Journal for Research in Mathematics Education
2007, Vol. 38, No. 1, 35–70

Curriculum Research: Toward a Framework for "Research-based Curricula"

Douglas H. Clements
University at Buffalo, State University of New York

Government agencies and members of the educational research community have petitioned for research-based curricula. The ambiguity of the phrase "research-based," however, undermines attempts to create a shared research foundation for the development of, and informed choices about, classroom curricula. This article presents a framework for the construct of research-based curricula. One implication is that traditional strategies such as market research and research-to-practice models are insufficient; more adequate is the use of multiple phases of the proffered Curriculum Research Framework.

Key words: Curriculum; Evaluation; Instructional intervention; Large scale studies; Naturalistic/ethnographic methods; Program/project assessment; Qualitative methods; Quasi-experimental design

Government agencies have recently emphasized the importance of evidence-based instructional materials.[1] It would be reasonable to assume that such evidence is easily available, because developers and publishers frequently characterize their curricula as based on research. However, the ubiquity and multifariousness of such characterizations, in conjunction with the ambiguous nature of the phrase

[1] For example, see (Feuer, Towne, & Shavelson, 2002; President's Committee of Advisors on Science and Technology—Panel on Educational Technology, 1997), the "No Child Left Behind" Act of 2001, signed into law by President Bush in January (Reeves, 2002, reports this act uses the term "scientific" or "scientifically" 114 times and the word "research" 246 times), the U.S. Dept. of Education calls for increasing randomized trials to 75% of all research studies (www.ed.gov/about/reports/strat/plan2002-07), Interagency Education Research Initiative (www.nsf.gov/pubs/2002/nsf02062/nsf02062.html), or the curriculum documents from adoption states such as Florida (see their "Major Priorities for Instructional Materials" at http://www.firn.edu/doe/instmat/home0015.htm). Of course, research reviews emphasize the need for scientific research as well (e.g., Kilpatrick, Swafford, & Findell, 2001; Walker, 1992).

This article was supported in part by the National Science Foundation under Grant No. ESI-9730804, "Building Blocks—Foundations for Mathematical Thinking, Pre-Kindergarten to Grade 2: Research-based Materials Development" and by the Institute of Educational Sciences (U.S. DOE, under the Interagency Educational Research Initiative, or IERI, a collaboration of the IES, NSF, and NICHHD) under Grant No. R305K05157 to D. H. Clements, J. Sarama, and J. Lee, "Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.

"research-based," discourages scientific approaches to curriculum development (allowing the continued dominance of nonscientific "market research") and undermines attempts to create a shared research foundation for the creation of, and informed choices about, classroom curricula. Describing and categorizing possible research bases for curriculum development and evaluation is a necessary first step in ameliorating these problems. The purposes of this article are to propose a framework for the construct of "research-based curricula" in mathematics and to discuss the ramifications for multiple relevant parties, including practitioners, curriculum developers, researchers, administrators, funding agencies, and policymakers.

CURRICULUM AND SCIENTIFIC RESEARCH

"Curriculum" has different meanings in different contexts (Beauchamp, 1986; Jackson, 1992; Pinar, Reynolds, Slattery, & Taubman, 1995; Walker, 2003). Although there are many definitions, there are only a few substantive distinctions among them (Jackson, 1992). This article focuses on curriculum as a specific set of instructional materials that order content used to support pre-K–grade 12 classroom instruction—what is often called the "available curriculum" (or potentially implemented curriculum, Schmidt et al., 2001), in contrast to the ideal, adopted, implemented, achieved, or tested curriculum (Burkhardt, Fraser, & Ridgway, 1990, pp. 5–6). Because my usage corresponds with historical (Beauchamp, 1981; Dewey, 1902/1976) and common uses as an available "course of study," reflected in dictionary definitions (Goodlad & Associates, 1979; Jackson, 1992), I shall refer to it hereafter without appending the adjective "available." In this meaning, curriculum is a written instructional blueprint and set of materials for guiding students' acquisition of certain culturally valued concepts, procedures, intellectual dispositions, and ways of reasoning (Battista & Clements, 2000; Beauchamp, 1981). The focus of the framework presented here is on the design and evaluation of a specific curriculum and thus involves one subtheory of curriculum theory (Beauchamp, 1981). As will be argued, basing curricula on scientific knowledge focuses the meaning considerably.

The isolation of curriculum development and educational research vitiates both (Clements & Battista, 2000; Clements, Battista, Sarama, & Swaminathan, 1997a; Lagemann, 1997; Sarama & Clements, in press). The two remain distinct: The goal of scientific research is the creation of knowledge, whereas the goal of curriculum development is the production of instructional materials. However,

---

Drafts of this article were presented at the National Clearinghouse for Comprehensive School Reform Annual Meeting on Comprehensive School Reform, Washington, DC, June 29, 2004, and at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 2004. The ideas expressed here were developed and tested in collaboration with Julie Sarama. Appreciation is expressed to Frank Lester, Martin A. Simon, Alan Schoenfeld, and Leslie Steffe for their comments on early drafts.

the minimal connection between them is one reason curriculum development in the United States does not reliably improve (Battista & Clements, 2000; Clements, 2002; Clements & Battista, 2000). For example, although knowledge is usually created during curriculum development, this knowledge is seldom explicated or published and thus is unavailable to the educational community (Gravemeijer, 1994b).

Scientific knowledge is valued because it offers reliable, self-correcting, documented, shared knowledge based on research methodology (NRC, 2002). Curriculum development can be a design science (Brown, 1992; Simon, 1969; Wittmann, 1995), with the dual goals of engineering a learning process and developing local theories (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). As a science, knowledge created during curriculum development should be both generated and placed within a scientific research corpus, peer reviewed, and published. Because scientific advances are ultimately achieved by the "self-regulating norms" of a scientific community over time, the goal cannot be the development of a single "ideal" curriculum 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 curricula should be based on research—as defined here. That is, all research is social and political (Latour, 1987), with researchers garnering support for their global perspectives, research issues, studies, and results, and thus is not free from social-historical movements, values, controversies, politics, competition, status hierarchies, and egotism. Because these factors affect research on curriculum, particularly in the realm of financial gain, the checks and balances of scientific research are essential to support full disclosure as well as progress.

Finally, curriculum research should not be limited to research-to-practice strategies. Similar strategies are included in the proposed framework. However, because any model limited to research-to-practice strategies constitutes a one-way translation of research results, it is flawed in its presumptions, insensitive to changing goals in the content area, and unable to contribute to a revision of the theory and knowledge on which it is built—the second critical goal of a scientific curriculum research program. Instead, a valid scientific curriculum development program should address the basic issues of *effect* and *conditions* across the three domains of practice, policy, and theory, as described in Table 1. To achieve these goals satisfactorily and scientifically, developers must draw from existing research so that what is already known can be applied to the anticipated curriculum; structure and revise the nature and content of curricular 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. Thus, research should be present in all phases of the curriculum development and research process, from James' (1958) initial scientific base to formative and summative evaluation (Brown, 1992), and thus be integrated into even the most creative processes (Dewey, 1929) to achieve the documentation of decisions and the ultimate checking of hunches and full reporting of all procedures (Cronbach

& Suppes, 1969). Such documentation requires a common language for connections between curriculum development and research.[2]

Although research literatures exist on the methods of various components of the framework described in this article, no single methodology encompasses its scope. For example, design experiments (Brown, 1992; Cobb et al., 2003; The Design-Based Research Collective, 2003), developed as a way to conduct formative research to test and iteratively refine educational designs based on principles derived from previous research (Collins, Joseph, & Bielaczyc, 2004), provide a theoretical basis for several components of development. However, design experiments are often limited to pilot or field testing (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004; NRC Committee, 2004, p. 75), have less emphasis on the development of curriculum per se, and do not adequately address the full range of questions or methods of the proposed framework. (Unknown to us until recently is the work of Bannan-Ritland, 2003, who proposes a wider framework, bringing the stage models from engineering design to educational research.) The emphasis in design experiments on theory and model development is important, but the proposed framework's main goals are the production of an effective curriculum and educational research answering a comprehensive set of questions (Table 1). The

[2] In some circumstances, other types of inquiry, such as historical research, will be required (Darling-Hammond & Snyder, 1992). In addition, a focus on scientific research should not be misconstrued as minimizing the relevance of approaches such as those taking aesthetic (Eisner, 1998), literary criticism (Papert, 1987), narrative (Bruner, 1986), phenomenological (Pinar et al., 1995), or humanistic (Schwandt, 2002) perspectives (Walker, 1992, argues that humanistic approaches would make greater contributions if they were more specific and thorough). Such approaches would complement the scientific research methods described here. Of course, no single scientific finding or set of findings should dictate pedagogy: "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 early formulation, our framework for curriculum development research rejects strict "rules" but values scientific research for its practical, and political, utility. Although the recent hermeneutic trend in the field of evaluation are valuable and complementary, the logic of practical wisdom (Schwandt, 2002), which rejects evaluating a published curriculum as defined here and focuses only on "lived human practice," "embraces the inherent ambiguity of life," and eschews scientific knowledge for "practical wisdom" (p. 12), cannot (is not designed) to answer the full suite of questions *as posed* (developing and evaluating a curriculum object that is to be widely disseminated), especially those of policy, outlined in Table 1, and so, at least at present, will not address the previously described needs of practitioners, publishers, and government agencies (NRC, 2002). Meeting such needs, in politically charged environs in which decisions have substantive financial and social ramifications, requires the reliable, self-correcting, documented, shared knowledge of scientific research. (Consistencies and the necessity of cross-fertilization between Schwandt's recommendations and the proposed framework are nevertheless numerous; i.e., the proposed framework was not designed to address the complete, complex field of curriculum theory and research, but is posited as a framework for including scientific research in curriculum development programs.) Finally, societal values and goals are substantive components of any curriculum (Confrey, 1996; Hiebert, 1999; NRC, 2002; Schwandt, 2002; Tyler, 1949); curriculum research cannot ignore or determine these components (Lester & Wiliam, 2002; Schwandt, 2002). Determining goals thus requires a dialectical process among all legitimate direct and indirect stakeholders (van Oers, 2003). Unlike groups such as the reconceptualists and poststructuralists (Pinar et al., 1995; Walker, 2003), however, I acknowledge limitations of science without rejecting its fundamental role.

Table 1
Goals of Curriculum Research

	Practice	Policy	Theory
Effects	a. Is the curriculum effective in helping children achieve specific learning goals? Are the intended and unintended consequences positive for children? (What is the quality of the evidence?— Construct and internal validity.) b. Is there credible documentation of both a priori research and research performed on the curriculum indicating the efficacy of the approach as compared to alternative approaches?	c. Are the curriculum goals important? d. What is the effect size for students? e. What effects does it have on teachers?	f. Why is the curriculum effective? g. What were the theoretical bases? h. What cognitive changes occurred and what processes were responsible? That is, what specific components and features (e.g., instructional procedures, materials) account for its impact and why?
Conditions	i. When and where? Under what conditions is the curriculum effective? (Do findings generalize?— External validity.)	j. What are the support requirements for various contexts?	k. Why do certain sets of conditions decrease or increase the curriculum's effectiveness? l. How do specific strategies produce previously unattained results and why?

recent NRC report on evaluating curricular effectiveness (NRC Committee, 2004), is consistent with several components of the proposed framework but did not focus on either curriculum development or formative evaluation. My position is that work using such methods as teaching experiments, design experiments, and curriculum evaluation should be synthesized into a coherent, complete curriculum framework.

The remainder of this article describes a framework for the development, study, and evaluation of research-based curricula. I first describe the framework, including its three categories of activities and 10 phases that are embedded within those categories. I then briefly review the relationship between this framework and extant mathematics curricula. The last two sections draw implications, suggest several caveats, and provide conclusions.

RESEARCH BASES FOR CURRICULA: A FRAMEWORK

Establishing, maintaining, and evaluating connections between curricula and research are problematic because many, if not most, developers and publishers claim to have based their curricula on research, but few fully explicate the claims. Without an established framework for understanding or evaluating these claims, educators turn to other criteria in developing and selecting curricula, and the potential for curriculum development and evaluation to build a coherent scientific knowledge base is left unrealized. I propose a *Curriculum Research Framework (CRF)* that builds upon many elements of previous works (e.g., Beauchamp, 1981; Clements & Battista, 2000; Cobb et al., 2003; Jackson, 1992; Tyler, 1949). The CRF specifies research methods in place of several nonscientific procedures and provides a coherent structure for development and evaluation in place of useful but separate techniques. As an example of the latter, Walker (1992) advocated strategies such as "simple, quick" field tests, which are practicable in classrooms and provide feedback to developers. I agree with these goals but contend that we also must contribute to theoretical and empirical work. To do so, we need to answer the questions in Table 1 within a research framework, with the goal of syncretizing the development of curricula, theories, empirical data, and implications (that communicate with researchers, designers, and practitioners). Further, I propose that curriculum research as described here provides an ideal context for building a scientific knowledge base for education and educational reform. The CRF includes 10 phases of the curriculum development research process that would warrant the claim that a curriculum is based on research. These 10 phases are classified into three categories (reflecting the three categories of knowledge required to meet Table 1's goals), as outlined in Table 2. The following sections describe the CRF's cyclic phases.

A Priori Foundations

1. *Subject Matter A Priori Foundation*. Establishing educational goals involves multiple considerations, not all of which involve scientific knowledge (recall footnote 2). This research phase contributes to the process by using scientific procedures to identify subject-matter content that is valid within the discipline and makes a substantive contribution to the mathematical development of students in the target population (cf. Tyler, 1949). That is, concepts and procedures of the domain should play a central role in the subject-matter domain per se (Tyler, 1949), build from the students' past and present experiences (Dewey, 1902/1976), and be generative in students' development of future understanding (for an explication and examples, see Clements, Sarama, & DiBiase, 2004).[3] Further, research on complementary components of competence should be considered, such as problem posing and problem solving, metacognition, and a positive disposition toward learning and using the subject-matter content (Baroody with Coslick, 1998; Schoenfeld, 2002). The NCTM *Standards* (2000) and *Curriculum Focal Points* (2006) were created by a dialectical process among many legitimate stakeholders and thus serve as a valuable starting point, as are comparisons to other successful curricula. These are scientific research-oriented strategies that constitute part of comprehensive content analyses (cf. NRC Committee, 2004). This phase does not determine a particular pedagogical approach, but the reviews should encompass valid and reliable measures. Ideally, one member of the research team is respon-

[3] There is a presage of the enormity of the challenge for the research community; for example, although large studies such as TIMSS and NAEP contribute to identifying areas of strengths and weaknesses, the generativity criterion requires extensive longitudinal work.

sible, in this and other phases, for taking a perspective of "standing outside," observing and documenting the curriculum development and research team's activities, decisions, and reasons for decisions (Lesh & Kelly, 2000).

2. *General A Priori Foundation*. Broad philosophies, theories, and empirical results on teaching and general curriculum issues are reviewed. For example, developers might start from an Ausubelian or "constructivist" perspective and proceed in any of several directions (Forman, 1993; Lawton, 1993). In addition, curriculum theory and research offer perspectives on students' and teachers' experiences with curricula, as well as on school and society (e.g., concerns for equity), that help establish general goals and directions (Pinar et al., 1995).

3. *Pedagogical A Priori Foundation*. Empirical findings on making specific types of activities educationally effective—motivating and efficacious—are reviewed to create general guidelines for the generation of activities. As one example, in designing software for young children, we consulted empirical data on features that appeared to make computer programs motivating (Escobedo & Evans, 1997; Lahm, 1996; Shade, 1994) and effective (Childers, 1989; Clements & Sarama, 1998; Lavin & Sanders, 1983; Murphy & Appel, 1984; Sarama, Clements, & Vukelic, 1996). Pedagogical strategies and curriculum structure are not determined fully by this line of reasoning, of course; intuition, and the art of teaching play roles (Confrey, 1996; 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, 1958, p. 24)

James treats research as an a priori foundation only—appropriate for this category (indeed, it can play a major contributing role, Tamir, 1988), but not encompassing the other categories.

Learning Model

The second category emphasizes *learning models*. Here, a tenet of the CRF comes into sharp focus: Although the CRF can be discussed in general, both the instantiations and the correlated research are inextricably based in subject matter content, which cannot simply be added post hoc to a general predetermined structure.

4. *Structure According to Specific Learning Models*. Activities are structured in accordance with domain-specific models of learning.[4] This might involve two interrelated aspects. First, activities may be designed to be consistent with empir-

---

[4] Design includes its own theories and processes. Examples are presented here only briefly (e.g., see Clements & Battista, 2000; Clements, Meredith, & Battista, 1992; Clements & Sarama, in press). The intent here is to present a curriculum *research* framework for the instantiation of different specific design models, some of which may be complementary or competitive (see, e.g., Bannan-Ritland, 2003; Cobb et al., 2003; The Design-Based Research Collective, 2003; Zaritsky, Kelly, Flowers, Rogers, & O'Neil, 2003).

Table 2
Categories and Phases of the Curriculum Research Framework (CRF)

Categories	Questions asked	Phases
A Priori Foundations. In variants of the research-to-practice model, extant research is reviewed and implications for the nascent curriculum development effort drawn.	What is already known that can be applied to the anticipated curriculum? Goals* — Phase b c f g — 1 b f g — 2 b f g — 3	Established review procedures (e.g., Light & Pillemer, 1984) and content analyses (NRC Committee, 2004) are employed to garner knowledge concerning the specific subject matter content, including the role it would play in students' development (*phase 1*); general issues concerning psychology, education, and systemic change (*phase 2*); and pedagogy, including the effectiveness of certain types of activities (*phase 3*).
Learning Model. Activities are structured in accordance with empirically based models of children's thinking and learning in the targeted subject-matter domain.	How might the curriculum be constructed to be consistent with models of students' thinking and learning (which are posited to have characteristics and developmental courses that are not arbitrary and therefore not equally amenable to various instructional approaches or curricular routes)? Goals — Phase b f h — 4	In *phase 4*, the nature and content of activities is based on models of children's mathematical thinking and learning (cf. James, 1958; Tyler, 1949). In addition, a set of activities (the hypothetical mechanism of the research) may be sequenced according to specific learning trajectories (Clements & Sarama, 2004c). What distinguishes phase 4 from phase 3, which concerns pedagogical a priori foundations, is not only the focus on the child's learning, rather than teaching strategies alone, but also the iterative nature of its application. That is, in practice, such models are usually applied and revised (or, not infrequently, created anew) dynamically, simultaneously with the development of instructional tasks, using grounded theory methods, clinical interviews, teaching experiments, and design experiments.

Table 2—*Continued*
Categories and Phases of the Curriculum Research Framework (CRF)

Categories	Questions asked	Phases
Evaluation. In these phases, empirical evidence is collected to evaluate the curriculum, realized in some form. The goal is to evaluate the appeal, usability, and effectiveness of an instantiation of the curriculum.	How can market share for the curriculum be maximized? Goals: b c f — Phase: 5	*Phase 5* focuses on marketability, using strategies such as gathering information about mandated educational objectives and surveys of consumers.
	Is the curriculum usable by, and effective with, various student groups and teachers? How can it be improved in these areas or adapted to serve diverse situations and needs? Goals: a b f h k l — Phase: 6 Goals: a b f h j k l — Phase: 7 Goals: a b f i j k l — Phase: 8	Formative *phases 6 to 8* seek to understand the meanings that students and teachers give to the curriculum objects and activities in progressively expanding social contexts; for example, the usability and effectiveness of specific components and characteristics of the curriculum as implemented by a teacher who is familiar with the materials with individuals or small groups (*phase 6*) and whole classes (*phase 7*) and, later, by a 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 *(phase 8)*. The curriculum is altered based on empirical results, with the focus expanding to include aspects of support for teachers.
	What is the effectiveness (e.g., in affecting teaching practices and ultimately student learning) of the curriculum, now in its complete form, as it is implemented in realistic contexts? Goals: a b d f j k l — Phase: 9 Goals: a b c d e f i j k l — Phase: 10	Summative *phases 9* and *10* both use randomized field trials and differ from each other most markedly on the characteristic of scale. That is, *phase 10* examines the fidelity or enactment, and sustainability, of the curriculum when implemented on a large scale, and the critical contextual and implementation variables that influence its effectiveness. 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. In addition, qualitative approaches continue to be useful for dealing with the complexity and indeterminateness of educational activity (Lester & Wiliam, 2002).

* Goals refer to the specific questions in Table 1, answers to which are the goals of the CRF.

ically based models of children's thinking and learning in the targeted subject-matter domain, which can substantially affect curriculum design by focusing it on teaching and learning (Tamir, 1988; Walker, 1992). As an example, based on research that indicates that young children can invent their own solutions to simple arithmetic problems (Baroody, 1987; Carpenter & Moser, 1984; Ginsburg, 1977 Kamii, 1985; Steffe & Cobb, 1988) and profit from doing so more than being introduced to arithmetic by being taught prescriptive procedures (Hiebert et al., 1997; Kamii & Dominick, 1998; Steffe, 1983, 1994), curricula have been crafted that pose problems in the forms of activities and games that ask children to figure out how to solve the problems and explain their solution strategies (Baroody with Coslick, 1998; Everyday Math, see Fuson, Carroll, & Drueck, 2000; Griffin & Case, 1997; Hiebert, 1999; Kamii & Housman, 1999), often using scaffolding techniques to guide their inventions (Mokros, 2003; van den Brink, 1991). As a specific illustration, Fuson (1997) described how a curriculum is based on a model of children's solving of word problems (as well as models of teaching, bilingual language use in word problem solving, and mathematizing children's stories). Briefly, a teacher begins with a story from a child and mathematizes that story to focus on the mathematical elements. Children pose questions and pose word problems as well as solve them. They retell a given story in their own words, as well as representing it through drawings. (In addition, the curriculum moves through increasingly difficult types of word problems based on the model, which anticipates the second aspect.)

Extant models may be available, although they vary in nature and degree of specificity. Especially when details are lacking, developers use grounded theory methods (Strauss & Corbin, 1990) (the methodology of grounded theory can provide critical theoretical bases to work in the early phases) and related methods such as clinical interviews to examine students' knowledge of the content domain, including conceptions, strategies, intuitive ideas, and informal strategies used to solve problems. The researchers set up a situation or task to elicit pertinent concepts and processes. Once a (static) model has been partially developed, it is tested and extended with teaching experiments, which present limited tasks and adult interaction to individual children with the goal of building models of children's thinking and learning (Steffe, Thompson, & Glasersfeld, 2000). Once several iterations of such work reveal no substantive variations, it is accepted as a working model.

Second, sets of activities may be sequenced according to learning trajectories (Simon, 1995) through the concepts and skills that constitute a domain of mathematics (Clements, 2002; Cobb & McClain, 2002; Gravemeijer, 1999). This strategy guides learning to be more effective and efficient and can help avoid the fragmentation common in U.S. textbooks, in which the number of short curricular strands are up to 10 times the potential number of topics (Valverde, Bianchi, Wolfe, Schmidt, & Houang, 2002). Learning trajectories might be based on the historical development of mathematics and observations of children's informal solution strategies (Gravemeijer, 1994b) or emergent mathematical practices of student groups (Cobb & McClain, 2002).

Our CRF emphasizes learning trajectories built upon natural developmental progressions identified in empirically based models of children's thinking and learning (Carpenter & Moser, 1984; Case, 1982; Griffin & Case, 1997; Steffe & Cobb, 1988). These learning trajectories are

> descriptions of children's thinking and learning in a specific mathematical domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking, created with the intent of supporting children's achievement of specific goals in that mathematical domain. (Clements & Sarama, 2004c, p. 83)

An example of such a learning trajectory is young children's development of geometric composition abilities. Research has confirmed a developmental progression in which children move through levels of thinking; from lack of competence in composing geometric shapes, they gain abilities to combine shapes (initially through trial and error and gradually by attributes) into pictures, and finally to synthesize combinations of shapes into new shapes, that is, composite shapes (Clements, Wilson, & Sarama, 2004). (For a description of all components of the learning trajectory, including instructional activities, see Fig. 1 in Clements & Sarama, in press.) The complete learning trajectory includes an explication of the mental constructions (actions-on-objects to meet specific goals or solve specific problems) and patterns of thinking that constitute children's thinking at each level, how they are incorporated in each subsequent level, and tasks aligned to each level (promoting movement to the succeeding level). The learning trajectories construct differs from instructional design based on task analysis because it is based not on a reduction of the skills of experts but on models of children's learning; expects unique constructions and input from children; involves self-reflexive constructivism; and involves continuous, detailed, and simultaneous analyses of goals, pedagogical tasks, teaching, and children's thinking and learning (with cognitive models describing specific processes and concepts involved in the construction of the goal mathematics across several distinct structural levels of increasing sophistication, complexity, abstraction, power, and generality). Such explication allows the researcher to test the theory by testing the curriculum (Clements & Battista, 2000), usually with teaching and design experiments (with the latter emphasizing intervention to support particular forms of learning, Cobb et al., 2003). To be scientific, these experiments must include conceptual analyses and theories that "do real design work in generating, selecting and validating design alternatives at the level at which they are consequential for learning" (diSessa & Cobb, 2004, p. 77).

Evaluation

5. *Market Research*. Market research is consumer-oriented research about the customer and what the customer wants. Because it is arguably the most common type of research in commercial curriculum development, I first consider market

research as typically conducted. (There is also market research that deals with how the publisher will design their message for promoting and selling the materials, which I will not discuss.) Such market research usually involves a close look at state standards, guidelines, and curricula (especially of the key adoption states, such as California, Florida, and Texas) and standardized tests. The publisher often creates prototype materials that are presented to "focus groups" in a geographically balanced sample of sites. These focus groups often are conducted by a separate facility so that the identity of the publisher is hidden. Facility personnel ask focus groups general questions about what they are looking for in a curriculum and specific questions about the prototype. Interviews, and especially large surveys of teachers and administrators, also are performed to seek general information about desired topics, assessments, and features. These strategies are complemented by meetings of the company's sales force, at which participants describe what customers are requesting (often a reaction to the current version of the product). Sometimes a sample chapter is provided to a sample of teachers, who provide feedback via a questionnaire.

Market research as typically conducted fails to meet the standards for scientific research. In contrast, *scientific* market research collects useful information about goals, needs, usability, and probability of adoption and implementation. In the United States, those who ignore concerns of publishers, teachers, and marketability in general often do not achieve wide adoption (Tushnet et al., 2000). To meet the needs of research and marketability, developers form early and sustained relationships with publishers to use findings from, or to conduct, *scientific* market research; that is, inquiry that is fully grounded in the disciplines, is in the public view, and is consciously documented or fully reported (Jaeger, 1988). This has the added advantage of connecting the scientific curriculum research to the types of information with which publishers are most familiar, thus bridging the gap between developers and publishers that is especially common for innovative materials (Tushnet et al., 2000). Such market research is conducted at several points in the developmental cycle, from the beginning, as a component of the *A Priori Foundations* phases, through the last phase of planning for diffusion (Rogers, 2003).

6. *Formative Research: Small Group*. Pilot testing with individuals or small groups of students is conducted on components (e.g., a particular activity, game, or software environment) or on sections of the curriculum. Early interpretive work evaluates components using a mix of model (or hypothesis) testing and model generation strategies, including design experiments, as well as grounded theory, microgenetic, microethnographic, and phenomenological approaches (Ginsburg, 1997; Pinar et al., 1995; Schoenfeld, Smith III, & Arcavi, 1993; Siegler & Crowley, 1991; Spradley, 1979; Steffe et al., 2000; Strauss & Corbin, 1990, note that specific methodologies are proffered as illustrations rather than prescriptions, a point to which I return in the final section). The goal is to understand the meanings that students give to the curriculum objects and tasks (Lincoln, 1992; Pinar et al., 1995).

Evaluating sections of the curriculum focuses on consonance between the actions of the students and the learning model or trajectory. If there are discrepancies, either the model, or the way in which this model is instantiated in the curriculum, should be altered. (This distinguishes this and all subsequent phases from traditional formative and summative evaluations, which do not necessarily connect to theory and do not typically create new theories, cf. Barab & Squire, 2004.) Do students use the tools provided (e.g., manipulatives, tables or graphs, software tools or features) to perform the actions, either spontaneously or with prompting? If the latter, what prompts or scaffolding strategies are successful? In all cases, are students' own actions-on-objects enactments of the desired cognitive operations (Steffe & Wiegel, 1994) in the way the model posits, or merely trial-and-error manipulation? Using the cognitive model and learning trajectories as guides, and the tasks as catalysts, the developer creates more refined models of the thinking of particular groups of students. Simultaneously, the developer describes what elements of the teaching and learning environment, such as specific scaffolding strategies, are observed as having contributed to student learning (Walker, 1992). The theoretical model may involve disequilibrium, modeling, social processes, practice, and combinations of these and other processes. The goal is to connect these processes with specific environmental characteristics and teaching strategies and student learning, and thus describes knowledge and abilities that are expected of the teacher.

As in all phases, equity must be considered (Confrey, 2000; NCTM, 2000). Thought should be given to the students who are envisioned as users and who participate in field tests; a convenience sample is often inappropriate, such as when a curriculum is designed for "all" or specifically at-risk students and yet the field-testing is done in affluent schools. The NRC report (2004) noted that one set of evaluation studies selected sites by advertisements in journals, resulting in samples mostly of white, middle-income, suburban populations. Previous reports' (Confrey, 2000; NRC Committee, 2004) recommendations that evaluations systematically include demographically representative student populations imply the need for appropriate samples in summative research, but the importance of representative populations when the structure and content of curricula are being formed also should be recognized explicitly. Systemic classroom and home participation patterns and sociocultural issues should be considered as well.

Phase 6 is often the most iterative research-design phase; sometimes evaluation and redesign may cycle in quick succession, often as much as every 24 hours (Burkhardt et al., 1990; Char, 1990; Clements & Sarama, 1995; Cobb et al., 2003). Tasks may be completely reconstituted, with edited or newly created ones tried the next day. Several classrooms may also be used so that revised lessons can be tested in a different classroom staggered to be 1–5 days behind in implementing the curriculum (Flagg, 1990).

With so many research and development processes happening, and so many possibilities, extensive documentation is required. Documentation must allow researchers to relate findings to specific components and characteristics of the curriculum. Field notes, and often audiotapes and videotapes (for microgenetic analysis), are collected. Computers might store data documenting students' ongoing activity. Solution-path recording is a particularly useful technique (Gerber, Semmel, & Semmel, 1994; Lesh, 1990). Solution paths can be re-executed and examined by the teacher, student, or researcher (and analyzed in many ways); they also can be modified. Issues such as the efficiency, simplicity, and elegance of particular solutions—even those that result in the same answer—can be assessed (Lesh, 1990). Techniques such as videorecording a mix of two inputs, traditional camera video, and computer screen output serve similar purposes. This 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 students' input. Often, students' free exploration of materials precedes the introduction of activities. In addition, the researcher interprets the contributions of children, and new tasks or questions are posed. One of the welcome but challenging features of curriculum research is that it studies what *could be*, unlike traditional research that tends to investigate *what is*. As such, it presents an invaluable counterpoint to research that invites confirmation bias and, instead, attempts to invent ways to produce previously unattained results (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). In sum, research in this phrase is rich with possibilities. Using the model of mathematics learning as a guide, and the tasks as a catalyst, the developer creates more refined models of particular students' thinking. Also collected is more detailed information about the worth of various features of the teaching and learning interventions, some of which will emerge from, and be mutually constituted by, the developer-teacher and the student. Valuable empirical data may be garnered from the interactions of the students with the tasks (writ large), the software, peers, the teacher-developer, and combinations of them. Developers may be teacher-researchers or engaged participant observers (NRC, 2002). This phase lays the groundwork not only for the final curriculum but also for professional support materials and instrumentation for later phases, such as student achievement and classroom observation measures.

7. *Formative Research: Single Classroom.* Although teachers are ideally involved in all phases of the CRF (in many projects, teachers are a central component of the research-and-development team), a special emphasis here is the process of curricular enactment (Ball & Cohen, 1996; Dow, 1991; Snyder, Bolin, & Zumwalt, 1992). For example, a goal of the curriculum may be to help teachers interpret students' thinking about the tasks and the content they are designed to teach; to support teachers' learning of that content, especially any topics that are new to teachers; and to provide guidance regarding the external representations of content that the materials use (Ball & Cohen, 1996). Thus, there are two research thrusts. First, classroom-based teaching experiments are used to track and evaluate student learning, with the goal of making sense of the curricular activities as they are experienced by individual students (Clements, Battista, Sarama, & Swaminathan, 1996a; for examples, see Clements, Battista, Sarama, Swaminathan, & McMillen, 1997b; Gravemeijer, 1994a; Pinar et al., 1995).

Extensive field notes and often videotapes are required so that students' performances can be examined repeatedly for evidence of their interpretations and learning, for reasons similar to those of the previous phase.

Second and simultaneously, the entire class is observed for information concerning the usability and effectiveness of the curriculum as well as for its character. Ethnographic participant observation is used to examine the teacher and students as they construct new types of classroom cultures and interactions together (Spradley, 1980). Such observation is critical, because events and properties emerge in such interactions that cannot be predicted or understood solely in terms of analyses of the components but must be understood as a complex system (Davis & Simmt, 2003; Herbst, 2003). Thus, the focus is on how the materials are used, how the teacher guides students through the activities, what characteristics emerge in various instantiations of the curriculum (class dynamics cannot be taken as a given; parents and the community are also considered), and, generally, how these processes are connected to both intended and unintended student outcomes.

This phase may involve teachers working closely with the developers. That is, the class may be taught either by a team including one of the developers and the teacher or by a teacher familiar with and intensively involved in curricula development. The goal is to examine learning in the context of the curriculum with teachers who can enact it consonant with the developers' vision, as opposed to ascertaining how the curriculum works in classrooms in general, which is one focus of the following phase. Achieving such initial "fidelity" should not be misinterpreted as following a script; indeed, many pedagogical approaches require creative, adaptive enactment. The philosophical foundations of the curriculum and of the researchers influence the interpretation of fidelity on a continuum from compliance to consonance of an individual enactment to a particular educational vision.

From the chosen perspective, this phase seeks an implementation similar to what Cronbach and others (1980) called a "superrealization"—a painstaking assessment of what the curriculum can accomplish at its best, as a nascent curriculum collaboratively constructed by the developers and teacher. Regular meetings of the teacher and research group are requisite. Written records and videotaping can also be used here as sources of data. Video from this and the following phases can also constitute an existence proof that is a particularly effective complement to other research data for practitioners, policymakers, and researchers. In preparation for the next phase, a near-final draft of the curriculum is completed and project-specific instruments, including measures of student achievement and fidelity of implementation (research on implementation moves from enactment to fidelity perspectives as the research questions change, cf. Snyder et al., 1992) as well as instruments to support qualitative data collection via classroom observation, are formalized.

8. *Formative Research: Multiple Classrooms*. Several classrooms are observed for information about the effectiveness and usability of the curriculum, with an emphasis on the usability and decision-making by such teachers and the conditions under which the curriculum is more or less effective, and how it might be

altered or complemented to better serve the latter conditions. Innovative materials often provide less support for teachers than the traditional materials with which they are familiar (Burkhardt et al., 1990), so such ecological research is especially important for reform curricula. Thus, the first of three main research questions for this phase is whether the supporting materials are flexible enough to support multiple situations, various modes of instruction (e.g., demonstration to a class, class discussion, small group work), and different modes and styles of management (e.g., how teachers track students' progress while using the materials, monitor students' problem solving with the materials, and assess student learning), as well as how the materials might do so better. Addressing this question goes beyond evaluating and increasing a curriculum's effectiveness; by employing strategies of condition seeking, it extends the research program's inoculation against confirmation bias (Greenwald et al., 1986). That is, by *trying to fail*, and thereby identifying the limiting, necessary, and sufficient conditions (and eventually designing to succeed within more configurations of conditions), researchers extend theory, curriculum effectiveness, and guidance to future design and empirical research work. Involving new researchers also helps protect against confirmation bias.

A second question is whether the materials support teachers if they desire to delve more deeply into their students' thinking and then teach differently (Remillard, 2000). A third set of questions ask which contextual factors support productive adaptations and which allow lethal mutations (Brown & Campione, 1996) and why, as well as how, the curriculum might be changed to catalyze the former and minimize the latter. Understanding how and why the curriculum works in various contexts is essential for theory development and for helping practitioners implement the curriculum in their local setting. As learning trajectories in curricula are always *hypothetical* learning trajectories (Simon, 1995) that must be realized in each classroom, so too is a curriculum a hypothetical path to teaching and learning that is sensitive to local contexts and interpretations (Herbst, 2003). No modification can proof a curriculum against such factors; developers provide support for as wide a variety of contexts as possible and document the effects of various contextual and implementation variables.

Again, ethnographic research (Spradley, 1979, 1980) is important, because teachers may agree with the curriculum's goals and approach, but their implementation of these may not be veridical to the developers' vision (Sarama, Clements, & Henry, 1998). This phase should determine the meanings that the various curricular materials have for both teachers and students. Materials for professional development are created, or revised, based on this research, and instrumentation for summative evaluations is revised and validated (e.g., fidelity of implementation measures are used in parallel with qualitative methods and the two are cross-validated; achievement measures are validated). In addition, qualitative methods may uncover previously ignored factors (variables) that provide a better *explanation* for a curriculum's effects and indicate what design features may provide a more efficacious curriculum.

Finally, another round of content analyses should inform revisions to the curriculum before summative evaluations begin. These should be conducted by multiple experts from different perspectives using approved procedures (NRC Committee, 2004).

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). Again in multiple classrooms (2 to about 10), pre- and posttest randomized experimental designs using measures of learning are used. Six issues are common for phases 9 and 10. First, standardized instruments (not necessarily standardized tests as commonly construed) must have been chosen or developed (usually incrementally, as described in the previous phases) as valid measures of the curriculum goals (NRC, 2002; NRC Committee, 2004). Often, this involves at least two assessment components: one that is a valid measure of the shared goals of the experimental and comparison curricula, and one that measures any unique goals of the experimental curriculum (which may involve categorical data; e.g., levels of thinking along a learning trajectory). In both cases, instruments should be sufficiently valid, reliable, and differentiated to measure nuanced differences in various content and process areas. Second, the design requires that the intervention is fully and explicitly described and able to be implemented with fidelity (reliably evaluated according to the definition of fidelity adopted, allowing analysis of data by various curricular components, and recognizing that some curricula may be implemented in nonstandard, but appropriate, ways, and that at the highest levels, the art of teaching does not yield easily to instrumental analysis). Experiments provide the most efficient and least biased designs to assess causal relationships, and most criticisms of them speak to misapplications and misinterpretations (Cook, 2002). For example, recognition that researchers cannot definitively test a theory and that both curriculum and research are social in nature (rejecting logical positivism) does not imply that experiments cannot contribute to evidence on causal claims.

Third, in a similar vein, the curriculum used in the comparison classrooms also should be fully and explicitly described, and ideally selected on a principled basis. Further, the use of a "traditional" curriculum as the only comparison will be less useful than involving a wider variety of comparison curricula, including other innovative curricula, and describing each comparison groups' curriculum and fidelity of implementation (NRC Committee, 2004). Fourth, the quantity and quality of mathematics instruction must be measured in all participating classrooms (e.g., via a classroom observation instrument that measures components such as the classroom culture, including the environment and the personal attributes of the teacher, and specific mathematics lessons, including mathematical focus, organization and teaching approaches, teaching and learning interactions, and assessment and instructional adjustment). Fifth, 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). Sixth and finally, if quasi-experimental designs only are possible, careful consideration of bias must be conducted to ensure comparability (e.g., of students, teachers, and classroom contexts, NRC Committee, 2004).

Experiments are conducted in conjunction with, and to complement, methodologies previously described. Other approaches, including qualitative work, are stronger if conducted within the context of a randomized experiment. For example, if teachers volunteer to implement the curriculum in a quasi-experimental design, neither quantitative nor qualitative techniques alone will easily discriminate between the effects of an intervention and the teachers' dispositions and knowledge that led to their decisions to volunteer.

Surveys of teacher participants also may be used to compare data collected before and after they have used the curriculum, as well as to collect such data as teacher's background, professional development, and resources. The combined interpretive and survey data also address whether supports are viewed as helpful by teachers and other caretakers 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 framework is essential; comparison of scores outside of such a framework, permitted in traditional curriculum evaluation, is inadequate. A related point is that the comparison curriculum must be selected deliberately, to focus on specific research issues. Further, connecting the curriculum objects and activities and the processes of curricular enactment, including all components of the implementation, to the outcomes is important for theoretical, development, and practical reasons. Also connected to outcomes are variables from the broader data collected (e.g., data procedure via classroom observation instruments, such as various components of high-quality teaching of mathematics). Similar connections should be made across experimental and comparison classrooms (e.g., using the aforementioned measures of the quantity and quality of mathematics instruction). Without such connections, there is an inadequate basis for contributing to theories of learning and teaching in complex settings, guiding future curriculum development, and implementing the curriculum in various contexts. Finally, statistical analyses should allow making those connections (NRC Committee, 2004) and provide estimates of the efficacy of curricula expressed as effect sizes.

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" (Elmore, 1996, p. 1; see also Berends, Kirby, Naftel, & McKelvey, 2001; Confrey, Bell, & Carrejo, in press; Cuban, 2001; Tyack & Cuban, 1995; Tyack & Tobin, 1992). Thus, with any curriculum, but especially with one that differs from tradition, evaluations must be conducted on a large

scale (after considering issues of ethics and practical consequences, see Lester & Wiliam, 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 to document the fidelity of the implementation and the effects of the curriculum across diverse contexts. 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 curriculum'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; student/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 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 curriculum. 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 nested within classrooms) relate outcome measures to 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 curriculum must be sustained and evaluated in multiple sites for more than 2 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 curriculum (Darling-Hammond & Snyder, 1992), with attention given to issues of adoption and diffusion of the curriculum (Fishman et al., 2004; Rogers, 2003; Zaritsky et al., 2003). The large expense and the great effort involved in meeting these criteria are other reasons why previous evaluation phases should be employed first; only effective curricula should be scaled up.

A final approach is nonscientific (as is typical market research) and often contrived, but it may be frequent in practice and thus is mentioned for completeness. It is *not* a component of the CRF. Following the creation of a curriculum, research results that are consistent with it are cited post hoc. I am not aware of any recorded information about such *Post Hoc Rationalization*, but have on multiple occasions been asked by publishers to write one or several pages of research-based justifications for completed curriculum materials, and more than one colleague has confided that this practice is common. Ideally, such justifications would constitute descriptions of a priori foundations or other phases that were veraciously used as the basis for the curriculum but never recorded. In this case, the justifications would merely be documentation that was, unfortunately, delayed. As argued previously, all phases should be recorded in detail and shared with the greater community as part of the research process. In contrast, the chronology and the structure within which the requests for *Post Hoc Rationalizations* are frequently made suggest that this "documentation" may often be spurious.

Given this variety of possibilities, claims that a curriculum 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 10 described and the results obtained with each.

CURRICULUM RESEARCH AND MATHEMATICS CURRICULA

Some of the phases of the CRF have been used and reported in extant mathematics curriculum projects. A brief description of examples suggests that those that use multiple phases make substantive, unique contributions to theory, research, and curriculum development. Evaluations suggest that curricula whose development employed more of the phases of the CRF, including rigorous, mostly qualitative research in early development, have had more positive effects on learning.

Mathematics education in the United States has a long history of connecting research with curriculum development to varying degrees (Schaff, 1960; Whipple, 1930). Authors of "Patterns of Arithmetic" (Braswell & Romberg, 1969) reviewed basic research on learning, gathered feedback from participating teachers, and conducted extensive large-scale summative research that included inventories of teachers and students, as well as achievement tests. Many mathematics curriculum projects of the 1950s and 1960s were based to varying degrees on a priori research, and many were successful, although they generated only small amounts of summative research (Davis, 1984).

Unfortunately, many widely used mathematics textbooks of recent decades have not built on that foundation. Commercially published, traditional textbooks dominate mathematics curriculum materials in U.S. classrooms and to a great extent determine teaching practices (Goodlad, 1984; Grouws & Cebulla, 2000; Kilpatrick et al., 2001; Schmidt et al., 2001; Woodward & Elliot, 1990), even in the context of reform efforts (Grant, Peterson, & Shojgreen-Downer, 1996). According to Ginsburg, Klein, and Starkey (1998), the most influential publishers are a few large conglomerates that usually have profit, rather than the mathematics learning of children, as their main goal. This leads them to painstakingly follow state curriculum frameworks, attempting to meet every objective of every state—especially those that mandate adherence to their framework. Thus, *unscientific market research* is chiefly used to determine content and approach. Focus groups of teachers frequently emphasize that reform movements are not based "in the real world," that drill and practice should predominate curricula, and that "good textbooks" are those that get one through mathematics as quickly and effortlessly as possible by supplying simple activities and familiar routines (Ginsburg et al., 1998). The result is a false sense of innovation and research foundation. This reveals

> the skill of publishers in including materials which appear to support the new aspects . . . presented in such a way as not to embarrass those who wish to continue teaching mathematics the way they have always done it. (Burkhardt et al., 1990, p. 16)

Authors and editors of these textbooks are often researchers and other knowledgeable professionals, however, and they influence the materials to various degrees. Further, publishers state that the recent governmental policies on research have motivated renewed emphasis on research, clearly shown in their materials (in conversations with publishers about their materials, July 2005, one had "no research," although it was "planned"; another included some research from each of the CRF's three categories; two described a priori research only, with one saying that they planned additional methods; one described a priori and a mix of nonrandomized summative research). However, the difficulty of uncovering the nature and extent of that influence supports the concern that many curricula are not developed using scientific methods (which by definition must include full reporting) and do not contribute to the research literature.

Even materials based on theory and research, when limited to a priori phases, may not be successful. For example, the van Hiele theory of levels of geometric thinking and phases of instruction (van Hiele, 1986) lends itself to the *subject matter a priori foundation*, and the *pedagogical a priori foundation* phases. In two studies, a curriculum based on the theory did not lead to better achievement than a traditional approach (Halat & Aspinwall, 2004; Han, 1986). This is another indication that the research-to-practice model alone is inadequate.

Several recent projects have employed more of the phases in the CRF, with positive results. One is Realistic Mathematics Education (RME), whose "developmental research" is an integration of design and research (Gravemeijer, 1994b). Their procedures are consistent with the proposed CRF's *A Priori Foundations* and *Learning Model* (focusing on learning trajectories) categories, as well some formative and summative evaluation methods (Gravemeijer, 1994a, 1994b, 1999). Collaborators with the RME developers (McClain, Cobb, Gravemeijer, & Estes, 1999), have similar philosophical and curriculum development perspectives (Cobb & McClain, 2002; Gravemeijer, Cobb, Bowers, & Whitenack, 2000). These developers have documented positive results both on wide-scale adoption of the Netherlands curriculum and on student outcomes.

Some of the units in the *Investigations in Number, Data, and Space* were based on several research phases, with findings reported in the literature (Battista & Clements, 1996, 1998; Battista, Clements, Arnoff, Battista, & Borrow, 1998; Clements et al., 1996a, 1997a; Clements, Sarama, & Battista, 1996b, 1998; Clements, Sarama, Battista, & Swaminathan, 1996c). Other units were built upon *a priori foundation* knowledge and informal research in classrooms. Without such approaches, we would not know about the substantial role of spatial structuring in learning about two- and three-dimensional space, including mapping and measuring those spaces (Battista et al., 1998; Sarama, Clements, Swaminathan, McMillen, & González Gómez, 2003), of the integration of body motions and abstract-symbolic notions in the learning of turn and angle measurement (Clements et al., 1996a), or the impact of curriculum activities on other cognitive abilities (e.g., doubling of scores on spatial visualization resulting from activities on motions and areas; Clements et al., 1997a), much less the specific gains on targeted mathematics achievement that these reports document.

Considered together, these recent projects show signs of using at least some phases of all three categories of the CRF. They illustrate that these disciplined, mostly qualitative, methods have provided a rigorous research basis for materials, which are documented to result in improved student performance. They confirm the importance of knowledge about the students for whom the curriculum was designed (Tamir, 1988). Important to the theme of the present article, for several of these projects instructional design served as a primary setting for the development of theory[5] (Battista & Clements, 1996; Clements et al., 1997b; Cobb, 2001; Gravemeijer, 1994b; Sarama & Clements, 2002; Yerushalmy, 1997).

Most of these curricula have also been used widely, but specific reporting of results of multiple class formative or summative research have only begun to appear (e.g., Mokros, 2003; Streefland, 1991; and Cobb's group is planning on working with 10 classrooms). There are, of course, many other evaluations, such as the *summative research: small scale* evaluations by Fuson and colleagues of their own curricula and of *Everyday Math* (Fraivillig, Murphy, & Fuson, 1999; Fuson et al., 2000; Fuson, Smith, & Lo Cicero, 1997) or the studies of *Connected Mathematics 2*.[6]

As we shall discuss in the final section, it may be impractical for every project to include each phase. However, it is possible. One curriculum was based explicitly on the CRF, with all 10 phases applied at least to some degree (albeit taking

[5] Although this is a main point of the article, it deserves special attention. One reviewer of a previous draft of this article said that creation of curricula, empirical research, and theory were different activities and that the manuscript should address only one.

[6] Space constraints prohibit describing the many relevant research-based projects from the fields of mathematics education (e.g., Clements, 2002b; Confrey, Castro-Filho, & Wilhelm, 2000; Confrey & Lachance, 2000; Hoyles & Noss, 1992; Hoyles, Noss, & Sutherland, 1989; Lehrer & Chazan, 1998; Lewis & Tsuchida, 1998; Stigler & Hiebert, 1999; Yerushalmy, 1997) and cognitive science (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Brown, 1992; Griffin & Case, 1997; Lehrer et al., 1998a; Lehrer, Jenkins, & Osana, 1998b), as well as different conceptions such as didactical engineering (Artigue, 1994).

twice the originally funded 4-year period; Clements & Sarama, 2004a; Sarama, 2004; Sarama & Clements, 2002). The first *summative research: small scale* evaluation of the Building Blocks curriculum resulted in effect sizes of 1.71 for number and 2.12 for geometry (Cohen's d; Clements & Sarama, in press). Effect sizes of the first of two *summative research: large scale* evaluations 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 approached the sought-after 2-sigma effect of individual tutoring (Bloom, 1984, albeit under good conditions, with considerable support for the teachers). Further, the research described the support conditions necessary to achieve such effects. As another example, commercial publishers are beginning to support more phases of the CRF, even if the methods are not always fully conducted or fully reported in the CRF's scientific fashion (e.g., two at www.phschool.com/Research/math). Thus, the CRF is practicable. Consider, with the hundreds of millions of dollars undoubtedly spent on developing and testing mathematics curricula *without producing satisfactory evaluation data* (NRC Committee, 2004), is it more impractical to use the proposed CRF or to spend such large sums *without* using it?

RAMIFICATIONS

There are several ramifications of the proposed framework and this line of argument.

1. *Using the multiple phases of the proposed Curriculum Research Framework (CRF) will help developers improve curricula and contribute to the field of curriculum research.* 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. This is why different methods are used in various phases of the CRF (cf. NRC, 2002, p. 4; NRC Committee, 2004). For example, although iterating through one or two of the phases here, such as phase 8, *might* lead to an effective curriculum, such iteration would not meet all the goals outlined in Table 1. 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.

Using the CRF not only documents whether the design is successful in attaining achievement goals, but also traces whether that success can be attributed to the posited theory-design connections. This necessitates developers accepting new responsibilities, such as expanding their knowledge of the subject matter, psychology, and cognitive science, instruction, implementation, and scaling up, as well as of the variety of scientific research methods in the CRF's phases. Even if multiple phases are used, if they are all *a priori foundations*, for example, they are inadequate. As noted, subtle differences in activities can enhance or sabotage effectiveness (Sarama, 2000; Martin A. Simon, personal communication, May 28, 2002). Achieving the goals of the CRF (see Table 1) requires refining and especially elaborating principles by ongoing research and development work that tracks the effectiveness of every specific implementation, consistently maintaining links to the hypothesized theories and models, through progressively expanding social contexts. Ensuring that the *research trajectory* described by the CRF is coherent and connected throughout the development process maintains unbroken threads of argumentation.

2. *Achieving the goals of CRF requires both qualitative and quantitative methodologies* (NRC Committee, 2004, makes similar recommendations, albeit for summative research only). In response to theorists who celebrate the "defeat of quantitative research in the curriculum field and the victory of qualitative research" (Pinar et al., 1995, p. 52), I paraphrase Mark Twain to say that the report of its death is greatly exaggerated. Both approaches can make valid, rigorous contributions to scientific research (Darling-Hammond & Snyder, 1992; Johnson & Onwuegbuzie, 2004; NRC, 2002; NRC Committee, 2004). Quantitative methodologies provide experimental results, garnered under conditions distant from the developers, that are useful in and of themselves and in that they can generate political and public support. Randomized experiments are more powerful and less biased than alternative designs and also can uncover unexpected and subtle interactions not revealed by qualitative investigations (Clements & Nastasi, 1988; Nastasi, Clements, & Battista, 1990; Russek & Weinberg, 1993).

Qualitative methodologies are important for three reasons. First, curriculum research seeks to understand individual students' interpretations and learning and how these change in the context of, and as a result of, interactions among teachers and students around a specific curriculum. Qualitative research describes the nature of the "it" when researchers ask, "Did it work?" (Erickson & Gutierrez, 2002); validity is suspect without this information (especially given the possibility of unintended and immeasurable outcomes; Taba, 1962; van Oers, 2003; Walker, 1992). Second, such research helps explain *why* it works and how and why it works differently in different contexts. Third, qualitative research in a triangulation context may serve to validate or invalidate quantitative results, more so than the inverse (Russek & Weinberg, 1993), and such methodologies complement experiments in ruling out alternate explanations (NRC, 2002). Experiments control a necessarily small fraction of an indefinite number of contextual variables, and one will rarely identify limiting or catalytic conditions and curricular features (including the aforementioned "subtle differences") optimally by considering only focal experimental variables (Greenwald et al., 1986). In summary, given its inherently complex and creative nature, its interpretive goals, the small number of students involved in many of its techniques, and the progressive breadth of concerns combined with the consistent need for sensitivity to new findings and insights, curriculum research requires qualitative methodologies and openness to emergent findings throughout the phases (Smith, 1983).

Quantitative and qualitative method are integrated throughout the CRF's phases. Every experiment benefits from collecting ethnographic data. Conversely, the validity of qualitative methodologies, such as case studies, is increased if they are conducted within the context of an experiment (Cook, 2002). Finally, the use of summative evaluation without other phases is usually premature, wasteful, and misleading. (The medical research model, oft-cited as the gold standard, uses randomized trials, especially large-scale experiments, only after nonrandom, discovery strategies, exploratory clinical research, dose-response trials, etc.; Giorgianni & Granna, 1999; Zaritsky et al., 2003.) Thus, although randomized experiments remain the best design for evaluation of causal interpretations, placing them in the context of a complete research framework mitigates the limitations and misuses of randomized experiments (The Design-Based Research Collective, 2003).

3. *Increasing academe's support for curriculum research will improve curricula, research, and the public's opinion of educational research*. There is a long history of bias against design sciences in academe (Simon, 1969; Wittmann, 1995). Increasing support is justified for at least two reasons. First, such research is legitimate science and has led to new directions in theory and empirical research in complex situations. Second, universities benefit as well as schools, because the approaches will prove practically useful and thus will legitimize educational research per se to a wide audience.

4. *Curriculum research could be more successful if funding agencies reconsidered funding requirements and time frames for this enterprise*. Curriculum research needs increased funding (Feuer et al., 2002). The proportion of funds currently allocated to research in education is inconsistent with virtually any other enterprise (Dow, 1991; President's Committee of Advisors on Science and Technology—Panel on Educational Technology, 1997; Schoenfeld, 1999). All phases of the CRF entail substantial costs. Paradoxically, using the full range of phases increases the justification for expending public funds, because the resultant curricula will be more effective and better documented; a substantive amount of valid research will be produced evaluating that curriculum and guiding future curriculum development, research, and theoretical efforts; and contextual and other implementation issues will be addressed. To realize these benefits, funding agencies could insist that those receiving funds propose and apply a coherent use of the CRF's phases, including the essential last step of sharing the research—addressing perhaps the worst sin of the curriculum development community.

Such funding suggests a concomitant reconsideration of the time such development requires. Usually in the development of curricula, there are deadlines, but any extra time that might exist is usually used to improve the product, rather than for reflection and research (Gravemeijer, 1994b). Funded curriculum projects usually are given implausible time frames that make such reflection and research (especially using the multiple methods in the CRF) nearly impossible, such as 5 years to develop 5 years of curriculum (Schoenfeld, 1999).

5. *To benefit from curriculum research, the entire education community needs to support and expect research-based curriculum development and to expect specific methods used and results obtained to be fully explicated*. Lack of a connec-

tion between research and curriculum development and adoption is a major reason that curriculum, and ultimately student achievement, in the United States do not improve (Battista & Clements, 2000; Clements, 2002; Clements & Battista, 2000) and that curriculum reforms usually fail, with "genuine achievements . . . thrown out along with excesses and failures" (Walker, 2003, p. 116). To have substantial benefit for all children, the educational community has to establish scientific research as a sine qua non of curriculum development and selection. Educators at all levels should insist that a full reporting of methods and findings accompany any curriculum proffered and should eschew curricula that do not have the support of at least a viable subset of the phases; the construct of "evidence-based" or "research-based" curricula is spurious without such criteria. This calls into question much of what is currently used in classrooms, which might be replaced as successful research-based curricula become available.[7]

CAVEATS AND CONCLUSIONS

Although I believe the proposed Curriculum Research Framework (CRF) has been and can be useful, it is inchoate and in need of further testing and elaboration. For example, the nature, basis, and procedures in the use of learning trajectories need to be clarified (Clements & Sarama, 2004b, discusses variations such as psychological vs. social perspectives). Prosaic issues such as the optimal amount of time or number of iterations of specific phases are underdetermined. Maintaining theoretical continuity between phases must be further addressed. Finally, phases that rely on design experiments are vulnerable to the weaknesses in those methods. Design experiments cannot control the many variables in their complex settings; the large amount of data collected can rarely be fully analyzed before the next cycle of revision, enactment, and analysis take place (Collins et al., 2004); and different participants may have different data and perspectives, so that ultimate paths and products may be to an extent arbitrary and generalization difficult (Kelly, 2004). Randomized trials have weaknesses that ameliorate many of these limitations. However, design experiments and other methods such as teaching experiments and classroom-based teaching experiments, which include conceptual and relational, or semantic, analysis, are theoretically grounded methodologies that can help accomplish what randomized trials cannot: Build models of the child's mathematics, of mental actions-on-objects, of learning, and of teaching interactions (Les Steffe,

[7] Being based on research does not, of course, guarantee success—evaluation being one reason to conduct research—nor does it speak to values and goals (cf. Hiebert, 1999; NRC, 2002), although, qua research, it *should be explicit about those values and goals*. Thus, the results of research remain only one criterion for curriculum selection. However, findings from multiple curriculum research methods that indicate that valued goals will be achieved should constitute the most important standard. In addition, fortunately, the research methods discussed here that include tight cycles of planning, instruction, and analysis, are consistent with the practices of teachers who develop broad conceptual and procedural knowledge in their students (Cobb, 2001; Fuson et al., 2000; Lampert, 1988; Simon, 1995; Stigler & Hiebert, 1999); therefore, the curriculum and findings are not only applicable to other classrooms but also support those practices.

personal communication, July 18, 2005). In summary, because the CRF includes a *coherent complement* of methods, built-in checks and balances address limitations of each method, with the focus on the *Learning Model* especially useful for maintaining a core focus.

In conclusion, a synthesis of curriculum development, classroom teaching, and research is necessary to contribute both to a better understanding of thinking, learning, and teaching and to progressive change in curricula. Without curriculum development projects, researchers would have fewer rich tasks, authentic settings, and theoretical problems. Such projects serve as sources of, and testing sites of, research ideas. Without concurrent research, curriculum developers and teachers would miss opportunities to learn about critical aspects of students' thinking and the particular features of software, curricula, and teaching actions that engender learning (including understandings of limitations on what a curriculum alone can "promise," given that curriculum enactment affects effectiveness). I believe that the CRF can help ameliorate these problems (Clements et al., 1997a; Schoenfeld, 1999). 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 curriculum, then research moves to the vanguard of educational innovation and results in substantive student achievement across the multiple goals of educational reform (NRC, 2002; Taba, 1962).

I argue that *curriculum research is one of the best ways to answer the three types of research questions* (NRC, 2002)*, descriptive, causal, and process, within a program that is synergistic, integrated, and complete*. Across the different phases, and within them, there are iterative cycles, each of which must "work" to proceed and reveal weaknesses if they do not work, and thus offer tests of construct validity that are both more frequent and more trustworthy than tests in most other approaches (cf. Johnson & Onwuegbuzie, 2004). Further, because it is result-centered, rather than theory-centered, the CRF minimizes seductive theory-confirming strategies that tend to insidiously replace the intended theory-testing strategies and maximizes strategies that attempt to produce specified patterns of data and thus mitigate confirmation bias, stimulating creative development of theory (Greenwald et al., 1986). This type of scientific research both constrains decisions to be consistent with what has been scientifically verified (James, 1958) and liberates, by broadening the range of possibilities (Dewey, 1929). The CRF makes the relationships among theory, research, design, and practice more salient and accessible to reflection.

I also argue that *curriculum should be produced and selected using all of CRF's phases that are appropriate* and that the more comprehensive the curriculum (e.g., compare a single module undergoing minor revisions to a complete pre-K to grade 8 mathematics curriculum built from the ground up), the more phases should be employed. Thus, all 10 CRF phases need not, and often cannot, be employed in every individual project (e.g., a single project may simply be evaluating a published curriculum). However, every curriculum should be based on a foundation of extant research and should proceed in the context of a coherent, dynamic research program that uses all the phases that are applicable and tractable. Decisions to omit certain phases should be made deliberately, and reasons for those decisions documented. Optimizing the contribution of both the curriculum and research produced, and avoiding pitfalls of randomized trials such as the premature experimental evaluation of an innovation, depends on using all relevant phases.

Although I believe these implications and guidelines are warranted, the main purpose of this article is to begin a discussion of a framework for the construct of "research-based curricula." Therefore, criticisms and alterations would be welcome, as well as agreements and applications.

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167–207.

Artigue, M. (1994). Didactical engineering as a framework for the conception of teaching products. In R. Biehler, R. W. Scholz, R. Strässer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 27–39). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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*, *16*(2), 6–8; 14.

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher*, *32*(1), 21–24.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*, 1–14.

Baroody, A. J. (1987). *Children's mathematical thinking*. New York: Teachers College Press.

Baroody, A. J., with Coslick, R. T. (1998). *Fostering children's mathematical power: An investigative approach to K–8 mathematics instruction*. Mahwah, NJ: Lawrence Erlbaum Associates.

Battista, M. T., & Clements, D. H. (1996). Students' understanding of three-dimensional rectangular arrays of cubes. *Journal for Research in Mathematics Education*, *27*, 258–292.

Battista, M. T., & Clements, D. H. (2000). Mathematics curriculum development as a scientific endeavor. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 737–760). Mahwah, NJ: Lawrence Erlbaum Associates.

Battista, M. T., Clements, D. H., Arnoff, J., Battista, K., & Borrow, C. V. A. (1998). Students' spatial structuring of 2d arrays of squares. *Journal for Research in Mathematics Education*, *29*, 503–532.

Beauchamp, G. A. (1981). *Curriculum theory*. Itasca, IL: F. E. Peacock Publishers.

Beauchamp, G. A. (1986). Curriculum theory: Meaning, development, and use. *Theory Into Practice*, *21*(1), 23–27.

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in new American schools: Three years into scale-up*. Santa Monica, CA: Rand Education.

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*, 4–16.

Bodilly, S. J. (1998). *Lessons from new American schools' scale-up phase*. Santa Monica, CA: RAND Education.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*, 125–230.

Braswell, J. S., & Romberg, T. A. (1969). *Objectives of patterns in arithmetic and evaluation of the telecourse for grades 1 and 3*. Madison, WI: Wisconsin Research and Development Center for Cognitive Learning.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in evaluating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In R. Glaser (Ed.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates.

Burkhardt, H., Fraser, R., & Ridgway, J. (1990). The dynamics of curriculum change. In I. Wirszup & R. Streit (Eds.), *Developments in school mathematics around the world* (Vol. 2, pp. 3–30). Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, *15*, 179–202.

Case, R. (1982). General developmental influences on the acquisition of elementary concepts and algorithms in arithmetic. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 156–170). Hillsdale, NJ: Lawrence Erlbaum Associates.

Char, C. A. (1990). *Click on the clock: Formative research in the development of mathematics software for young children* (No. Report No. 90–1). Newton, MA: Educational Development Center.

Childers, R. D. (Cartographer). (1989). *Implementation of the writing to read instructional system in 13 rural elementary schools in southern West Virginia. 1988–89 annual report*.

Clements, D. H. (2002). Linking research and curriculum development. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 599–636). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D. H., & Battista, M. T. (2000). Designing effective software. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 761–776). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1996a). Development of turn and turn measurement concepts in a computer-based instructional unit. *Educational Studies in Mathematics*, *30*, 313–337.

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

Clements, D. H., Battista, M. T., Sarama, J., Swaminathan, S., & McMillen, S. (1997b). Students' development of length measurement concepts in a Logo-based unit on geometric paths. *Journal for Research in Mathematics Education*, *28*(1), 70–95.

Clements, D. H., Meredith, J. S., & Battista, M. T. (1992). Design of a Logo environment for elementary geometry. In W. Geeslin & K. Graham (Eds.), *Proceedings of the sixteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (vol. 1, p. 152). Durham, NH: Program Committee.

Clements, D. H., & Nastasi, B. K. (1988). Social and cognitive interactions in educational computer environments. *American Educational Research Journal*, *25*, 87–106.

Clements, D. H., & Sarama, J. (1995). Design of a Logo environment for elementary geometry. *Journal of Mathematical Behavior*, *14*, 381–398.

Clements, D. H., & Sarama, J. (1998). *Building blocks—foundations for mathematical thinking, pre-kindergarten to grade 2: Research-based materials development* [National Science Foundation, grant number ESI-9730804; see www.Gse.Buffalo.Edu/org/buildingblocks/]. Buffalo, NY: State University of New York at Buffalo.

Clements, D. H., & Sarama, J. (2002). Design of microworlds in mathematics and science education. *Journal of Educational Computing Research*, *27*(1&2), 1–6.

Clements, D. H., & Sarama, J. (2004a). *Building blocks* for early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 181–189.

Clements, D. H., & Sarama, J. (2004b). Hypothetical learning trajectories [special issue]. *Mathematical Thinking and Learning*, *6* (2).

Clements, D. H., & Sarama, J. (2004c). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*, 81–89.

Clements, D. H., & Sarama, J. (in press). Effects of a preschool mathematics curriculum: Summary research on the *building blocks* project. *Journal for Research in Mathematics Education*.

Clements, D. H., Sarama, J., & Battista, M. T. (1996b). Development of turn and turn measurement concepts in a computer-based instructional unit. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the eighteenth annual meeting of the North American Chapter of the International*

Group for the Psychology of Mathematics Education (Vol. 2, pp. 547–552). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Clements, D. H., Sarama, J., & Battista, M. T. (1998). Development of concepts of geometric figures in a specially-designed Logo computer environment. *Focus on Learning Problems in Mathematics*, *20*, 47–64.

Clements, D. H., Sarama, J., Battista, M. T., & Swaminathan, S. (1996c). Development of students' spatial thinking in a curriculum unit on geometric motions and area. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the eighteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 217–222). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Clements, D. H., Sarama, J., & DiBiase, A.-M. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Cobb, P. (2001). Supporting the improvement of learning and teaching in social and institutional context. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 455–478). Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Cobb, P., & McClain, K. (2002). Supporting students' learning of significant mathematical ideas. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 154–166). Oxford, England: Blackwell.

Cohen, D. K. (1996). Rewarding teachers for student performance. In S. H. Fuhrman & J. A. O'Day (Eds.), *Rewards and reforms: Creating educational incentives that work*. San Francisco: Jossey Bass.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, *13*(1), 15–42.

Confrey, J. (1996). The role of new technologies in designing mathematics education. In C. Fisher, D. C. Dwyer, & K. Yocam (Eds.), *Education and technology, reflections on computing in the classroom* (pp. 129–149). San Francisco: Apple Press.

Confrey, J. (2000). Improving research and systemic reform toward equity and quality. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 87–106). Mahwah, NJ: Lawrence Erlbaum Associates.

Confrey, J., Bell, K., & Carrejo, D. (in press). *Systemic crossfire: What implementation research reveals about urban reform in mathematics*. Austin, TX: Systemic Research Collaborative for Education in Mathematics, Science, and Technology (SYRCE), University of Texas at Austin.

Confrey, J., Castro-Filho, J., & Wilhelm, J. (2000). Implementation research as a means to link systemic reform and applied psychology in mathematics education. *Educational Psychologist*, *35*, 179–191.

Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis*, *24*, 175–199.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., et al. (Eds.). (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.

Cronbach, L. J., & Suppes, P. (Eds.). (1969). *Research for tomorrow's schools: Disciplined inquiry for education*. New York: Macmillan.

Cuban, L. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.

Darling-Hammond, L., & Snyder, J. (1992). Curriculum studies and the traditions of inquiry: The scientific tradition. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 41–78). New York: Macmillan.

Davis, B., & Simmt, E. (2003). Understanding learning systems: Mathematics education and complexity science. *Journal for Research in Mathematics Education*, *34*, 137–167.

Davis, R. B. (1984). *Learning mathematics: The cognitive science approach to mathematics education*. Norwood, NJ: Ablex.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8.

Dewey, J. (1902/1976). The child and the curriculum. In J. A. Boydston (Ed.), *John Dewey: The middle works, 1899–1924. Volume 2: 1902–1903* (pp. 273–291). Carbondale, IL: Southern Illinois University Press.

Dewey, J. (1929). *The sources of a science of education*. New York: Liveright Publishing Corp.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences*, *13*(1), 77–103.

Dow, P. B. (1991). *School house politics*. Cambridge, MA: Harvard University Press.

Eisner, E. W. (1998). The primacy of experience and the politics of method. *Educational Researcher*, *17*(5), 15–20.

Elmore, R. F. (1996). Getting to scale with good educational practices. *Harvard Educational Review*, *66*, 1–25.

Erickson, F., & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher*, *31*(8), 21–24.

Escobedo, T. H., & Evans, S. (1997, March). A comparison of child-tested early childhood education software with professional ratings, Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, *31*(8), 4–14.

Fishman, B., Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences*, *13*, 43–76.

Flagg, B. (1990). *Formative evaluation for educational technology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Forman, G. E. (1993). The constructivist perspective to early education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (2nd ed., pp. 137–155). New York: Merrill.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in *everyday mathematics* classrooms. *Journal for Research in Mathematics Education*, *30*, 148–170.

Fullan, M. G. (1992). *Successful school improvement*. Philadelphia: Open University Press.

Fuson, K. C. (1997). Research-based mathematics curricula: New educational goals require programs of four interacting levels of research. *Issues in Education*, *3*(1), 67–79.

Fuson, K. C., Carroll, W. M., & Drueck, J. V. (2000). Achievement results for second and third graders using the *standards*-based curriculum *everyday mathematics*. *Journal for Research in Mathematics Education*, *31*, 277–295.

Fuson, K. C., Smith, S. T., & Lo Cicero, A. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education*, *28*, 738–760.

Gerber, M. M., Semmel, D. S., & Semmel, M. I. (1994). Computer-based dynamic assessment of multidigit multiplication. *Exceptional Children*, *61*, 114–125.

Ginsburg, H. P. (1977). *Children's arithmetic: How they learn it and how you teach it*. Austin, TX: Pro-ed.

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

Ginsburg, H. P., Klein, A., & Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of child psychology. Volume 4: Child psychology in practice* (pp. 401–476). New York: John Wiley & Sons.

Giorgianni, S. J., & Granna, J. (1999). The importance of innovation in pharmaceutical research. *The Pfizer Journal*, *7*(2).

Goodlad, J. I. (1984). *A place called school*. New York: McGraw-Hill.

Goodlad, J. I., & Associates. (1979). *Curriculum inquiry: The study of curriculum practice*. New York: McGraw-Hill.

Grant, S. G., Peterson, P. L., & Shojgreen-Downer, A. (1996). Learning to teach mathematics in the context of system reform. *American Educational Research Journal*, *33*, 509–541.

Gravemeijer, K. P. E. (1994a). *Developing realistic mathematics instruction*. Utrecht, The Netherlands: Freudenthal Institute.

Gravemeijer, K. P. E. (1994b). Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education*, *25*, 443–471.

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

Gravemeijer, K. P. E., Cobb, P., Bowers, J., & Whitenack, J. (2000). Symbolizing, modeling, and instructional design. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 225–274). Mahwah, NJ: Lawrence Erlbaum Associates.

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, *93*, 216–229.

Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, *3*(1), 1–49.

Grouws, D. A., & Cebulla, K. J. (2000). Elementary and middle school mathematics at the crossroads. In T. L. Good (Ed.), *American education: Yesterday, today, and tomorrow (volume ii)* (pp. 209–255). Chicago: University of Chicago Press.

Halat, E., & Aspinwall, L. (2004, April). Van Hiele theory based curriculum in geometry; performance and gender. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Han, T.-S. (1986). The effects on achievement and attitude of a standard geometry textbook and a textbook consistent with the van Hiele theory. *Dissertation Abstracts International*, 47, 3690A.

Herbst, P. (2003). Using novel tasks in teaching mathematics: Three tensions affecting the work of the teacher. *American Educational Research Journal*, *40*, 197–238.

Hiebert, J. C. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, *30*, 3–19.

Hiebert, J. C., Carpenter, T., Fennema, E. H., Fuson, K. C., Wearne, D., Murray, H. G., et al. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hoyles, C., & Noss, R. (1992). A pedagogy for mathematical microworlds. *Educational Studies in Mathematics*, *23*, 31–57.

Hoyles, C., Noss, R., & Sutherland, R. (1989). Designing a Logo-based microworld for ratio and proportion. *Journal of Computer Assisted Learning*, *5*, 208–222.

Jackson, P. W. (1992). Conceptions of curriculum and curriculum specialists. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 3–40). New York: Macmillan.

Jaeger, R. M. (1988). Survey research methods in education. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (pp. 303–340). Washington: American Educational Research Association.

James, W. (1958). *Talks to teachers on psychology: And to students on some of life's ideas*. New York: Norton.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Kamii, C. K. (1985). *Young children reinvent arithmetic: Implications of Piaget's theory*. New York: Teachers College Press.

Kamii, C. K., & Dominick, A. (1998). The harmful effects of algorithms in grades 1–4. In L. J. Morrow & M. J. Kenney (Eds.), *The teaching and learning of algorithms in school mathematics* (pp. 130–140). Reston, VA: National Council of Teachers of Mathematics.

Kamii, C. K., & Housman, L. B. (1999). *Young children reinvent arithmetic: Implications of Piaget's theory* (2nd ed.). New York: Teaching College Press.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences*, *13*(1), 115–128.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Lagemann, E. (1997). Contested terrain: A history of education research in the united states. *Educational Researcher*, *26*(9), 5–18.

Lahm, E. A. (1996). Software that engaged young children with disabilities: A study of design features. *Focus on Autism and Other Developmental Disabilities*, *11*(2), 115–124.

Lampert, M. (1988). *Teaching that connects students' inquiry with curricular agendas in schools. Technical report*. Cambridge, MA: Educational Technology Center, Harvard Graduate School of Education.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

Lavin, R. J., & Sanders, J. E. (1983). *Longitudinal evaluation of the c/a/i computer assisted instruction title 1 project: 1979–82*. Chelmsford, MA: Merrimack Education Center.

Lawton, J. T. (1993). The Ausubelian preschool classroom. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (2nd ed., pp. 157–177). New York: Merrill.

Lehrer, R., & Chazan, D. (Eds.). (1998). *Designing learning environments for developing understanding of geometry and space*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., Jacobson, C., Thoyre, G., Kemeny, V., Strom, D., Horvarth, J., et al. (1998a). Developing understanding of geometry and space in the primary grades. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 169–200). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., Jenkins, M., & Osana, H. (1998b). 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 Associates.

Lesh, R. A. (1990). Computer-based assessment of higher order understandings and processes in elementary mathematics. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 81–110). Washington, DC: American Association for the Advancement of Science.

Lesh, R. A., & Kelly, A. E. (2000). Multitiered teaching experiments. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 197–230). Mahwah, NJ: Lawrence Erlbaum Associates.

Lester, F. K., Jr., & Wiliam, D. (2002). On the purpose of mathematics education research: Making productive contributions to policy and practice. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 489–506). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewis, C. C., & Tsuchida, I. (1998). A lesson is like a swiftly flowing river: How research lessons improve Japanese education. *American Educator*, 12, 14–17; 50–52.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lincoln, Y. S. (1992). Curriculum studies and the traditions of inquiry: The humanistic tradition. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 79–97). New York: Macmillan.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

McClain, K., Cobb, P., Gravemeijer, K. P. E., & Estes, B. (1999). Developing mathematical reasoning within the context of measurement. In L. V. Stiff & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K–12* (pp. 93–106). Reston, VA: National Council of Teachers of Mathematics.

Mohrman, S. A., & Lawler III, E. E. (1996). Motivation for school reform. In S. H. Fuhrman & J. A. O'Day (Eds.), *Rewards and reform: Creating educational incentives that work* (pp. 115–143). San Francisco: Jossey-Bass.

Mokros, J. R. (2003). Learning to reason numerically: The impact of *investigations*. In S. Senk (Ed.), *Standards-based school mathematics curricula. What are they? What do students learn?* (pp. 109–132). Mahwah, NJ: Lawrence Erlbaum Associates.

Murphy, R. T., & Appel, L. R. (1984). *Evaluation of writing to read*. Princeton, NJ: Educational Testing Service.

Nastasi, B. K., Clements, D. H., & Battista, M. T. (1990). Social-cognitive interactions, motivation, and cognitive growth in Logo programming and CAI problem-solving environments. *Journal of Educational Psychology*, *82*, 150–158.

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*. Reston, VA: Author.

National Research Council. (2002). *Scientific research in education*. In R. J. Shavelson & L. Towne (Eds.). Washington, DC: National Research Council, National Academy Press.

NRC Committee. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations*. Washington, D.C.: Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education, The National Academies Press.

Papert, S. (1987). Computer criticism vs. technocentric thinking. *Educational Researcher*, *16*(1), 22–30.

Pinar, W. F., Reynolds, W. M., Slattery, P., & Taubman, P. M. (1995). *Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses*. New York: Peter Lang.

President's Committee of Advisors on Science and Technology—Panel on Educational Technology. (1997). *Report to the president on the use of technology to strengthen K–12 education in the United States*. Washington, DC: Author.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, *53*, 109–120.

Reeves, D. B. (2002, May 8). Galileo's dilemma. *Education Week*, p. 44.

Remillard, J. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *Elementary School Journal*, *100*, 331–350.

Rogers, E. M. (2003). *Diffusion of innovations* (4th ed.). New York: The Free Press.

Russek, B. E., & Weinberg, S. L. (1993). Mixed methods in a study of implementation of technology-based materials in the elementary classroom. *Evaluation and Program Planning*, *16*, 131–142.

Sarama, J. (2000). Toward more powerful computer environments: Developing mathematics software on research-based principles. *Focus on Learning Problems in Mathematics*, *22*(3&4), 125–147.

Sarama, J. (2004). Technology in early childhood mathematics: *Building blocks™* as an innovative technology-based curriculum. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 361–375). Mahwah, NJ: Lawrence Erlbaum Associates.

Sarama, J., & Clements, D. H. (2002). *Building blocks* for young children's mathematical development. *Journal of Educational Computing Research*, *27*(1&2), 93–110.

Sarama, J., & Clements, D. H. (in press). Linking research and software development. In K. Heid & G. Blume (Eds.), *Technology in the learning and teaching of mathematics: Syntheses and perspectives*. New York: Information Age Publishing, Inc.

Sarama, J., Clements, D. H., & Henry, J. J. (1998). Network of influences in an implementation of a mathematics curriculum innovation. *International Journal of Computers for Mathematical Learning*, *3*, 113–148.

Sarama, J., Clements, D. H., Swaminathan, S., McMillen, S., & González Gómez, R. M. (2003). Development of mathematical concepts of two-dimensional space in grid environments: An exploratory study. *Cognition and Instruction*, *21*, 285–324.

Sarama, J., Clements, D. H., & Vukelic, E. B. (1996). The role of a computer manipulative in fostering specific psychological/mathematical processes. In E. Jakubowski, D. Watkins & H. Biske (Eds.), *Proceedings of the eighteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 567–572). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*, 265–283.

Schaff, W. L. (1960). Selected annotated bibliography. In *Instruction in arithmetic. Twenty-fifth yearbook* (pp. 320–354). Washington, D.C.: National Council of Teachers of Mathematics.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Schoenfeld, A. H. (1999). Looking toward the 21st century: Challenge of educational theory and practice. *Educational Researcher*, *28*, 4–14.

Schoenfeld, A. H. (2002). Research methods in (mathematics) education. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 435–487). Mahwah, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A. H., Smith III, J. P., & Arcavi, A. (1993). Learning: The microgenetic analysis of one students' evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in instructional psychology* (vol. 4, pp. 55–175). Mahwah, NJ: Lawrence Erlbaum Associates.

Schwandt, T. A. (2002). *Evaluation practice reconsidered*. New York: Peter Lang.

Shade, D. D. (1994). Computers and young children: Software types, social contexts, gender, age, and emotional responses. *Journal of Computing in Childhood Education*, *5*(2), 177–209.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*, 606–620.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: The M.I.T. Press.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145.

Smith, J. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*, *12*(3), 6–13.

Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402–435). New York: Macmillan.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart & Winston.

Spradley, J. P. (1980). *Participant observation*. New York: Holt, Rinehart & Winston.

Steffe, L. P. (1983). Children's algorithms as schemes. *Educational Studies in Mathematics*, *14*, 109–125.

Steffe, L. P. (1994). Children's multiplying schemes. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 3–39). Albany, NY: SUNY Press.

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

Steffe, L. P., Thompson, P. W., & Glasersfeld, E. v. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267–306). Mahwah, NJ: Lawrence Erlbaum Associates.

Steffe, L. P., & Wiegel, H. G. (1994). Cognitive play and mathematical learning in computer microworlds. *Journal of Research in Childhood Education*, *8*(2), 117–131.

Stigler, J. W., & Hiebert, J. C. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Taba, H. (1962). *Curriculum development: Theory and practice*. New York: Harcourt Brace Jovanovich, Inc.

Tamir, P. (1988). The role of pre-planning curriculum evaluation in science education. *Journal of Curriculum Studies*, *20*, 257–262.

Tushnet, N. C., Millsap, M. A., Abdullah-Welsh, N., Brigham, N., Cooley, E., Elliot, J., et al. (2000). *Final report on the evaluation of the national science foundation's instructional materials development program*. Arlington, VA: National Science Foundation.

Tyack, D., & Cuban, L. (1995). *Tinkering towards utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Tyack, D., & Tobin, W. (1992). The "Grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, *31*, 453–479.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

van den Brink, F. J. (1991). Realistic arithmetic education for young children. In L. Streefland (Ed.), *Realistic mathematics education in primary school* (pp. 77–92). Utrecht, The Netherlands: Freudenthal Institute, Utrecht University.

van Hiele, P. M. (1986). *Structure and insight: A theory of mathematics education*. Orlando, FL: Academic Press.

van Oers, B. (2003). Learning resources in the context of play. Promoting effective learning in early childhood. *European Early Childhood Education Research Journal*, *11*, 7–25.

Walker, D. F. (1992). Methodological issues in curriculum research. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 98–118). New York: Macmillan.

Walker, D. F. (2003). *Fundamentals of curriculum: Passion and professionalism* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Weiss, I. R. (2002, April). Systemic reform in mathematics education: What have we learned? Paper presented at the research presession of the 80th annual meeting of the National Council of Teachers of Mathematics, Las Vegas, NV.

Whipple, G. M. (1930). *The twenty-ninth yearbook of the National Society for the Study of Education: Report of the society's committee on arithmetic*. Bloomington, IL: Public School Publishing Co.

Wittmann, E. C. (1995). Mathematics education as a "design science." *Educational Studies in Mathematics*, *29*, 355–374.

Woodward, A., & Elliot, D. L. (1990). Textbook use and teacher professionalism. In K. J. Rehage, D. L. Elliot, & A. Woodward (Eds.), *Textbooks and schooling in the United States (eighty-ninth yearbook of the National Society for the Study of Education, part 1)* (pp. 178–193). Chicago: University of Chicago Press.

Yerushalmy, M. (1997). Emergence of new schemes for solving algebra word problems. In E. Pehkonen (Ed.), *Proceedings of the 21st conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 165–178). Lahti, Finland: University of Helsinki.

Zaritsky, R., Kelly, A. E., Flowers, W., Rogers, E., & O'Neil, P. (2003). Clinical design sciences: A view from sister design efforts. *Educational Researcher*, *32*(1), 32–34.

Author

Douglas Clements, Professor, University at Buffalo, State University of New York, Graduate School of Education, 212 Baldy Hall (North Campus), Buffalo, NY 14260; clements@buffalo.edu