20 Design Experiments and Curriculum Research

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

Introduction

Building curriculum is a complex problem of creative engineering. At first blush, design experiments could appear to be the most appropriate strategy for connecting curriculum development to research. In this chapter, I argue that design experiments should play an important role, but also that a complete research program requires the inclusion of additional, complementary, research methods.

Curriculum development is not always connected to research, with unfortunate consequences for students and for the field (Battista & Clements, 2000; Clements & Battista, 2000). Ideally, curriculum development is a design science (Brown, 1992; Simon, H. A., 1969; Wittmann, 1995) with the goals of engineering a learning process and developing local instructional theories (Cobb et al., 2003). As a science, knowledge created during curriculum development should be generated and published within a scientific community. Because scientific advances are achieved ultimately by the self-regulating norms of a scientific community over time, the goal cannot be to develop a single ideal curriculum but, rather, dynamic problem solving, progress, and advancement beyond the present limits of competence (Dewey, 1929; Tyler, 1949). Science cannot tell you what your goals are (Hiebert, 1999); neither can it generate an ideal approach (James, 1958).

On the other hand, another implication is that curricula should be based on research, albeit using a complete framework as proposed here. The reason is that, contrary to a popular notion, research is not conducted by objective detached scientists. All research is social/political (Latour, 1987), with researchers attempting to gather support for their perspectives, research issues, methodologies, and interpretations. Particularly when financial gain is involved, such biases affect curriculum development and research; therefore, the checks and balances of science are essential to support progress and full disclosure.

What goals should we hold for a complete program of research on curriculum development? What methods should be used? When should they be used?

Goals and Research Activities for Scientifically-Based Curricula

A complete, scientific, curriculum development program should address two types of questions—about effects and conditions—in three domains: policy, practice, and theory (Clements, 2007). In the domain of policy, we ask whether the curriculum goals are important, what the effects are on teachers, and what the size of the effect is for different populations. Questions of conditions include the support requirements for various topics and variations in the effects and the sizes of the effects for different populations.
In the domain of practice, we need to know if a curriculum is effective in helping children achieve specific learning goals and to identify both its intended and unintended consequences. Conditions under which it is effective must be identified. Theoretical questions include why the curriculum is effective, what theoretical bases were used, to what degree they were explanatory, what cognitive changes occurred, and what processes were responsible. Conditional questions include why certain sets of conditions decrease or increase the curriculum’s effectiveness, how specific strategies produce previously unattained results, and why?

To answer these questions, we contend that writers and scientists must conduct four related types of research and development activities. Researchers do the following: (a) draw implications from existing research so that what is known can be applied to the anticipated curriculum, (b) develop curricular components in accordance with models of children’s thinking and learning in a domain, (c) conduct formative evaluations of research and refine the curriculum in a series of progressively expanding social contexts, and (d) perform summative evaluations, also in expanding contexts, to provide warranted evidence for claims of effectiveness.

The Role of Design Experiments

Design experiments play a significant role in research and development activities in a scientifically-based curriculum development program, but they are not sufficient in themselves. Design experiments were developed as formative research tools, to test and refine educational designs based on principles derived from previous research (Brown, 1992; Cobb et al., 2003; Design-Based Research Collective, 2003). Thus, they provide a useful set of tools for several phases of a complete research and development program (Clements, 2007). However, their unique focus confines their contribution to only some phases of such a curriculum research program, for at least four reasons. First, design experiments are often limited to pilot-testing (Fishman et al., 2004; NRC Committee, 2004: 75). Second, they tend to place less emphasis on the development of a curriculum itself and more on the development of local theories (or more abstract principles). Third, they do not address adequately the full range of questions or methods of the proposed framework (but see Bannan-Ritland, 2003). Fourth, they can be swept awash by too much data and too many interacting variables and possibilities, with too little theoretically-determined constraints. Fortunately, design experiments and other established methods, such as research reviews, teaching experiments, and formative and summative curriculum evaluation, complement each other. A comprehensive framework integrating all these methods is described in the following section.

A Curriculum Research Framework

The Curriculum Research Framework (CRF) consists of ten phases of the development research process that warrant claiming that a curriculum is based on research (this section summarizes and thus draws heavily upon the full description of the CRF in Clements, 2007). These ten phases are classified into four categories reflecting the types of questions asked to meet the previously described goals. The following sections describe the CRF’s cyclic phases.
A Priori Foundations

Activities in the first category establish the a priori foundations for the curriculum. A main activity is generating, or applying, extant, scientific research reviews to curriculum questions. These help to achieve the policy goal of establishing the importance of the curriculum’s educational objectives, the practice goal of credible documentation of the a priori research indicating the efficacy of the approach, and the theory goal of determining why the curriculum is effective and the validity of its theoretical bases. This category contains three interrelated phases. These phases could be, but are not always, used in design research projects.

Subject Matter A Priori Foundation

Establishing educational goals involves multiple considerations that include, but also go beyond, science. Socially determined values and goals are important determinants of any curriculum (Hiebert, 1999; Schwandt, 2002; Tyler, 1949). Thus, determining goals requires a dialectical process among all the legitimate, direct, and indirect interested parties (van Oers, 2003). This is consistent with the approaches of the reconceptualists and the poststructuralists (Pinar et al., 1995; Walker, 2003). In contrast to their position, I acknowledge that scientific approaches have limitations, but maintain that they make critical contributions, even in determining curricular goals. This research phase contributes to the process by identifying the subject-matter content that is valid within the discipline and makes a substantive contribution to the mathematical development of students in the target population (cf. Tyler, 1949). In other words, concepts and procedures of the domain that are being considered for inclusion in the curriculum should play a central role in the subject-matter domain itself (Tyler, 1949). They should build from the students’ past and current experiences (Dewey, 1902/1976). Finally, they should be generative in students’ development of future understanding (for an explanation and examples, see Clements et al., 2004). In mathematics, documents from the National Council of Teachers of Mathematics (NCTM; 2000, 2006) were created by a dialectical process among many legitimate interested parties, and thus serve as a valuable starting point, as are comparisons to other successful curricula and reviews of research. These research-oriented strategies constitute parts of comprehensive content analyses (cf. NRC Committee, 2004).

Other types of inquiry may be needed to complement scientific research, from perspectives such as historical (Darling-Hammond & Snyder, 1992), literary criticism (Papert, 1987), narrative (Bruner, 1986), aesthetic (Eisner, 1998), phenomenological (Pinar et al., 1995), or humanistic (Schwandt, 2002). 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: 19)

One final note applies to this and all the other phases. Ideally, one member of the
research team is responsible for taking a perspective of “standing outside,” observing and documenting the curriculum development and research team’s activities, decisions, and the reasons for their decisions (a “triarchic” design, see Lesh & Kelly, 2000).

General A Priori Foundation

Broad philosophies, theories, and empirical results on teaching and general curriculum issues are reviewed. Curriculum theory and research are studies for guidelines on students’ and teachers’ experiences with curricula, as well as on school and society, all of which helps to establish general education goals and directions (Pinar et al., 1995).

Pedagogical A Priori Foundation

Empirical findings on making specific types of activities educationally effective, that is, both motivating to students and effective in supporting students’ learning, are summarized to create general guidelines for the generation of instructional designs and activities. Intuition and creative work complement scientific information (Dewey, 1929; Hiebert, 1999; James, 1958).

Learning Model

The second category involves the determination and development of learning models. This phase addresses mainly the theory goal of identifying the cognitive changes that occurred and the processes, including the specific curricular components and features that are responsible for the changes. Here, it becomes clear that, although the CRF can be discussed in general, applied research and development are based in specific, subject-matter content, which cannot be added effectively to a general structure.

Structure According to Specific Learning Models

In this phase, the curriculum’s activities are structured to be consistent with domain-specific models of children’s thinking and learning. This may involve two strategies. First, activities might be designed to be consistent with empirically-based models of children’s thinking and learning, which can affect curriculum design substantially by focusing it on teaching and learning (Tamir, 1988; Walker, 1992). For example, research indicates that young children can invent their own solutions to simple arithmetic problems (Baroody, 1987; Carpenter & Moser, 1984; Ginsburg, 1977; Kamii & Housman, 1999; Steffe & Cobb, 1988) and that this benefits them more than direct instruction on prescriptive procedures (Hiebert et al., 1997; Kamii & Dominick, 1998; Steffe, 1994). Based on these findings, some curricula pose problems and ask children to figure out how to solve the problems and explain their solutions (Baroody, with Coslick, 1998; Fuson et al., 2000; Hiebert, 1999; Kamii & Housman, 1999).

Initial cognitive models may be available for a particular domain a priori. However, when models do not exist or do not give sufficient detail developers might use grounded theory methods (Strauss & Corbin, 1990) or clinical interviews to examine students’ knowledge of the content domain. Once a model has been developed, it can be tested and extended with teaching experiments, to build models of children’s thinking and learning that also suggest guidelines for instructional activities (Steffe & Thompson, 2000).

In the second strategy of this phase, sets of activities are developed and sequenced to
complete learning trajectories (Simon, M. A. 1995)—paths of learning and teaching the concepts and skills of the goal domain (Clements, 2002; Cobb & McClain, 2002; Gravemeijer, 1999). Such learning trajectories have been based on the historical development of mathematics and observations of children’s informal solution strategies (Gravemeijer, 1994b) and on emergent mathematical practices (Cobb & McClain, 2002). The CRF’s learning trajectories build directly upon the previously discussed cognitive models; that is, upon developmental progressions identified in empirically-based models of children’s thinking and learning (Carpenter & Moser, 1984; Case, 1982; Griffin & Case, 1997; Steffe & Cobb, 1988). Such 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, 2004: 83)

This detailed theoretical structure permits researchers to test theories by testing curricula (Clements & Battista, 2000). Such testing frequently involves teaching experiments and design experiments. Design experiments are preferred, as they emphasize a greater variety of types of interventions developed to support specific types of learning (with the latter emphasizing intervening to support particular forms of learning [Cobb et al., 2003]). As a science, these experiments include conceptual analyses and theories that “do real design work in generating, selecting and validating design alternatives at the level at which they are consequential for learning” (diSessa & Cobb, 2004: 77).

Design experiments are, therefore, a prime method of this; a critical phase of curriculum development. The CRF requires that all a priori foundation phases are conducted before and/or alongside design experiments and, most critically, that theoretically-based learning trajectories underlie the design experiments themselves.

Evaluation (Formative)

The next four phases, in the third category of evaluation, involve collecting specific empirical evidence in marketing and formative evaluations. The first of these, market research, addresses the policy goal of establishing the importance of the curriculum’s objectives and, to a limited extent, the theory goal of explaining why the curriculum is effective.

Market Research

Market research is about what customers want. Traditional market research is the most common type of research in commercial curriculum writing. It involves surveying state standards, guidelines, and curricula and tests, especially of the key adoption states, as well as widely-administered standardized tests. Prototype materials are prepared and presented to focus groups. Traditional market research fails to meet the standards for scientific research, but eschewing it as mere commercialism, as researchers often do, is unwise. In the United States especially, the products of those who ignore the concerns of publishers and teachers are not widely adopted (Tushnet et al., 2000). To meet the needs of research and marketability, researchers and developers should conduct studies that...
are grounded in the disciplines and reported fully (Jaeger, 1988). This has the additional advantage of connecting the scientific curriculum research to information with which publishers are comfortable, helping to bridge the gap between developers and publishers that is especially problematic for innovative curricula (Tushnet et al., 2000). Such market research might be conducted as a component of the a priori foundation phases through the last phase of planning for diffusion (Rogers, 2003).

The following three phases are types of formative evaluation. They address the practice goal of determining where the curriculum is and is not effective in helping children achieve specific learning goals and the theoretical goal of identifying cognitive changes and the responsible environmental attributes and processes. In addition, they begin to address conditions: for policy, what the support requirements are; for practice, under what conditions the curriculum is effective; and for theory, why certain sets of conditions decrease or increase the curriculum’s effectiveness and how specific strategies produce previously unattained results and why. These phases often involve repeated cycles of design, enactment, analysis, and revision (Clements & Battista, 2000), with increasing sizes of the populations and the research variables.

**Formative Research: Small Group**

In this phase, developers conduct pilot testing with individuals or small groups, evaluating particular curricular components, such as an activity, game, software environment, or longer section of the curriculum. Early interpretive work evaluates components using a mix of model- (or hypothesis-) testing and model-generation strategies. Therefore, design experiments play a major role, but methods also might include grounded theory, microgenetic, microethnographic, and phenomenological approaches (Pinar et al., 1995; Spradley, 1979; Steffe et al., 2000; Strauss & Corbin, 1990). The objective is understanding the meaning that students give to the curriculum objects and tasks and to alter these as necessary to better assist children in moving along the learning trajectory (Lincoln, 1992; Pinar et al., 1995).

Emphasis is placed on determining the accuracy of the learning trajectory, realized in practice. Such checking distinguishes this and subsequent phases from traditional formative and summative evaluations, which do not typically maintain connections to theory or create new theories (cf. Barab & Squire, 2004). The developer creates more refined models of the thinking of particular groups of students and describes what elements of the teaching and learning environment are observed as having contributed to students’ learning (Walker, 1992). Such connections help describe the knowledge and abilities that are expected of the teacher, laying the foundation for generating teacher support materials in the final curriculum.

This phase includes repeated, close, cycles of evaluation and redesign, sometimes within twenty-four hours (Burkhardt et al., 1990; Clements & Sarama, 1995; Cobb et al., 2003). Design experiments may be conducted in multiple classrooms, allowing revised lessons to be tested in a classroom staggered to be one to five days behind in implementing the curriculum (Flagg, 1990). Extensive documentation, such as field notes and audio-tapes or video-tapes, allows researchers to relate the findings to specific components and characteristics of the curriculum and specifically evaluate components of the design that were based on intuition and subconscious beliefs.
Although, ideally, teachers are involved in all of the phases, the emphasis in this phase is the process of curricular enactment (Ball & Cohen, 1996). First, classroom-based, teaching experiments are used to track and evaluate students’ learning, with the goal of making sense of the curricular activities as they are experienced by individual students (Gravemeijer, 1994a; Pinar et al., 1995).

Second, the class is observed to evaluate the usability and effectiveness of the curriculum. Ethnographic participant observation is used heavily to examine the teacher and students as they interact to establish a classroom culture (Spradley, 1980). Such observation is important because events and properties cannot always be predicted or understood solely from analyses of the components, but emerge within a complex system (Davis & Simmt, 2003). Focus is on how teachers use materials and guide students through activities, what classroom processes are prominent and how these processes connect to intended and unintended outcomes. In this phase, the learning trajectory per se remains the focus of the formative evaluation; therefore, teachers who can enact the curriculum in reasonable harmony with the developers’ vision—often the developers or teachers working closely with them—implement the curriculum.

**Formative Research: Multiple Classrooms**

In contrast to the previous phase, several classrooms involving teachers not closely connected to the developers are observed to learn about the effectiveness and usability of the curriculum. The emphasis turns to the curriculum’s usability, conditions that influence the curriculum’s effectiveness, and how it might be revised to serve all teachers and students better. Innovative materials, including those created in research and development projects, many times provide less support than the traditional materials with which teachers are familiar (Burkhardt et al., 1990), increasing the need for ecologically-valid studies for such materials. Understanding how and why the curriculum works in various contexts aids researchers in developing theory and practitioners in implementing the curriculum in local settings. Ethnographic research (Spradley, 1979, 1980) is useful, especially as teachers may appear to agree with a curriculum’s goals and approach, but their implementation may be inconsistent with the developers’ intention (Sarama et al., 1998).

**Evaluation (Summative)**

The final two phases are forms of summative evaluation. They address the policy goal of establishing the size of the effect for teachers and for students, the practice goal of documenting whether the curriculum is effective in helping children achieve learning goals, and the theory goal of identifying why the curriculum is effective. In addition, they address the conditions: *in policy*, the support requirements for various contexts; *in practice*, the conditions under which the curriculum is effective; and *in theory*, explanations as to why certain sets of conditions decrease or increase the curriculum’s effectiveness and how and why specific strategies produce previously unattained results.

**Summative Research: Small Scale**

In this phase, researchers evaluate what can be achieved with typical teachers under realistic circumstances (Burkhardt et al., 1990; Rogers, 2003). In four to ten class-
rooms, pre- and post-test, randomized, experimental designs using measures of learning are used. Those on the forefront of innovative research and curriculum development often eschew such experiment; however, they are the most efficient and least biased designs to determine causal relationships, and criticisms often confuse limitations of the design with misapplications of it (Cook, 2002). Further, innovators may reject logical positivism, noting that theories cannot be tested definitively and that curriculum development and research are social in nature. However, this does not imply that experiments cannot contribute to evidence on causal claims. It does imply that experiments should be carefully designed to have greater explanatory power, by connecting specific processes and contexts to outcomes so that moderating and mediating variables are identified (ibid.).

In the CRF, experiments are conducted in conjunction with methods described previously. These methods, including qualitative work, are actually stronger if conducted within the context of a randomized experiment. For example, neither quantitative nor qualitative techniques alone will discriminate easily between the effects of an intervention and the teachers’ dispositions and knowledge that led to their decision to volunteer for a quasi-experimental study. Experimental designs require that the intervention is described fully and explicitly and can be implemented with fidelity, according to the definition adopted. The curriculum used in the comparison classrooms should be selected on a principled basis and described completely. In addition, the quantity and quality of mathematics instruction must be measured in all participating classrooms with common measures. In addition, a combination of survey and interpretive information helps determine whether teachers view the extant supports as adequate and whether the intervention has altered their teaching practices.

Such summative research extends traditional summative evaluations. In the CRF, theoretical frameworks are a sine qua non. Comparison of the scores outside a framework is inadequate. In addition, connecting specific attributes of the curriculum, its enactment, and outcomes provides an adequate basis for contributing to theories of learning and teaching in complex settings and advising future curriculum development, as well as for implementing the curriculum in diverse contexts.

**Summative Research: Large Scale**

Increased attention has been given to the unique problem of scaling up a curriculum or other innovation that has proven successful in small-scale evaluations (Berends et al., 2001; Cuban, 2001; Elmore, 1996; Tyack & Tobin, 1992). With any curriculum, but particularly innovative curricula, researchers conduct evaluations to address this problem directly. Issues include the curriculum’s effects in contexts where implementation is usually expected to vary widely (Cook, 2002) and the critical variables, including the contextual variables—settings, such as urban/suburban/rural; type of program; class size; teachers’ characteristics; students’/families’ characteristics—and the implementation variables—engagement in professional development opportunities; fidelity of implementation; leadership and support; peer relations; and incentives used (Berends et al., 2001; Cohen, 1996; Elmore, 1996; 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 and a series of analyses such as hierarchical linear modeling provides correct estimates of the effects and standard errors when the data are collected at several levels. Because no set of experimental variables is complete or appropriate for each situation, qualitative methods supplement these analyses to ascertain the significant meanings, relationships, and critical variables that affect
implementation and effectiveness (Lincoln & Guba, 1985) and thereby link meaningfully implementation processes to learning outcomes.

To complete this phase, the curriculum must be sustained and evaluated in multiple sites for more than two years (Berends et al., 2001; Fishman et al., 2004) and evaluations must be confirmed by researchers unrelated to the developers of the curriculum (Darling-Hammond & Snyder, 1992), considering issues of adoption and diffusion of the curriculum (Fishman et al., 2004; Rogers, 2003).

Conclusions
Design experiments are an important component of curriculum research and development activity within the proposed Curriculum Research Framework (CRF). However, this framework shows that other research and development strategies are necessary to meet the goals of a complete curriculum research and development program. Design experiments cannot control the many variables in their complex settings; the large amount of data collected rarely can be analyzed fully before the next cycle of revision, enactment, and analysis takes place (Collins et al., 2004); and different participants may have different data and perspectives, so that the ultimate paths and products may be arbitrary to an extent and generalization may be difficult (Kelly, 2004). Experiments (randomized trial designs) provide some of what design experiments cannot. In addition, the use of phases in the A Priori Foundations and Learning Model categories of the CRF provide useful constraints and theoretical groundings for design experiments.

Conversely, design experiences, as well as other methods such as teaching experiments and classroom-based teaching experiments, can help accomplish what randomized trials cannot. In the CRF context, these methods include conceptual and relational, or semantic, analysis, and thus are theoretically grounded. As such, they allow researchers to build models of the child’s mathematics, of mental actions on objects, of learning, and of teaching interactions (L. Steffe, personal communication, July 18, 2005). Because it includes a coherent complement of methods, the CRF has built-in checks and balances that address the limitations of each method, with concentration on the learning model especially useful for maintaining theoretical and scientific foci.

In summary, embedding design experiments within the CRF encourages the conduct of a comprehensive research and development program. Such a program is theoretically sound and scientifically defensible. Further, it addresses the full range of curriculum research questions, about effects and conditions in the three domains of policy, practice, and theory (Clements, 2007). CRF-based programs thus contribute to both the research field and produce a curriculum that is a contribution to practice.

Acknowledgments
This chapter was supported in part by the National Science Foundation under Grant No. ESI-9730804, and by the Institute of Educational Sciences under Grant No. R305K05157. Any opinions, findings, conclusions, or recommendations expressed in this chapter are those of the author and do not reflect necessarily the views of the National Science Foundation or the Institute of Educational Sciences.
References


