phase 10 examines the fidelity or enactment, and sustainability, of the product when implemented on a large scale, and the critical contextual and implementation variables that influence its effectiveness.

In both phases, experimental or carefully planned quasi-experimental designs, incorporating observational measures and surveys, are useful for generating political and public support, as well as for their research advantages. The main such advantage is the experiments provide the most efficient and least biased designs to assess causal relationships (Cook, 2002). In addition, qualitative approaches continue to be useful for dealing with the
complexity and indeterminateness of educational activity (Lester & Wiliam, 2002). This mixed methods approach, synthesizing the two approaches, is a powerful pairing.

The cluster randomized design requires that the product is well described and able to be implemented with fidelity. Also, the curricula or practices used in the comparison classrooms should be fully and explicitly described, and ideally selected on a principled basis. To do so, the quantity and quality of the environment and teaching must be measured in all participating classrooms. Experiments should be designed to have greater explanatory power by connecting specific processes and contexts to outcomes so that moderating and mediating variables are identified (Cook, 2002). Finally, if quasi-experiment designs only are possible, careful consideration of bias must be conducted to ensure comparability (e.g., of children, teachers, and classroom contexts, National Research Council, 2004).

**Phase 9: Summative Research: Small Scale**

In this phase, researchers evaluate what can actually be achieved with typical teachers under realistic circumstances (Burkhardt et al., 1990; Rogers, 2003). In a few classrooms, from about 4 to about 10, researchers conduct pre- and posttest cluster randomized experimental designs using measures of learning. As stated, experiments are conducted in to complement methodologies previously described. Qualitative work is stronger if conducted within the context of a randomized experiment. For example, if teachers volunteer to implement the product in a quasi-experimental design, neither quantitative nor qualitative techniques alone will easily discriminate between the effects of the implementation of the product and the teachers' dispositions and knowledge that led to their decisions to volunteer.

Surveys and interviews of teacher participants also may be used to compare data collected before and after they have used the product, as well as to collect such data as teacher's background, professional development, and resources. The combined interpretive and survey information also evaluates whether supports are viewed as adequate by teachers and whether their teaching practices have been influenced. Do before-and-after comparisons indicate that they have learned about children's thinking in specific subject matter domains and adopted new teaching practices? Have they changed previous approaches to teaching and assessment of the subject matter?

Such research is similar to, but differs from, traditional summative evaluations. A theoretical frame is essential; comparison of scores outside theory, permitted in traditional evaluation, is inadequate. A related point is that the comparison curriculum or practices must be selected deliberately, to focus on specific research issues. Further, connecting the product's objects and activities and the processes of enactment, including all components of the implementation, to the outcomes is important for theoretical,
development, and practical reasons. Variables from the broader data collected should be linked to child outcomes. Links also should be made across experimental and comparison classrooms. Without such, there is an inadequate basis for contributing to theories of learning and teaching in complex settings, guiding future research as well as implementations of the product in various contexts. Finally, statistical analyses performed on the appropriate unit of analysis, often the classroom or school, should allow making those connections (National Research Council, 2004) and provide estimates of the efficacy of the product expressed as effect sizes.

The first summative (phase 9) evaluation of Building Blocks resulted in significant differences, with effect sizes of 1.71 for number and 2.12 for geometry (Clements & Sarama, 2007b). Effect sizes of the first of two large-scale evaluations (phase 10) ranged from .46 (compared to another research-based curriculum) to 1.11 (compared to a "home grown" control curriculum). Achievement gains of the experimental group were thus comparable to the sought-after 2-sigma effect of individual tutoring (Bloom, 1984).

**Phase 10: Summative Research: Large Scale**

Commonly known is the "deep, systemic incapacity of U.S. schools, and the practitioners who work in them, to develop, incorporate, and extend new ideas about teaching and learning in anything but a small fraction of schools and classrooms" (see also Berends, Kirby, Nettel, & McKelvey, 2001; Cuban, 2001; Elmore, 1996, p. 1). Thus, with any product, but especially one that differs from tradition, evaluations must be conducted on a large scale (after considering issues of ethics and practical consequences, see Lester & William, 2002; Schwandt, 2002). Such research should use a broad set of instruments to assess the impact of the implementation on participating children, teachers, program administrators, and parents, as well as document the fidelity of the implementation and effects of the product across diverse contexts (from Clements, 2007). That is, unlike the treatment standardization necessary to answer the questions of previous phases, here it is assumed that implementation fidelity will vary (often widely, with research indicating that people who take advantage of all program components are more likely to benefit, Ramey & Ramey, 1998), with the questions centering around the product's likely effects in settings where standard implementation cannot be guaranteed (Cook, 2002).

A related goal is to measure and analyze the critical variables, including contextual variables (e.g., settings, such as urban/suburban/rural; type of program; class size; teacher characteristics; child/family characteristics) and implementation variables (e.g., engagement in professional development opportunities; fidelity of implementation; leadership, such as principal leadership, as well as support and availability of resources, funds,
and time; peer relations at the school; "convergent perspectives" of the researcher developers, school administrators, and teachers in a cohort; and incentives used) (Berends et al., 2001; Cohen, 1996; Elmore, 1996; Fullan, 1992; Mohrman & Lawler III, 1996; Sarama et al., 1998; Weiss, 2002). A randomized experiment provides an assessment of the average impact of exposure to a product. A series of analyses (e.g., hierarchical linear modeling, or HLM, that provide correct estimates of effects and standard errors when the data are collected at several levels; that is, repeated observations nested within individual children, children nested within classrooms) relate outcome measures with a set of target contextual and implementation variables, critical for identifying moderating and mediating variables (appropriate units of analysis—such as the class—should be defined and should be identical to the unit used for random assignment). Ideally, because no set of experimental variables is complete or appropriate for each situation, qualitative inquiries supplement these analyses. From the wide breadth of documents, including field notes, theoretical notes (methodological and personal journals), drafts of research literature syntheses, and the like, researchers conduct iterative analyses, to determine the significant meanings, relationships, and critical variables that affect implementation and effectiveness (Lincoln & Guba, 1985) and thus meaningfully connect implementation processes to learning outcomes.

Finally, summative evaluations are not complete until two criteria are met. First, the product must be sustained and evaluated in multiple sites for more than two years, with full documentation of the contextual and implementation variables, including practical requirements, procedures, and costs (Berends et al., 2001; Bodilly, 1998; Borman, Hewes, Overman, & Brown, 2003; Fishman et al., 2004; Fullan, 1992). Second, evaluations must be confirmed by researchers unrelated to the developers of the product (Darling-Hammond & Snyder, 1992), with attention given to issues of adoption and diffusion of the product (Fishman et al., 2004; Rogers, 2003; Zaritsky, Kelly, Flowers, Rogers, & O’Neil, 2003). The large expense and effort involved in meeting these criteria is another reason that previous evaluation phases should be employed first; only effective programs should be scaled up.

Given this variety of possibilities, claims that a product is based on research should be questioned to reveal the nature and extent of the connection between the two, including the specific phases used of the ten described and the results obtained with each.

We built a new model to scale up. The TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development) model has the goal of increasing math achievement in young children, especially those at risk, by means of a high-quality implementation of the Building Blocks, with all aspects of the curriculum—mathematical content,
Methods for Developing Scientific Education • 741

pedagogy, teacher’s guide, technology, and assessments—based on a common core of learning trajectories. For example, we performed a “gold standard” Randomized Cluster Trial (RCT) in three states. Forty-two schools serving low-resource communities were randomly selected and randomly assigned to three treatment groups using a randomized block design involving 1,375 preschoolers in 106 classrooms. Teachers implemented the intervention with adequate fidelity. Pre- to posttest scores revealed that the children in the TRIAD/BUILDING BLOCKS group learned more mathematics than the children in the control group (effect size, $g = 0.72$, Clements & Sarama, 2007b). They also developed better language competencies (e.g., effect sizes ranging from .16 to .36, Sarama, Lange, Clements, & Wolfe, 2012).

CONCLUSIONS AND IMPLICATIONS

In this final section, we describe ramifications of our Framework. First, theoretical purity is less important than a consideration of all relevant theories and empirical work. The complexity of the field often creates a Babel of disciplines (Latour, 1987) in which the lack of communication prevents progress. This is one conceit researcher developers can ill afford. Instead, they must meld academic issues and practical teaching demands no less than a serious consideration of what researchers and teachers from other philosophical positions experience and report. This does not imply inconsistent positions. It does imply that overzealous applications (often misinterpretations and overgeneralizations) can limit practical effectiveness. As merely one illustration, constructivism does not imply that practice is not necessary and does not dictate specific pedagogical practices (Clements, 1997; M. A. Simon, 1995).

Second, particular research designs and methods are suited for specific kinds of investigations and questions, but can rarely illuminate all the questions and issues in a line of inquiry (cf. National Research Council, 2002, p. 4; 2004). This is why different methods are used in various phases of the Framework (Clements, 2007). For example, although iterating through one or two of the phases might lead to an effective product and high-quality research, this would not meet all the goals of an integrated research and development program. As a simple example, the curriculum might be effective in some settings, but not others, or it might be too difficult to scale up. Moreover, we would not know why the curriculum is effective.

Third, the Framework is resource intensive. Some might argue that using multiple stages and phases are logistically or practically infeasible. Just producing satisfactory evaluation data (National Research Council, 2004) is costly. Consider, with the hundreds of millions of dollars undoubtedly spent on developing and testing products without it impracticable to use
the proposed framework? We argue, paradoxically, that it is impractical to spend such sums without using it.

Fourth, the education community should support and heed the results of research frameworks such as the one proposed. Given the grounding in both comprehensive research and classroom experience, the curricular products and empirical findings of such integrated research and development programs should be implemented in classrooms. Researcher developers should follow models and base their development on the findings and lessons learned from these projects. Administrators and policy makers should accept and promote curricula based upon similar research-based models. Educators at all levels should eschew software that is not developed consonant with research on children’s learning and that does not have the support of empirical evaluation. This would eliminate much of what is presently used in classrooms. This is a strong position, but one that may avoid a backlash against the use of computers in education, and the use of innovative curricula in general, and that will, we believe, ultimately benefit children.

Fortunately, the design models discussed here, with their tight cycles of planning, instruction, and analysis, are consistent with the practices of teachers who develop broad conceptual and procedural knowledge in their children (Cobb, 2001; Lampert, 1988; M. A. Simon, 1995; Stigler & Hiebert, 1999). Therefore, the product and findings are not only applicable to other classrooms but also support exactly those practices.

Fifth, and in a similar vein, universities should legitimize research programs such as these. There is a long history of bias against design sciences. As professional schools, including the independent engineering schools, are more and more absorbed into the general culture of the university, they hanker after academic respectability. In terms of the prevailing norms, academic respectability calls for subject matter that is intellectually tough, analytic, formalizable, and teachable. In the past, much, if not most, of what we knew about design and about the artificial sciences was intellectually soft, intuitive, informal, and cookbooky. Why would anyone in a university stoop to teach or learn about designing machines or planning market strategies when he could concern himself with solid-state physics? The answer has been clear: he usually wouldn’l (H. A. Simon, 1969, pp. 56-57)

In particular, the more that schools of education in prestigious research universities “have rowed toward the shores of scholarly research the more distant they have become form the public schools they are bound to serve” (Clifford & Guthrie, 1988, p. 3). This is a dangerous prejudice, and one that we should resist. Education might be seen largely as a design science, with a unique status and autonomy (Wittmann, 1995). “Attempts to organize... education by using related disciplines as models miss the point because they overlook the overriding importance of creative design for conceptual and
practical innovations” (Wittmann, 1995, p. 363). The converse of this argument is that universities benefit because the approaches described here will prove practically useful, they will legitimize academic research per se.

In summary, traditional research is conservative; it studies “what is” rather than “what could be.” When research is an integral component of the design process, when it helps uncover and invent models of children’s thinking and builds these into a creative product, then research moves to the vanguard in innovation and reform of education.

NOTES

1. This paper was supported in part by the Institute of Educational Sciences (U.S. Department of Education) under Grants No. R305K05157 and R305A110188 and by the National Science Foundation, under grant No. DRL-1020118 and by the James C. Kennedy Institute for Educational Success and the Marsico Institute for Early Learning and Literacy at the Morgridge College of Education. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

2. Rather, subitizing is amenable to a simple presentation. A complete account, including the multiple theories and studies on the innate processes that underlie it, the role of learning and development (including language) and so forth, would be chapter or even book length (Sarama & Clements, 2009). Not discussed here, but represented somewhat is Figure 3, are such features as arrangement of objects and even the type of object (visual, auditory, etc.) that can be subitized. Thus, none of this is actually “simple,” merely simplified for our purposes here.

REFERENCES


Duckworth, E. (1979). Either we’re too early and they can’t learn it or we’re too late and they know it already: The dilemma of “applying Piaget.” Harvard Educational Review, 49, 297–312.


