Curriculum Development Research:

Toward a Framework for “Research-based Curricula”

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

Draft. Please provide feedback, but do not cite or share.

This paper 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." 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 National Science Foundation.

Forms of this paper were presented at the following conferences.


Appreciation is expressed to Frank Lester, Julie Sarama, Martin A. Simon, Alan Schoenfeld, and Leslie Steffe for their comments on early drafts.
Abstract

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 methods of the proffered curriculum development research framework.
Government agencies have recently emphasized the importance of evidence-based instructional materials. It would be reasonable to assume 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 “research-based,” discourages scientific approaches to curriculum development (and allows the continued dominance of non-scientific “market research”); undermines attempts to create a shared research foundation for the creation of, and informed choices of, classroom curricula; and reinforces the bias against design sciences, such as curriculum development, in academe (H. A. Simon, 1969; Wittmann, 1995). 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,” using mathematics as the subject matter domain for illustration, and to discuss the ramifications for multiple relevant parties, including practitioners, curriculum developers, researchers, administrators, funding agencies, and policy makers.

**Curriculum and Scientific Research**

“Curriculum” has different meanings in different contexts (Beauchamp, 1986; Jackson, 1992; Pinar, Reynolds, Slattery, & Taubman, 1995; Walker, 2003). While 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 PreK-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 throughout the paper without appending the adjective “available” each time. 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 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.

Isolation of curriculum development and educational research vitiates both (Clements & Battista, 2000; Clements, Battista, Sarama, & Swaminathan, 1997; 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, the minimal connection between them in too many cases is one reason curriculum development in the U.S. does not reliably improve (Battista & Clements, 2000; Clements, 2002). For example, although knowledge is usually created during curriculum development, it is usually not explicated or published (Gravemeijer, 1994b) and thus is unavailable to the educational community.

Like “curriculum,” “scientific research” also has different meanings. Scientific research includes the observation, description, analysis, experimental investigation, and
theoretical explanation of phenomena. Scientific knowledge is accepted as more reliable than commonsense knowledge because the way in which it is developed is explicit and repeatable. “Our faith [in scientific research] rests entirely on the certainty of reproducing or seeing again a certain phenomenon by means of certain well defined acts” (Valéry, 1957, p.1253, as quoted in Glasersfeld, 1987). These acts are the methods of acquiring scientific knowledge. Scientific method is disciplined inquiry (Cronbach & Suppes, 1969; Dewey, 1929; Feuer et al., 2002; Shulman, 1997). “Inquiry” suggests that the investigation’s goal is answering specific questions. “Disciplined” suggests that the investigation should be guided by concepts and methods from disciplines and connected to relevant theory in those disciplines, and also that it should be in the public view so that the inquiry can be inspected and criticized. The use of research methods, and the conscious documentation and full reporting of these processes—data collection, argumentation, reasoning, and checking for counterhypotheses—distinguishes disciplined inquiry from other sources of opinion and belief (Cronbach & Suppes, 1969; National Research Council, 2002; Shulman, 1997).

Thus, scientific knowledge is valued because it offers reliable, self-correcting, documented, shared knowledge based on research methodology (Mayer, 2000; National Research Council, 2002). Curriculum development is a design science (Brown, 1992; H. A. Simon, 1969; Wittmann, 1995) and knowledge created during curriculum development should be both generated and placed within a scientific research corpus, peer reviewed, and published. However, I do not promote a simple, causal, deterministic view of science. Nor, as the remainder of this article should make clear, do I privilege narrow views of scientific research, such as quantitative experiments (Dewey, 1929).
Consider the following polar positions.

For some, only when educational researchers agree about how one goes about creating knowledge in educational research will education start producing “reliable knowledge” (Zinman, 1978) and become a “proper science.” For others, however, the very nature of the educational activity—the complexity of the objects of study—means that educational research can never become a “science” in the traditional (and narrow) sense (Lester & Wiliam, 2002, p. 490).

Although I agree that the nature of educational research is complex and interpretive, I do not accept the “traditional” and “narrow” sense of the term “science” or believe that research can produce “truth” or a single correct view. I take a constructivist perspective that scientific research, as disciplined inquiry, provides reliable ways of dealing with experiences and pursuing and achieving goals (Glasersfeld, 1987). In the research process, conditions under which a phenomenon occurs are documented, a hypothetical mechanism for the phenomenon leads to a prediction of an event that has not yet been observed, and the conditions which the under which the mechanism should lead to the observation of the predicted event are generated (Glasersfeld, 1987, p. 117, summarizing Maturana’s formulation). In curriculum development research, the mechanism centers on a curriculum and its implementation.

Further, scientific research involves co-mutual influences within social and political contexts (Cobb, 2001) and is itself social and political (Latour, 1987), with researchers garnering support for their global perspectives, research issues, individual studies, and even results. That is, scientific research is not free from social-historical movements, values, controversies, politics, competition, status hierarchies, and egotism. Scientific knowledge advances are ultimately achieved by the “self-regulating norms” of a scientific community over time. One implication is that the goal cannot be to develop 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. Given that all such factors affect curriculum as well—probably to a much greater degree, 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, scientific research is necessary but not sufficient, for high-quality curriculum development.²

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

James argues that scientific knowledge is applied artfully to create teaching materials. Such research-to-practice methods are included in the framework. However, because it constitutes one-way translations of research results, a strict research-to-practice model 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 two basic issues—effect and conditions—in three domains, 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’ initial scientific base to formative and summative evaluation (Brown, 1992), and thus be integrated even into the most creative phases (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.

A final set of issues regarding the emphasis on scientific research should be addressed before describing the proposed framework. 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), 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 it’s 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 (National Research Council, 2002). To meet such needs, in politically-charged environs in which decisions have substantive financial and social ramifications, require 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.)

Finally, societally-determined values and goals are substantive components of any curriculum (Confrey, 1996; Hiebert, 1999; National Research Council, 2002; Schwandt, 2002; Tyler, 1949); curriculum research can not ignore or determine these components (Lester & Wiliam, 2002; Schwandt, 2002). Although some have criticized the scientific tradition as necessarily objectifying children (Lincoln, 1992), I believe that science as defined here does not do this inherently (see also Darling-Hammond & Snyder, 1992).
Nor can science determine value-free goals; my own constructivist position, with roots in Piaget, Dewey, and Vygotsky, include democratic values as a basis for an interactionalist and constructivist framework that values the dynamic negotiation of the needs of the individual and the democratic society—that is the realization of individual potential within a context of democratic rights and responsibilities.

**Research Bases for Curricula: A Framework**

To summarize, establishing, maintaining, and evaluating connections between curricula and research are problematic because many, if not most, developers and publishers claim that their curricula are based on research but few explicate the claims. Without an established framework for understanding or evaluating these claims, educators turn to other criteria in developing and selecting curricula. In addition, the potential of the curriculum development and evaluation context for building a coherent scientific knowledge base for education is left unrealized without such a framework. I propose a framework for *curriculum development research*. The framework builds upon many elements of previous works (e.g., Beauchamp, 1981; Jackson, 1992; Tyler, 1949), specifying research methods in place of several non-scientific components and processes and provides a coherent structure for development and evaluation in place of useful but separate techniques. For example, Walker (1992) advocated methods such as “simple, quick” field tests, which will not contribute to the theoretical and research corpus but are practicable in classrooms and provide feedback to developers. I agree with his goals, but find the compromise inadequate in that we also can, and need to, build empirical corpus and theoretical frameworks. His methods may work for producing an effective curriculum expeditiously, but the lack of systemacticity and documentation leaves many
questions, and, more importantly, does not provide knowledge that would be invaluable to future curriculum development efforts. In contrast, I contend that a scientifically-based curriculum requires answering the questions in Table 1 within a curriculum development research framework, so as to build comprehensive theories and implications that communicate with researchers, designers, and practitioners. Further, I propose that curriculum development research as proposed here provides an ideal site for building a scientific knowledge base for education and educational reform. The framework includes ten valid methods for asserting that a curriculum is based on research, which can be roughly ordered by the chronology of typical curriculum development (although these methods are cyclic) and classified into three categories (reflecting the three categories of knowledge required to meet Table 1’s goals, as previously described: existing research, models of children’s thinking and learning in a domain, and evaluation), as described in Table 2. The following sections describe the curriculum development research methods; the first three establish *a priori foundations* for the curriculum.

*A Priori Foundations*

1. **Subject Matter A Priori Foundation.** Establishing educational goals involves multiple considerations that go beyond science (see endnote 2). In this method, research contributes to the process by identifying subject-matter content that is developmentally appropriate and interesting to students in the target population (cf. Tyler, 1949). The research identifies those domains, learning about which would make a substantive contribution to students' development. That is, the 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). Further, research on complementary components of competence should be considered; in mathematics, for example, these might include problem posing and problem solving, metacognition, and a positive disposition toward learning and using the subject-matter content (Baroody & Coslick, 1998; Schoenfeld, 2002).

2. General A Priori Foundation. Broad philosophies, theories, and empirical results on learning and teaching are reviewed and considered when creating the curriculum using standards for research reviews (Light & Pillemer, 1984). For example, developers might start from an Ausubelian or “constructivist” perspective and proceed in any of several directions (Forman, 1993; Lawton, 1993). The researcher-developer team documents the decisions made and reasons for them. Ideally, one member of the research team is responsible, in this and all other methods, for taking a perspective of “standing outside,” observing, and documenting all activities of the curriculum development and research team (Lesh & Kelly, 2000).

3. Pedagogical A Priori Foundation. Empirical findings on making specific types of activities educationally effective—motivating and efficacious—serve as 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 method (indeed, research indicates it plays a major contributing role, Tamir, 1988), but not encompassing other (succeeding) methods.

Learning Model

A related method, the sole method in the second category, emphasizes learning models. Here, a tenet of this framework comes into sharp focus: Although the framework 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. This might involve two interrelated aspects. First, activities may be structured to be consistent with empirically-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 from mathematics, 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 #935) and profit from doing so more than from 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 and 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 (Investigations in Number, Data, and Space, see Mokros, 2003; van den Brink,
1991). As a specific illustration, Fuson (1997) describes 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. The curriculum moves through increasingly difficult types of word problems
based on the model, which leads to the next aspect of structuring activities in accordance
with models of learning. Extant models may be available, although they vary in degree of
specificity. Especially when details are lacking, developers use clinical interviews and
observations 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 exploratory
teaching. 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 (M.
A. Simon, 1995) through the concepts and skills that constitute a domain of mathematics (Clements, 2002; Cobb & McClain, 2002; Gravemeijer, 1999). These learning trajectories may be built upon natural developmental progressions identified in empirically-based models of children’s thinking and learning (Carpenter & Moser, 1984; Case, 1982; Griffin & Case, 1997). 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 strands are up to ten times the potential number of topics (Valverde, Bianchi, Wolfe, Schmidt, & Houang, 2002). 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, 2004b). An example is young children’s development of geometric composition abilities (See Fig 1). 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 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, usually using clinical interviews and teaching experiments as needed in this method (Clements & Battista, 2000).

Creation and use of such models always implies conceptual analysis; “the actual process of thinking remains invisible and so do the concepts it uses and the raw material of which they are composed” (Glasersfeld, 1987, p. 77). Therefore, the scientific processes (documenting conditions, hypothesizing mechanisms, predicting events, etc.) must be carefully followed and recorded. In addition, models of teaching and learning may also consider a wide variety of individual (e.g., prerequisite knowledge), social, and contextual aspects (e.g., Bauersfeld, 1980; Cobb, 2001; McClain & Cobb, 2001; Schofield, 1995; Secada, 1992). Indeed, throughout the methods, a concerted effort should be made to view the curriculum and the teaching and teaching process through multiple conceptual lenses (Schoenfeld, 2002), examining both underlying assumptions and data from as many alternate (ideally competing) perspectives as possible (Lester & Wiliam, 2002). This method, then, presents rich opportunities to add to the educational knowledge base.

The remaining six methods, in the third category, evaluation, involve collecting specific empirical evidence in marketing, formative, and summative research.

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 method
of research on commercial curriculum, 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.) This usually involves a close look at state standards, guidelines, and curricula (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 are often conducted by a research 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, features, assessment, and so forth. 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 (less frequently in the past decade), 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 curriculum research. In contrast, scientific market research could collect useful information about usability and probability of adoption and implementation. In the U.S., 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 could form early and sustained relationships with publishers to use findings from, or 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 would have the added advantage of connecting the scientific curriculum research to the types of information publishers are most familiar with, thus bridging the gap between developers and publishers that is especially common for innovative materials (Tushnet et al., 2000).

The following three methods are also forms of formative evaluation. In contrast to the prototype testing of typical market research, these methods often involve repeated cycles of design, enactment, analysis, and revision (Clements & Battista, 2000), with increasing grain size of the populations and the research variables.

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 teaching experiences, microgenetic, and microethnographic approaches (Ginsburg, 1997; Schoenfeld, Smith III, & Arcavi, 1993; Siegler & Crowley, 1991; Spradley, 1979; Steffe, Thompson, & Glasersfeld, 2000, note that specific theories and methodologies are proffered as illustrations rather than prescriptions, a point to which I return in the final section). The goal is to understand the meaning that students give to the curriculum objects and tasks (Lincoln, 1992).

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. 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 type is 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. DC: This is WAY SHORT of good description of LT; get into processes,

Often this is the most iterative research-design method; sometimes evaluation and redesign may cycle in quick succession, often as much as every twenty-four hours (Burkhardt et al., 1990; Char, 1990; Clements & Sarama, 1995; Cobb, 2001). Tasks may be completed 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 (see also 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. Videotapes (for microgenetic analysis), audiotapes, and field notes 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 also 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 begins with a model-testing approach, there remains significant adaptation to students’ input. Often, a free exploration phase 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 development research is that it studies what could be, unlike traditional research, which 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, the teaching experiments, especially the latter ones during this phase, are rich with both design and research 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. Also collected is more detailed information about the worth of various features of the teaching and learning interventions, some of which will be knowledge emerging from, and 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 (National Research Council, 2002).

7. Formative Research: Single Classroom. Although teachers are ideally involved in all methods (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). 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; support teachers' learning of that content, especially any topics that are new to teachers; and 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 help 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, 1996; for examples, see Clements, Battista, Sarama, Swaminathan, & McMillen, 1997; Gravemeijer, 1994a). Such classroom-based teaching experiments serve similar research purposes as traditional teaching experiments (Steffe et al., 2000), but are conducted in a naturalistic classroom setting. Videotapes and extensive field notes are required so that students' performances can be examined repeatedly for evidence of their interpretations and learning, for reasons similar to that of the previous method.

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 heavily to examine the teacher and students as they construct new types of classroom cultures and interactions together (Spradley, 1980). This 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 (B. Davis & Simmt, 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 method 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 with a high fidelity of implementation, as opposed to ascertaining how the curriculum works in classrooms in general, which is one focus of the following method. Thus, this is similar to what Cronbach and others (1980) called a “superrealization”—a painstaking assessment of what the curriculum can accomplish at its best. Regular meetings of the teacher and research group are requisite. Videotaping can also be used here as a source of data. All 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, policy makers, and researchers.

8. Formative Research: Multiple Classrooms. In several classrooms, the entire class is observed for information about the effectiveness and usability of the curriculum,
but more emphasis is placed 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 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 students' learning), as well as how they might be changed to 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.

A second question is whether the materials support teachers if they desire to delve
more deeply into their students’ thinking and teach differently. A third question is 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 own local setting. Again, ethnographic research (Spradley, 1979, 1980) is important, especially 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 method should determine the meaning that the various curricular materials have for both teachers and students. Materials for professional development are created, or revised, based on this research. In addition, qualitative methodologies 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.

9. **Summative Research: Small Scale.** This method evaluates 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. There are several critical assessment issues. First, standardized instruments (not necessary standardized tests as commonly construed) must have been chosen or developed (usually incrementally in the previous phases) as valid measures of the curriculum goals (National Research Council, 2002). Often, this involves at least two assessment components, one that is as 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., of 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 (with that fidelity reliably measured, recognizing that some curricula may be implemented in nonstandard, but appropriate, ways). 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.

Experiments are conducted in conjunction with, and to complement, methodologies previously described. Other approaches, including qualitative work, are stronger if conducted within a randomized experiment. For example, if teachers volunteer to implement the curriculum in a quasi-experimental design, neither quantitative nor qualitative techniques 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. The combined interpretive and survey data address whether such 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 more 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 to the outcomes is important for theoretical, development, and practical reasons. 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.

10. Summative Research: Large Scale. Commonly known is the “deep, systemic incapacity of U.S. schools, and the practitioners who work in them, to develop, incorporate, and extend new ideas about teaching and learning in anything but a small fraction of schools and classrooms” (see also Berends, Kirby, Naftel, & McKelvey, 2001; Confrey, Bell, & Carrejo, in press; Cuban, 2001; Elmore, 1996, p. 1; Tyack & Cuban, 1995; Tyack & Tobin, 1992). Thus, with any curriculum, but especially one that differs from tradition, evaluations must be conducted on a large scale (after considering issues of ethics and practical consequences, see Lester & 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 document the fidelity of the implementation of the curriculum across diverse contexts. That is, unlike the treatment standardization necessary to answer the questions of the previous methods, here it is assumed that implementation fidelity will vary (often widely), 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, 1991, 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. However, not all intervention classrooms will implement the program with equal veracity (research indicates that people who take advantage of all program components are more likely to benefit, Ramey & Ramey, 1998). Therefore, a series of analyses (e.g., hierarchical linear modeling, or HLM) would relate outcome measures with a set of target contextual and implementation variables, critical for identifying moderating and mediating variables (appropriate units of analysis—such as the class—should be defined). Further, the use of the “traditional” curriculum as the only comparison is unacceptable; instead, researchers should involve a wider variety of comparison curricula, including other innovative curricula, and should describe each comparison groups’ curricula and fidelity of implementation. 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).

Finally, summative evaluations are not complete until three criteria are met. (a) The curriculum has been implemented in multiple sites for more than two years, with full documentation of the contextual and implementation variables, including practical requirements, procedures, and costs. (b) Achievement data has been collected for more than two years, including longitudinal data. For both these criteria, note that two years of complete implementation may mean many previous years of preparation; for example, a project may implement a complete elementary curriculum one grade at a time and thus take more than 6 years to meet these criteria. (c) Evaluation have been confirmed by independent researchers unrelated to the developers of the curriculum (Darling-Hammond & Snyder, 1992).

A final approach is non-scientific (as is typical market research) and often contrived, but may be frequent in practice, and thus is mentioned for completeness. However, it is not a component of the proposed framework.

Post Hoc Rationalization. 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 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 methods that were veraciously used as the basis for the curriculum but never recorded. In this case, the justifications would merely be reporting that was, unfortunately, delayed. As argued previously, all methods should be recorded in detail and shared with the
greater community as part of the research process. In contrast, the chronology and
structure within which the requests for *Post Hoc Rationalizations* are frequently made
suggests that this method 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 methods used of the ten described and the results obtained with
each.

**Curriculum Development Research and Mathematics Curricula**

Mathematics education in the U.S. has a long history of connecting research with
curriculum development to varying degrees using various methods (further, curriculum
specialists such as Bobbitt, 1924; e.g., Schaff, 1960; Tyler, 1949, broadly described
certain research methods; Whipple, 1930, the latter describes “techniques of research”
that presage the methods described here). For example, the “Patterns of Arithmetic”
program (Braswell & Romberg, 1969) included several research strategies, including
basic research on learning and gathering feedback from participating teachers.
Researchers also 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 research,
and many were successful, although they generated only small amounts of summative
research (R. B. Davis, 1984). However, they often used mostly a priori foundation
methods and a scattering of the others. Further, research-based curricula have not been
nationally implemented over a sustained period (R. B. Davis, 1984). (Recall that many
other types of curricula, from guidelines, such as the National Council of Mathematics’
standards (1989; 2000), to “methods books” for teachers, have been influenced by research. Here we focus on “available curricula” for students.)

Unfortunately, many widely used mathematics textbooks of recent years have not built on that foundation (even to the extent of using a priori foundation methods) and have rarely employed the other research methods. Documentation of research methods used and results obtained by developers and publishers is often unavailable; thus, descriptions must rely on informal interviews, partial reports, and inferences. Brief descriptions are presented to further the discussion of these issues.

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, & Shoigreen-Downer, 1996). Ginsburg, Klein, and Starkey (1998) state that 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 (for both student and teacher) as possible by supplying simple activities and familiar routines (Ginsburg et al., 1998). The result is a false sense of innovation and research foundation (also research,
especially psychological research, is implicitly or explicitly used to establish a general a priori foundation. This reveals “the skill of publishers in including materials which appear to support the new aspects of the curriculum that are needed for adoption, 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).

What of materials based on theory and research? Even application of theories that are born in instruction, when limited to a general framework, 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 methods. In two studies, a curriculum based on the van Hielian theory, featuring informal experiences before formal arguments, 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 methods in the framework, with positive results. One is Realistic Mathematics Education (RME), whose “developmental research” is an integration of design and research (Gravemeijer, 1994b). A team begins the process by conducting an anticipatory thought experiment to formulate a hypothetical learning trajectory. Then a preliminary design is elaborated, refined, and adjusted in a series of intense cyclic processes of deliberations on and trials of instructional activities (Gravemeijer, 1999). Finally, the knowledge gained is used to construct an optimal instructional sequence. The goal is to develop and describe the local instruction theory (a more general description of the learning trajectories that emerged in specific classrooms) that underlies this entire instructional sequence and to justify it with
both theoretical deliberations and empirical data (Gravemeijer, 1994a, 1994b, 1999). The ideal is that such a local instruction theory will provide a framework that teachers can use to construe hypothetical learning trajectories that fit their own classroom situations. Recent collaborators with the Netherlands developers (McClain, Cobb, Gravemeijer, & Estes, 1999), Cobb and his colleagues have similar philosophical and curriculum development perspectives (Cobb & McClain, 2002; Gravemeijer, Cobb, Bowers, & Whitenack, 2000). Results reported in these publications document dramatic positive results both on wide-scale adoption of the Netherlands curriculum and on student outcomes.

The *Investigations in Number, Data, and Space* curriculum, a K-5 reform-based mathematics program. Some of the units (Akers, Battista, Goodrow, Clements, & Sarama, 1997; Battista & Clements, 1995a, 1995b; Clements, Battista, Akers, Rubin, & Woolley, 1995; Clements, Battista, Akers, Woolley et al., 1995; Clements, Russell, Tierney, Battista, & Meredith, 1995) were based on several research methods, with findings reported in the literature (Battista & Clements, 1996, 1998; Battista, Clements, Arnoff, Battista, & Borrow, 1998; Clements, Battista et al., 1996; Clements, Battista, Sarama, & Swaminathan, 1997; Clements, Sarama, & Battista, 1996, 1998; Clements, Sarama, Battista, & Swaminathan, 1996). (In contrast, many of the other units were built upon *a priori foundation* knowledge and informal research in classrooms. As with any curriculum, they may be as or more effective, but there is less documentation of their effectiveness.) 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, Battista et al., 1996), 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, Battista, Sarama, & Swaminathan, 1997), much less the specific gains on targeted mathematics achievement that these reports document.

Considered together, these recent projects show signs of using subject matter a priori foundation; general a priori foundation; pedagogical a priori foundation; structure according to specific learning model; formative research: small group; and formative research: single classroom, and, for some, summative evaluations as well. 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, reports an experiment showing that curriculum plans were significantly different for those teams given access to information about students). Important to the theme of the present paper, for several of these projects, instructional design served as a primary setting for the development of theory (Battista & Clements, 1996; Clements, Battista, Sarama, Swaminathan et al., 1997; 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). A summative research: small scale clinical randomized evaluation of one curriculum development project that was based explicitly on the framework described here resulted in effect sizes of 1.71 for number and 2.12 for geometry (Cohen’s d, Clements & Sarama, 2003, 2004a; Sarama & Clements, 2002). Achievement gains of the experimental group were thus comparable to the sought-after 2-sigma effect of individual tutoring (Bloom, 1984). Space constraints prohibit describing the many relevant research-based projects from the fields of mathematics education (e.g., Confrey, Castro-Filho, & Wilhelm, 2000; Confrey & Lachance, 2000; Hoyles & Noss, 1992; Hoyles, Noss, & Sutherland, 1989; see also numerous chapters in Lehrer & Chazan, 1998, and several articles in an upcoming special issue of the Journal of Educational Computing Research, “Design of Microworlds in Mathematics and Science Education”; Lewis & Tsuchida, 1998; Stigler & Hiebert, 1999; Yerushalmy, 1997) and cognitive science (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Brown, 1992; Griffin & Case, 1997; Lehrer, Jacobson et al., 1998; Lehrer, Jenkins, & Osana, 1998), and different conceptions such as didactical engineering (Artigue, 1994). The purpose here is not to be exhaustive, but to be illustrative of the less common methods that have been used and reported, and suggest that those that use multiple methods from the framework make substantive, unique contributions to theory, research, and curriculum development.

**Ramifications**

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

*Using the multiple methods in the proposed framework will help developers*
improve curricula and contribute to the field of curriculum research.

Good research is a matter not of finding the one best method but of carefully framing that question most important to the investigator and the field and then identifying a disciplined way in which to inquire into it that will enlighten both the scholar and his or her community. (Shulman, 1997, p. 4)

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 framework (cf. National Research Council, 2002, p. 4). For example, although iterating through one or two of the methods here, such as method 8, might lead to an effective curriculum, this 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. Without methods 9 and 10, the evidence on these and other matters, such as support requirements, would be lacking.

Using all the curriculum development research methods not only documents if 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, including the literature of the subject matter, psychology, and cognitive science, instruction, implementation, and scaling up. One example is the use of learning trajectories, employed in many successful projects (e.g., Clements, 2002; Cobb & McClain, 2002; Fuson et al., 2000; Gravemeijer, 1994b; M. A. Simon, 1995; Stigler & Hiebert, 1999).

More fundamental is the knowledge and use of the variety of scientific research methods in the framework (Kilpatrick, 1992, argues that, historically, the most successful
researchers have used a variety of techniques; cf. Mayer, 2000). Even if multiple methods are used, if they are all *a priori foundations*, for example, they are inadequate. Subtle differences in activities can enhance or sabotage effectiveness (Sarama, 2000, Martin A. Simon, personal communication, May 28, 2002). Achieving the goals of curriculum development research (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 proposed framework is coherent and connected throughout the development process maintains unbroken threads of argumentation.

Further, achieving the goals of curriculum development research requires both qualitative and quantitative methodologies. Both can make valid, rigorous contributions to scientific research (Darling-Hammond & Snyder, 1992; National Research Council, 2002). 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. 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). Experiments can 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).

Qualitative methodologies are as important for at least four reasons. First, at its heart, curriculum development research is concerned with the interacting interpretations
of quite different individuals. It seeks to understand individual students’ understanding 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 to answer why it works, and how and why it works differently in different contexts, essential for the sine qua non of research—adding to the knowledge base of curriculum development. Third, the valid measurement of certain variables (e.g., at the class and school level of an HLM analysis) Fourth, 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 (National Research Council, 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 development research requires qualitative methodologies and openness to emergent findings throughout the methods (Smith, 1983).

To repeat, then, a mixture of quantitative and qualitative methods is necessary,
and not just as separate, but also as integrated, studies. For example, every experiment benefits from collecting ethnographic data within all the groups. Conversely, the validity of qualitative methodologies, such as case studies, is increased if they are conducted within the context of an experiment (Cook, 2002). In a similar vein, the use of summative evaluation without other methods is usually premature, wasteful, and misleading. While randomized experiments remain the best design for evaluation of causal interpretations, and the need for summative, causal evaluations of curricula is constant, the limitations and misuses of randomized experiments (The Design-Based Research Collective, 2003) are mitigated by placing them in the context of a complete curriculum development research program as described here. The findings still can be generalized, but those generalizations are not as powerful alone as when they are combined with other approaches and recognition of the inherent complexity of educational science and the necessity of understanding human interactions (Berliner, 2002; Darling-Hammond & Snyder, 1992).

As a final word, the desirability of theoretical coherence suggests that researchers consider multiple methodologies in light of their epistemological and theoretical underpinnings (Lagemann, 1997; Paul & Marfo, 2001). I use constructivism (Glasersfeld, 1987) as a basic philosophical grounding (Clements & Battista, 1990; Clements, Battista, Sarama, & Swaminathan, 1997; Clements, Sarama et al., 2004) and employ specific theories and methodologies that have developed in mathematics education as its own research tradition (Kelly & Lesh, 2000); of course, different research and development teams will have their own foundations. Methodologies should be judged in the light of standards for theories, models, and results, such as descriptive power; explanatory power;
Curriculum Research  p. 39

scope; predictive power; rigor and specificity; falsifiability; replicability, generality, and
trustworthiness; and multiple sources of evidence (Schoenfeld, 2002, pp. 456-464).

Increasing university support for curriculum development 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 (H. A. Simon, 1969). Increasing
support and encouragement is justified at least two reasons. First and most directly, such
research, as described here, is legitimate science, characterized by firm theoretical
grounding and rigorous research methodologies. Indeed, curriculum development work in
complex learning situations has led to new directions in theory and empirical research.
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.
Similarly, editors and reviewers of research journals could legitimize the methods of
curriculum development research. Research using larger scale methods is more likely to
be accepted, with other methods misconstrued as merely “trying out teaching ideas,”
rather than research (Deborah Leong, personal communication, June 16, 2002). Such
biases discourage the conduct of research using these methods and limit feedback from
peers during critical phases of development.

As a field, curriculum developmental research could be more successful if funding
agencies reconsidered time frames and funding requirements for this enterprise. At least
until a much larger body of research-based development is created, there is a need for
increased funding for curriculum development research (Feuer et al., 2002). The
proportion of funds presently 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). Both curriculum development and experimental designs are costly, and implementing all methods in the framework is even more so. Paradoxically, using the full range of methods 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, in both development and design domains, 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 curriculum development research methods, 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. In the development of traditional 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). Curriculum projects that are funded usually are given implausible time frames that make such reflection and research (especially using the multiple methods in the present framework) nearly impossible, such as five years to develop five years of curriculum (Schoenfeld, 1999).

*To benefit from curriculum development research, the entire education community needs to demand and support research-based curriculum development—and demand that the specific methods used and results obtained are fully explicated.* Lack of a connection between research and curriculum development and adoption is a major
reason that curriculum, and ultimately student achievement, in the U.S. do not reliably improve (Battista & Clements, 2000; Clements, 2002) 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 (here defined broadly as disciplined inquiry) 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 eschew curricula that do not have the support of curriculum development research as defined here—the construct of “evidence-based” or “research-based” curricula is spurious without such criteria. This calls into question much of what is presently used in classrooms, which might be replaced as successful research-based curricula become available.7

Final Words

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, the 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. I believe that curriculum development research can help ameliorate these problems (Clements, Battista, Sarama, & Swaminathan, 1997; 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 (Whitenack, 1995), then research moves to the vanguard of educational innovation and results in substantive student achievement across the multiple goals of educational reform (National Research Council, 2002; Taba, 1962).

I argue that no curriculum should be produced without curriculum development research. I also argue that curriculum development research is one of the best ways to answer the three types of research questions (National Research Council, 2002), descriptive, causal, and process, within a program that is synergistic, integrated, and complete. Across the different methods, 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. Further, because it is result-centered, rather than theory-centered, curriculum development research 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). Curriculum development research makes the relationships among theory, research, design, and practice more salient and accessible to reflection.

All ten scientific curriculum development research methods need not, and often cannot, be employed in every project. However, curriculum development 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 methods that are applicable and tractable. Decisions to omit certain methods 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 experiments such as the premature evaluation of an innovation, depends on using all relevant methods.

While I believe these implications and guidelines are warranted, the main purpose main of this article is to begin a discussion of a framework for the construct of “research-based curricula.” Therefore, criticisms and alternative frameworks would be as welcome as agreements.
Table 1

*Goals of Curriculum Development Research*

<table>
<thead>
<tr>
<th>Practice</th>
<th>Policy</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the curriculum effective in helping children achieve specific learning goals? (What is the quality of the evidence?—Construct and internal validity.) (6-10)*</td>
<td>Are the curriculum goals important? (1, 5, 10)</td>
<td>Why is the curriculum effective? (all) What were the theoretical bases? (1, 2, 3)</td>
</tr>
<tr>
<td>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? (all)</td>
<td>What is the effect size? (9, 10)</td>
<td>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? (4, 6, 7)</td>
</tr>
<tr>
<td><strong>Conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When and where?—Under what conditions is the curriculum effective? (Do findings generalize?—External validity.) (8, 10)</td>
<td>What are the support requirements (7) for various contexts? (8-10)</td>
<td>Why do certain sets of conditions decrease or increase the curriculum’s effectiveness? (6-10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How do specific strategies produce previously unattained results and why? (6-10)</td>
</tr>
</tbody>
</table>

* Numbers in parentheses refer to the methods that contribute to answering each question.
Table 2

*Categories and Methods of Curriculum development research*

<table>
<thead>
<tr>
<th>Categories</th>
<th>Questions Asked</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A Priori Foundations.</em> In variants of the research-to-practice model, extant research is reviewed and implications for the nascent curriculum development effort drawn.</td>
<td>What is already known that can be applied to the anticipated curriculum?</td>
<td>Established review procedures (e.g., Light &amp; Pillemer, 1984) are employed to garner knowledge concerning psychology, education, and systemic change in general (method 1); the specific subject matter content, including the role it would play in students’ development (method 2); and pedagogy, including the effectiveness of certain types of activities (method 3).</td>
</tr>
<tr>
<td><em>Learning Model.</em> Activities are structured in accordance with empirically-based models of children’s thinking and learning in the targeted subject-matter domain</td>
<td>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)?</td>
<td>In method 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. What distinguishes method 4 from method 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, which is why it is classified separately from the a priori foundation methods.</td>
</tr>
</tbody>
</table>
**Evaluation.** In these methods, 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?

Is the curriculum usable by, and effective with, various sizes of student groups and various teachers? How can it be improved in these areas?

Formative methods 6 to 8 seek to understand the meaning that students and teachers give to the curriculum objects and activities in progressively expanding social contexts (methods 6, 7, and 8); 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 (method 7) and, later, by a diverse group of teachers (method 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?

Summative methods 9 and 10 differ from each other most markedly on the characteristic of scale. That is, method 10 examines the fidelity and sustainability of the curriculum when implemented on a large scale, and the critical contextual and implementation variables that influence its effectiveness. Experimental or quasi-experimental designs are useful for generating political and public support, as well as for their research advantages. Calculation and reporting of effect sizes and confidence intervals are considered sine qua non for these methods. In addition, qualitative approaches continue to be useful for dealing with the complexity and indeterminateness of educational activity (Lester & Wiliam, 2002).
Figure Caption

Figure 1. A Learning Trajectory for Composition of Geometric Shapes (Clements, Wilson, & Sarama, 2004)
**Pre-Composer.** Children manipulate shapes as individuals, but are unable to combine them to compose a larger shape. In free-form “make a picture” tasks, shapes often do not touch. In puzzle tasks, shapes do not match simple outlines. The instructional task (illustrated on the computer; similar tasks are presented with manipulatives and paper outlines or wooden form puzzles) uses outlines in which children can simply match shapes without turn or flip motions.

**Piece Assembler.** Children at this level are similar to Pre-Composers, but they can place shapes contiguously to form pictures. In free-form tasks, each shape used represents a unique role, or function in the picture (e.g., one shape for one leg). Children can fill simple frames using trial and error, but have limited ability to use turns or flips to do so; and have difficulty with nonsimple regions (i.e., those requiring multiple shapes). The instructional task first provides substantial spatial support for placements of individual shapes.

<table>
<thead>
<tr>
<th>Level</th>
<th>Examples (above, free-form pictures; below, puzzles)</th>
<th>Instructional Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Composer</td>
<td><img src="image1.png" alt="Diagram of shapes" /></td>
<td><img src="image2.png" alt="Instructional Task" /></td>
</tr>
<tr>
<td>Piece Assembler</td>
<td><img src="image3.png" alt="Diagram of shapes" /></td>
<td><img src="image4.png" alt="Instructional Task" /></td>
</tr>
</tbody>
</table>
shapes, but not every shape outline is provided, and then fades the support to include somehow more open areas (as illustrated here).

*Picture Maker.* In free-form tasks, children can concatenate shapes to form pictures in which several shapes play a single role (e.g., a leg might be created from several contiguous squares), but use trial and error and do not anticipate creation of new geometric shapes. For puzzle tasks, shapes are chosen using gestalt configuration or one component such as side length. If several sides of the existing arrangement form a partial boundary of a shape (instantiating a schema for it), the child can find and place that shape. If such cues are not present, the child matches by a side length. The child may attempt to match corners, but does not possess angle as a quantitative entity, so they try to match shapes into corners of existing arrangements in which their angles do not fit. Rotating and flipping are used, usually by trial-and-error, to try different arrangements (a “picking and discarding” strategy). Instructional tasks have considerable “open” areas in which shape selection is ambiguous.
Shape Composer. Children combine shapes to make new shapes or fill puzzles, with growing intentionality and anticipation. Shapes are chosen using angles as well as side lengths. Eventually, the child considers several alternative shapes with angles equal to the existing arrangement. Rotation and flipping are used intentionally (and mentally, i.e., with anticipation) to select and place shapes. The child has imagery of the component shapes, although imagery of the composite shape develops within this level. Instructional tasks (here, solving similar problems multiple ways) encourage higher levels in the hierarchy not described here, involve substitutions (three higher levels are described in Clements, Sarama, & Wilson, 2001).
References


Gravemeijer, K. P. E. (1994b). Educational development and developmental research in


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. (University Microfilms No. DA8628106)


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


52.


---

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 recent Interagency Education Research Initiative (http://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,
Curriculum Research 69

research reviews emphasize the need for scientific research as well (e.g., Kilpatrick, Swafford, & Findell, 2001; Walker, 1992).

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), or narrative (Bruner, 1986) 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. Indeed, most of the methods Walker categories as “humanistic” or “professional” could and usually should be included in scientific research as I define it 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). Finally, and perhaps most important, societally-determined values and goals are substantive components of any curriculum (Confrey, 1996; Hiebert, 1999; National Research Council, 2002; Tyler, 1949; Walker, 2003); curriculum research can not ignore or determine these components (cf. Lester & Wiliam, 2002). Although some have criticized the scientific tradition as necessarily objectifying children (Lincoln, 1992), I believe that science does not do this inherently (cf. Darling-Hammond & Snyder, 1992). Nor can or should science determine value-free goals; my own constructivist position, with roots in Piaget, Dewey, and Vygotsky, include democratic values as a basis for an interactionalist and constructivist framework that values the dynamic negotiation of the needs of the individual and the democratic society—that is the realization of individual potential within a context of democratic rights and responsibilities. Determining goals thus requires a dialectical process among all legitimate direct and indirect stakeholders (van Oers, 2003). [brief author example of the “who” and “how” to address goals and values here] Unlike groups such as the reconceptualists (Walker, 2003), however, I acknowledge limitations of science without rejecting its fundamental role.

There is a presage of the enormity of the challenge for the research community; for example, the generativity criterion requires extensive longitudinal work.

Design includes its own theories and processes. Examples are presented here only briefly (e.g., [author]). 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.

Although this is a main point of the paper, it deserves special attention. A reviewer of a previous draft of this manuscripts said that creation of curricula, empirical research, and theory were different activities and the manuscript should address one or the other.
Although a constructivist foundation is consistent with the qualitative goal of “seeking understanding” (Smith, 1983), curriculum development, as a design science, is also concerned with inference and prediction. Thus, quantitative techniques play a role whose contribution and limitations I understand from a constructivist perspective (and I take simple categorical descriptions as inference vs. understandings as useful ways to distinguish research corpora but not dichotomies). The major arguments concerning curriculum, however, do not depend on these assumptions. Curriculum research does necessarily put the scientist in a central role; that is, the philosophical perspective and methodologies of the researcher must be considered when interpreting and evaluating the findings.

Being based on research does not, of course, guarantee success—evaluation being one reason to conduct research—not does it speak to values and goals (cf. Hiebert, 1999; National Research Council, 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; M. A. Simon, 1995; Stigler & Hiebert, 1999); therefore, the curriculum and findings are not only applicable to other classrooms but also support those practices.