7 Multitiered Design Experiments in Mathematics, Science, and Technology Education

Richard A. Lesh
Indiana University

Anthony E. Kelly
George Mason University

Caroline Yoon
University of Auckland

Introduction

This chapter describes an evolving new class of collaborative research methods called multitiered design experiments. After a brief introduction to the general concept of such studies, the chapter is divided into four sections. The first compares and contrasts the origins of the term design research in cognitive science versus mathematics and science education. The second describes how design research requires a reconceptualization of many traditional notions about researchers’ roles, subjects, relevant theories, and the nature of results. The third gives a concrete example of a multitiered design experiment that investigated the development of knowledge for three interacting types of subjects—students, teachers, and experts who know about the kinds of mathematical knowledge needed for success beyond school. The fourth describes two common design flaws in design experiments; then, several quality assurance principles are described which should apply to the ways that such studies are conducted.

Origins of the Term Design Experiment in Mathematics and Science Education

In the cognitive sciences and learning sciences, the term design research generally is considered to have been introduced in two articles: Ann Brown’s 1992 article in the *Journal of the Learning Sciences* about theoretical and methodological challenges in creating complex interventions in classroom settings and Alan Collins’ 1992 article describing steps toward a design science of education in the volume *New Directions in Educational Technology*, edited by Scanlon and O’Shea. In each of these articles, the basic notions of design research were borrowed from design sciences such as architecture or engineering where many of the most important kinds of systems that need to be understood were created by humans and where there is a long history of intertwining theory development with artifact development. In Brown’s case, design research methodologies were introduced explicitly to help increase the relevance of theory to practice—or, more precisely, to increase the relevance of cognitive science laboratory experiments to teaching, learning, and problem-solving activities in school classrooms. But, in Collins’ case, the opposite influences also were apparent; that is, being a software developer as well as a theory developer (i.e., a researcher), Collins was aware that the
best practices of the best practitioners often are more sophisticated and powerful than
the best theories of the best theorists. Consequently, a key reason for Collins to adopt
design research methods seems to have been to enable cognitive science researchers to
benefit from the wisdom of practitioners.

In mathematics and science education, research methods that are coming to be called
design experiments certainly were influenced by concerns similar to those that were
emphasized by both Brown and Collins (Collins, 1992; Hawkins & Collins, 1992); but,
other factors and concerns were perhaps even more significant (Lesh, 2002; Lesh &
Kelly, 2000; Lesh & Lamon, 1994). For example, like Brown and Collins, the multi-
tiered design experiments that are emphasized in this chapter were intended to reduce
the gap between theory development and practice—often by involving teachers and
other practitioners as co-researchers. Another goal was to increase the cumulativeness
of research by focusing on coordinated sequences of studies that build on one another
over longer periods of time. This latter characteristic is important because few realistically
complex problems in mathematics or science education are likely to be solved using
only single isolated studies. Consequently, the kinds of multitiered design experiments
that are emphasized in this chapter were developed explicitly to enable multiple
researchers representing multiple practical and/or theoretical perspectives to work
together at multiple sites over year-long periods.

Among mathematics and science educators, the origins of design research methods
also were more gradual and less explicit than among cognitive scientists. For example,
in the Handbook of Research Design in Mathematics and Science Education (Kelly &
Lesh, 2000), many of the innovative research methods that were described as having
been pioneered by mathematics and science educators could have used the language
development of design experiments to describe their work. One reason this was true surely was
because, like Collins, most productive mathematics and science education researchers
also function as curriculum developers, software developers, program developers,
teacher developers, and/or student developers (i.e., teachers). In other words, most
researchers in mathematics and science education also are practitioners in addition to
being researchers. Therefore, they tend to be heavily engaged in design enterprises
where knowledge development and artifact development interact. Consequently, it was
natural for them to adopt general ways of working and thinking that were borrowed
from the design sciences. Furthermore, because most productive mathematics and sci-
ence education researchers received their early research training primarily in the natural
sciences, more than in the cognitive or social sciences, they were accustomed to the
culture of design and modeling research where: (a) models for describing complex
systems are some of the most important products of research, (b) models are not true or
false but instead are useful, not useful, or relatively useful (under some circumstances
that need to be specified), (c) useful models of complex situations usually must draw on
more than a single theoretical or practical perspective, and (d) useful models usually
need to be developed using a series of iterative modeling cycles. In particular, they were
familiar with laboratory experiences that involve modeling and measuring “things”
that cannot be observed directly—and “things” whose nature may be changed signifi-
cantly through the process of observation. Consequently, they were familiar with
research whose goals involved investigating and measuring “objects”—ranging from
neutrinos to gravitational fields—whose existences make little sense apart from the
systems in which they are hypothesized to function.

In spite of the fact that mathematics and science educators tend to have natural
affinities for design research methods, it was only in later documents, such as a chapter
by one of the authors in the Handbook of International Research in Mathematics
Experiments in Mathematics, Science, and Technology Education 133

*Education* (Lesh, 2002), that the language of design experiments began to be used to describe this work. In earlier work that involved the design of a complex artifact and iterative cycles of development, mathematics and science educators tended to describe their work using the language of teaching experiments (Cobb, 2000; Kelly & Lesh, 2000; Simon, 2000; Steffe & Thompson, 2000; Verschaffel et al., 1999), clinical interviews (Clement, 2000; Lesh et al., 2000), action research (Confrey, 2000), participant observation studies (Ball, 2000), iterative video-tape analyses (Lesh & Lehrer, 2000), or projects in which curriculum development, software development, program development, or teacher development played central roles (Clements & Battista, 2000; Cline & Mandinach, 2000; Doerr & Lesh, 2002; Koellner-Clark & Lesh, 2003; Schorr & Lesh, 2003). One reason why such language was used was because the preceding methods evolved out of Piaget’s biology-inspired theories of cognitive development (Lesh, 2002) or out of Soviet-style teaching experiments (Kilpatrick et al., 1969), more than out of laboratory experiments associated with computer-inspired, information-processing theories in cognitive science. Thus, compared with notions of design research that prevail in the cognitive sciences and learning sciences, mathematics and science educators tend to think of such methods and models in ways that are far more organic and less rooted in machine metaphors (or computer metaphors) for characterizing the conceptual systems used by students, teachers, or researchers (Lesh & Doerr, 2003).

Finally, in mathematics and science education research, the nature of design research was influenced strongly by the fact that many of the most important subjects that we seek to understand involve systems that are complex, dynamic, interacting, and continually adapting. For example, such systems often involve many levels and types of interacting agents that have partly conflicting goals; and, they also often involve feedback loops, second-order effects, and emergent properties of the systems as a whole that are not derived merely from properties of the parts of the systems. Consequently, these systems generally cannot be characterized using only a list of simple input–output rules or simple declarative statements. As a result, in mathematics and science education, the most important products of design research seldom are reducible to simple hypotheses that are tested or to simple questions that are answered. Instead, the products of design research often are models, or related conceptual tools, which are not true or false but whose success is determined by criteria such as usefulness (in a specified situation), shareability (with others), and reusability (in other situations) (Lesh, 2002). In other words, appropriate quality assurance procedures are similar to those used in the design sciences—ranging from architecture to aeronautical engineering—where: (a) researchers and developers are comfortable with demands for a high level of accountability, (b) development usually involves both design and experimentation, (c) development usually draws on more than one theoretical perspective, (d) research teams typically involve people with diverse practical or theoretical perspectives, and (e) rigorous cycles of testing and revision generate auditable trails of documentation that contribute to the development of both knowledge and complex artifacts (which may range from spacecraft to biological ecosystems).

An Introduction to Multitiered Design Experiments

In addition to the characteristics of design research that were emphasized in the preceding section, the kind of design experiments that we wish to emphasize here often involve multiple tiers because they were created explicitly to investigate the interacting development of several levels or types of subjects (students, teachers, researchers, and developers) each of whom can be understood only incompletely, from the authors’
perspectives, if the development of the others is not taken into account. For example, in a typical multitiered design experiment: (a) students develop models for making sense of a problem-solving situation (which was designed by teachers and/or researchers), (b) teachers develop models for making sense of students’ modeling activities, and (c) researchers develop models for making sense of the interactions among teachers’ and students’ modeling activities.

The authors’ multitiered investigations involve design because they often investigate the nature of subjects’ ways of thinking by engaging them in activities that entail the parallel and interactive development of two distinct, but closely related, types of products:

1. Complex artifacts (or tools) that are needed for a specific purpose that lies (to some extent) outside the prejudices imposed by specific theories. Therefore, the assessment of success is not determined completely by the theories that are used to produce the relevant artifacts.

2. Underlying conceptual systems (e.g., theories or models) that are embodied in the relevant artifacts (or tools). Therefore, when the artifacts (or tools) are tested and revised, the underlying conceptual systems also are tested and revised.

Finally, the kind of multitiered design investigations that we emphasize tend to involve experimentation because the development of both the artifacts and the underlying conceptual systems usually involves a series of iterative cycles of testing and revision. Consequently, each temporary “product” tends to be the nth step in a continuing series.

This is especially true in mathematics and science education because many of the “subjects” that we seek to understand involve systems that are partly products of human creativity (or design) and because as soon as we develop better understandings of such systems, this understanding is used to change them. As a result, inferences about the ultimate directions for development tend to emerge from trends that become apparent in auditable trails of documentation that are generated across time—not from fragments of information collected at any given moment. In fact, the kind of multitiered design experiments that we use sometimes have been characterized as longitudinal development studies in mathematically enriched learning environments (Kelly & Lesh, 2000).

The Nature of Researchers, Subjects, Results, and Relevant Theories Found in Multitiered Design Experiments in Mathematics and Science Education

In some ways, the kinds of systems that need to be understood in mathematics and science education are even more complex than many of the physical systems found in the physical sciences. For example, the developing conceptual systems of both teachers and students are complex in their own right but their development involves even more complexity because they interact. Consequently, such complexity often creates the need to reconceptualize traditional notions and assumptions about researchers’ roles, subjects, the nature of results, and theories used to explain results. These topics are expanded below.

Researchers

Instead of emphasizing only the one-way transmission of research into practice (or practice into research), multitiered design experiments in mathematics and science
education often emphasize bidirectional interactions and feedback loops that involve many levels and types of participants (students, teachers, researchers, curricula designers, policy-makers). Furthermore, many participants may play more than a single role (e.g., participant-observers, researcher-developers, teacher-researchers), and a given participant’s roles often range from identifying and formulating problems to be addressed, to gathering, filtering, and organizing information that is most relevant, to interpreting results, to other key roles in the research process.

Subjects

Most of the subjects that need to be understood are (or involve) complex adaptive systems (Casti, 1994; Hmelo et al., 2000; Holland, 1995)—not necessarily in the strict mathematical sense—but at least in the general sense that these systems are dynamic, interacting, self-regulating, and adapting continually. Furthermore, among those systems that are most important for mathematics and science educators to investigate, many do not occur naturally (as givens in nature) but, instead, are products of human construction. Also, many cannot be isolated because their entire nature tends to change if they are separated from the systems in which they function; many of their most important characteristics cannot be observed directly but can be known only indirectly through their effects on other agents or events; and most participate in “communities of practice” (meta-agents), where interactions involving agents and meta-agents also are significant. Two examples are given below to describe subjects when design experiments are useful:

1. The primary subjects may be programs of instruction. In large systems that characterize complex programs of instruction, the hearts of important subsystems often are the conceptual systems that various agents use to create productive teaching, learning, or problem-solving experiences. Complexities arise not only because “within-agent” (conceptual) systems are complex, but also because of “among-agent” (interaction) systems where: (a) multiple agents have conflicting agendas, (b) trade-offs need to be considered among factors such as costs and quality, (c) feedback loops lead to second-order effects that outweigh first-order effects, and (d) emergent characteristics of the system as a whole cannot be derived from characteristics of the individual agents within the systems. The behaviors of such systems may be inherently unpredictable in the sense that exceedingly small differences between two systems often result in radically different outcomes because of resonances associated with patterns of interaction. Similarly, when they are observed, or when information is generated about them, changes often are induced that make researchers (and assessments) integral parts of the systems being investigated or measured.

2. The primary subjects may be students’ or teachers’ conceptual systems; and, instead of focusing on only procedural capabilities (e.g., “What kinds of computations are students proficient at doing?” or “What behaviors characterize effective teaching?”), relevant research may focus on conceptual capabilities (e.g., “What kind of situations can students describe mathematically so that available computational skills can be used?”). But, when students or teachers develop an interpretation of a situation, their interpretation abilities tend to change; and, this is especially likely to happen if the interpretations are embodied in sharable and reusable tools (Lesh & Doerr, 2003). So, tasks that are used to observe students’ interpretation abilities also tend to induce changes in these interpretation abilities; and again, researchers tend to become integral parts of the systems being investigated.
The second of the preceding examples illustrates why simple-minded aptitude treatment interaction studies did not work—especially for cases that involved interpretation abilities. The prediction was that when a person with attributes \((a,b,c, \ldots n)\) encountered a treatment with characteristics \((\alpha,\beta,\chi, \ldots \eta)\) then the result would be \((X,Y,Z)\)—or more simply, when \(A\) goes in, \(X\) comes out. Such simplistic input–output models might work nicely for systems that involve no interacting agents and no feedback loops; but, they have not proven to work well for complex adaptive systems whose most essential characteristics are that they cannot be modeled adequately using nothing more than simple input–output rules.

For similar reasons, simple input–output models are not likely to be useful for describing the kind of complex adaptive systems that characterize most successful programs of instruction. For such programs, the action tends to be in the interaction, in feedback loops, and in adaptability; and, the kind of functions that need to be monitored and manipulated tend to be dynamic and systemic—not static, rigid, and piecemeal.

**Implications**

When researchers design something, the underlying design (or conceptualization) often is one of the most important parts of the product that is produced because adaptability depends heavily on the quality and clarity of these underlying design principles.

Similarly, when researchers construct something, the underlying constructs (or concepts and conceptual systems) often are among the most important parts of the products that are produced. Or, when researchers model something, the underlying conceptual model may be among the most important parts of the result that is produced. Consequently, in all such cases, relevant research typically must include knowledge development as well as artifact development; and, in research that involves complex adaptive systems, useful results seldom can be expected to be reducible to lists of tested hypotheses or answered questions. For example, some of the most important pieces of knowledge that need to be produced tend to be models (or other types of conceptual tools) for constructing, describing, or explaining complex systems.

For the preceding kinds of reasons, for the kinds of design research that are emphasized in mathematics and science education, it is important to distinguish between: (a) model-developing studies and model-testing studies, (b) hypothesis-generating studies and hypothesis-testing studies, and (c) studies aimed at identifying productive questions that should be investigated versus those that are aimed at answering questions that practitioners (such as policy-makers) already consider to be priorities. For example, it may do little good to use sophisticated control group procedures and highly objective pre- and post-test designs to show that “it works” (for a given instructional program or curriculum innovation), if it is not clear what “it” is and what “works” means. And, this is especially true in cases where it is clear that nothing works unless relevant agents make it work.

An analogy can be made with the world of cooking. For example, does a recipe in a cookbook work? For whom? For what purposes? Under what conditions? Probably, if any recipe could claim the prize for working best (without regard to the preferences of the customers, the competencies of the cook, the quality of the ingredients and equipment, the weather, or the climate), it would be the recipe for **Toll House** chocolate chip cookies. Yet, if several mathematicians are given the same cookie recipe to follow at home, the results are sure to be several considerably different products. Conversely, if several superb cooks make **Toll House** chocolate chip cookies, they are sure to modify
the recipe to suit their own preferences, the characteristics of their resources and cooking environments, and the preferences of the people who are expected to eat the cookies. In fact, malleability, not rigidity, tends to be one of the most important hallmarks of a great recipe. In other words, a good recipe, like a good instructional activity or instructional program, usually needs to be shareable and reusable; it also should be easy to modify to fit a variety of teachers, students, and circumstances. Of course, cooking becomes much less algorithmic when it involves varying degrees of fresh fruit, vegetables, meat, herbs, and so on. Tasting and adjusting become as important as following the rules.

The results of multitiered design experiments often are conceptual tools that have some of the same characteristics as tools such as chainsaws, airplanes, automobiles, or cookbooks. There is no such thing as the best saw, plane, car, or cookbook. Each is designed for specific purposes, for specific people who will use them, for specific conditions, and under certain constraints (such as those that involve costs and available materials). Tools with different profiles of characteristics may be most useful for different purposes, people, preferences, and circumstances; to make judgments about usefulness, trade-offs often need to be considered that involve conflicting factors such as costs and quality. Likewise, every subject (or system) that is studied in multitiered design experiments has a complex profile of characteristics; no profile is “good” in all situations. In fact, successful subjects tend to be those who can manipulate their own profiles to fit changing circumstances, not those who adopt a single, fixed, and rigid profile. Partly for these reasons, useful assessment results generally need to look more like the information that is given in consumer guidebooks (for purchasing automobiles or other commodities) or in reports monitoring the progress of complex businesses (which typically use a variety of benchmarks, indicators, metrics, and other devices for tracking progress toward the goals specified in strategic plans).

A second way in which the results of multitiered design experiments are similar to tools such as saws, airplanes, cars, and cookbooks is that they tend to induce significant changes in the situations where they are intended to be used; in other words, their existence usually leads to new problems, new opportunities, and new purposes for similar tools in the future. Therefore, the existence of new tools generally leads to the need for even newer tools. This is why it is important to realize that results in the form of conceptual tools are not final but are really the nth step of an ongoing process.

Relevant Theories

In fields such as engineering, or in other mature design sciences that are heavy users of mathematics, science, and technology, experienced researchers consistently emphasize that realistically complex problems usually cannot be solved by relying on only a single theory, a single discipline, or a single textbook topic area. Why? Woody Flowers, one of the creators of the Massachusetts Institute of Technology’s famous undergraduate robotics design laboratories, answers this question in a recent personal communication when he noted: “The primary characteristic of a design project is that you always have conflicting goals (e.g., quality and costs) and too few resources (e.g., time and money). Trade-offs need to be considered, and compromises need to be made.” Similarly, in education, a classroom teaching decision that seems to involve only issues of cognitive science at first may turn into a classroom management issue, which turns into an administrative policy issue, and so on. As Dewey and other pragmatists argued long ago, in realistically complex decision making, what is needed is a framework capable of integrating ways of thinking from multiple points of view (Dewey, 1938, 1981).
Contrary to the naïve wishes of those who hope to use design research as a methodology for translating (a single) theory into practice, and contrary to the naïve claims of those who insist that each research project should be based on (a single) theory, design research in realistically complex settings should be expected to be multidisciplinary, multiaudience and multistakeholder.

With the preceding perspectives in mind, it is unrealistic to expect one-to-one correspondences between research projects and solved problems. It is at best a half-truth to claim that research projects should answer questions that are priorities for practitioners (such as teachers or policy-makers) (Lesh, 2002). To generate useful solutions for realistically complex problems, it is more reasonable to expect a two-step process in which: (a) research should inform theory (or coherent bodies of knowledge), and (b) theory should inform practice. Yet, even this conception of the relationship of theory to practice is overly simplistic. This is because few realistically complex problems are likely to be resolved by drawing on only a single theory. So, the kind of knowledge that can be expected to be most useful may be models which explicitly integrate ideas drawn from multiple practical and theoretical perspectives.

In model-development research, cumulativeness (shareability, reusability, modifiability) is a primary factor to emphasize in order to increase the quality of research. This is why useful research should be expected to design for shareability, reusability, and modifiability—rather than simply testing for these characteristics. It also is why emerging new research designs in mathematics and science education often emphasize collaborations over extended periods of time, at multiple sites, with multiple interacting subjects, and with investigators representing a variety of practical or theoretical perspectives. Finally, it is why coordinated clusters of studies need to be designed to study interactions among: (a) student development, (b) teacher development, (c) the development of relevant learning communities, and (d) the development of productive materials, tools, and programs for instruction and assessment. In the next section, a concrete example is given of a multitiered design experiment that studied three such levels of complex systems.

An Example of a Multitiered Design Experiment: Case Studies for Kids

During the last decades leading into the twenty-first century, tectonic-like shifts have occurred in traditional disciplines and new hyphenated fields of study have emerged, creating a demand for new kinds of abilities and expertise. The kinds of employees who are most sought-after in job interviews tend to be those who are proficient at mathematizing (or making sense of) complex systems, those who are capable of adapting rapidly to continually changing conceptual tools, and those who are able to work productively and communicate effectively in diverse teams of specialists working in complex multi-stage projects (Lesh et al., 2006). Furthermore, the kinds of mathematical conceptualization and communication abilities that are needed often involve more visualization abilities and multimedia representational fluency and less traditional computation.

At Purdue University’s Center for Twenty-first Century Conceptual Tools (TCCT), multitiered design experiments have been used to investigate such questions as: What is the precise nature of these new types of problem-solving situations? What new understandings and abilities are needed, even in traditional topic areas such as calculus, geometry, or statistics? How do these understandings develop? What can be done to facilitate their development? This section describes three levels of one multitiered design experiment that involved the development and implementation of Case Studies for
Kids, a set of problem-solving activities for middle-school students that are simulations of real-life situations where mathematical thinking is useful beyond school in a technology-based age of information (Lesh et al., 2002). For each tier, the reasons are described for which it was found useful to adopt design experiment methodologies and how the design process resulted in multiple revisions and testing.

**Tier One: Developing the Knowledge and Abilities of Researchers and/or Developers**

A major goal of the TCCT Center was to investigate the nature of the most important understandings and abilities in mathematics, science, language, and literacy that are most significant in order to provide foundations for success beyond school in a technology-based age of information (Lesh et al., 2003). In other words, the TCCT Center investigated how the 3Rs (Reading, wRiting, and aRithmetic) may need to be reconceptualized to meet the demands of the new millennium.

Early research on such questions typically involved either interviewing experts or observing problem-solvers in real-life problem-solving, or decision-making, situations (Lesh et al., 2006), often with the aid of video-tapes. One shortcoming of such research has been that choices about *where* to observe (in grocery stores? engineering firms? internet cafés?), *who* to observe (farmers? cooks? shoppers? computer programmers? baseball fans?), *when* to observe (only when they are calculating with numbers?), and *what* to count as mathematical or scientific thinking (on video-tapes or naturalistic observations) inevitably exposed preconceived notions about what it means to think mathematically or scientifically and about the nature of real-life situations in which mathematical or scientific thinking is useful.

A second way to enlist the opinions of experts has been more pragmatic and seems to focus on development more than research. For example, in 1989, the *Curriculum and Evaluation Standards for School Mathematics*, published by the National Council of Teachers of Mathematics, ignited a “decade of the standards” in which professional and public organizations in virtually every curricular area generated similar statements of instructional goals at the local, state, or federal levels. Most of these standards documents were developed by school people and academics, and the overriding concerns focused on making incremental improvements in the traditional curricula. Parents, policy-makers, business leaders, and other community leaders seldom were involved; such questions as what is needed for success beyond school seldom were treated as priorities. When it came to implementation, many schools formed blue-ribbon panels of local teachers to convert national standards to local versions that would not be viewed as top-down impositions and to write new test items or performance assessment activities that were aligned with these “local” standards. More often than not, school districts merely adopted existing standardized tests, which claimed to be aligned with their standards, and applied pressure on teachers to teach to these tests by promoting accountability.

Research conducted at the TCCT Center acknowledged that the knowledge of experts is crucial if the 3Rs are to be reconceptualized. But how can this knowledge be used in ways that challenge preconceived notions of mathematical thinking and real-life situations and do not lead to top-down attempts to reform curricula, resulting in standardized testing of superficial abilities? The authors chose to adopt a design experiment approach, enlisting teams of evolving experts (e.g., teachers, parents, policy-makers, mathematics and science education researchers, and professors in future-oriented fields that are heavy users of mathematics, science, and technology) as co-investigators and
co-designers of Case Studies for Kids (Lesh & Doerr, 2002). Each of the evolving experts is recognized as having important views that should be considered. Yet, the methodology also recognizes that different experts often hold conflicting views; none has exclusive insights about the truth, and all of them tend to evolve significantly when they go through a series of iterative, testing-and-revising cycles that mold and shape the final conclusions that are reached.

The evolving experts work in teams of three to five people to express their current ways of thinking about what is needed for success beyond school in forms that must be tested and revised repeatedly based on formative feedback and consensus-building (Lesh, 2002). They are given certain design specifications that provide “ends-in-view” for the kinds of tasks they were challenged to produce. These design specifications require the teams to consider: (a) the nature of real-life, problem-solving situations in which mathematical thinking is needed for success, (b) the nature of deeper and higher-order understandings and abilities that contribute to success in such situations, and (c) how to identify a broader range of students whose extraordinary abilities and achievements are apparent in real-life, problem-solving situations, even though their performance has been unimpressive in past situations involving traditional textbooks, tests, and teaching (Lesh & Lamon, 1994). Tasks that fail to be useful for these purposes are rejected or revised, no matter how appealing they may seem in other ways.

Thus, in TCCT’s ongoing evolving expert studies, the design processes that are used are similar to those that applied scientists use to design bridges, automobiles, ecosystems, and other complex systems. This approach recognizes that respecting the views of teachers, parents, professors, researchers, and other experts does not mean that any of their views should be accepted passively. Furthermore, it encourages participants to go beyond preconceived ways of thinking without presupposing that the nature of the improvements is known in advance.

Tier Two: Developing the Knowledge and Abilities of Teachers

In the book, Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching (Lesh & Doerr, 2003), many of the research studies that are reported investigated the nature of teachers’ developing knowledge and abilities (Doerr & Lesh, 2003; Koellner-Clark & Lesh, 2003; Schorr & Koellner-Clark, 2003; Schorr & Lesh, 2003) as well as interactions between teachers’ development and students’ development. Most of these teacher-level studies used research methods that satisfy the criteria described in the first sections of this chapter. In other words, they could have been called teacher-level design experiments because teachers repeatedly expressed their current ways of thinking in the form of complex conceptual tools that they designed, tested, and revised many times in order to meet specific teaching needs, and the tools were designed to be shareable (with others) and reusable (in a variety of situations).

Various concerns led to the previous investigations. First, many of the authors in the models and modeling book (Lesh & Doerr, 2003) have contributed significantly to the large and impressive research base that has developed about the nature of students’ developing mathematical knowledge and abilities in such topic areas as early algebra, rational numbers, and proportional reasoning. Although a great deal is known about the development of students’ understandings and abilities in these content areas, relatively little is known about the nature of teacher-level knowledge and abilities in them. For example, it certainly is clear that teacher-level understandings of ratios and proportions should involve more than student-level understandings. But what is the nature of
these additional, deeper, or higher-order understandings? In addition to understanding how relevant concepts develop logically (e.g., how can the concepts be defined logically?), teachers also probably need to understand how they develop psychologically (e.g., what representational systems contribute to their meanings in the minds of typical children?), historically (e.g., what issues led to their development in the history of science?), instructionally (e.g., what prerequisites exist in alternative curricular materials that are available?), and pragmatically (e.g., for what purposes, and in conjunction with what other ideas, are the ideas useful in real-life, problem-solving situations?). Also, what else should be involved in the understandings and abilities of outstanding teachers? How do these understandings and abilities develop? What kinds of experiences promote their development? Further, how can deeper and higher-order achievements be assessed or documented?

A second type of concern that led to the models and modeling book’s teacher-level studies was related to the fact that it does very little good for evolving experts to provide curricular materials such as the Case Studies for Kids described in the previous section unless teachers are able to recognize and make use of the thinking that their students exhibit. For example, even if student-level tasks are designed so that students reveal a broad range of mathematical understandings and abilities, there is no guarantee that teachers will recognize these achievements. Teachers who are skillful at guiding students toward teachers’ ways of thinking are not necessarily skillful at recognizing students’ ways of thinking. Teachers who can do things right do not do the right things at the right time, necessarily. Clearly, teaching expertise entails seeing at least as much as doing. But what is the nature of productive ways of thinking in situations such as: (a) analyzing complex pieces of work that students produce and giving them feedback about their strengths, weaknesses, and ways to improve, (b) observing processes and roles that are productive as students work on complex projects and giving them feedback on ways to improve their future performance, or (c) generating a concise summary of students’ alternative ways of thinking as they are reflected in oral reports describing the results of their work?

A third concern that influenced the teacher-level research in the models and modeling book (Lesh & Doerr, 2003) focused on the assessment of teachers’ knowledge and abilities. Teaching expertise does not appear to be reducible to a checklist of skills, and such checklists do not appear to improve significantly by adding lists of vague dispositions and performances. Nonetheless, many attempts to assess teaching expertise have supplemented what can be documented using short-answer tests by requiring teachers to produce portfolios that contain artifacts that are intended to document complex understandings and abilities. Many of these portfolios are similar to personal diaries, even though it is well known that teachers who talk articulately are not necessarily the same as those who produce results (Schorr & Lesh, 2003). Consequently, video-tapes often are used to show actual teaching episodes. However, almost anybody is able to perform capably in a few situations. Therefore, specific video-taped episodes seldom yield persuasive evidence about general capabilities, and doing something right in one situation, is quite different from doing the right things in a variety of situations.

As a result, in the teacher-level research that is reported in the models and modeling book, teachers’ ways of thinking were investigated as they worked collaboratively to design shareable, reusable, thought-revealing tools for implementing Case Studies for Kids tasks in the classroom. This design process often generated auditable trails of documentation that revealed important characteristics of teachers’ developing knowledge and abilities. In these cases, the design specifications required teachers to use formative feedback and consensus-building to make judgments about directions for
development that are increasingly “better” without feeling that they must converge toward a naïve conception of “best.” One of the most effective ways that researchers have found to promote teachers’ development is to help them become wiser about the nature of their students’ ways of thinking about the major themes in the topic areas they are trying to teach (e.g., Carpenter et al., 1989). Furthermore, by focusing on the design of conceptual tools that deal with classroom decision-making issues, teachers’ everyday teaching experiences can be turned into learning experiences. Such design tasks can provide the basis for effective types of on-the-job, classroom-based, teacher development activities; byproducts of these learning activities also produce documentation that reveals the nature of the understandings and abilities that develop. In particular, these byproducts include theory-based and experience-tested prototypes for materials, programs, and procedures to emphasize the following goals:

- Early democratic access to powerful conceptual tools (constructs, conceptual systems, capability amplifiers) that enable all students to achieve extraordinary results in simulations of real-life, problem-solving situations that are typical of those in which mathematical and scientific thinking is needed (after their school years) in the twenty-first century.
- Experiences in which students (who may be either adults or children, in or out of school) both develop and document deeper and higher-order understandings related to complex achievements that seldom are assessed on brief, easy-to-score, standardized tests.
- Ways to identify a broader range of students whose exceptional abilities and achievements often have not been apparent in settings involving traditional tests, textbooks, and teaching by recognizing the importance of a broader array of the knowledge and abilities needed for success in future-oriented fields ranging from agriculture to business to engineering to medicine and to other fields that are becoming increasingly heavy users of mathematics, science, and technology.

**Tier Three: Developing the Knowledge and Abilities of Students**

Two of the biggest concerns that led to the use of design research methods in studying the development of students’ knowledge and abilities were: (a) how to elicit and document the complex kinds of thinking that occur when students work on complex, problem-solving activities that require them to develop powerful conceptual systems, not traditional procedural kinds of answers, and (b) how to ensure that the development of the conceptual systems being elicited from the students is not guided by a teacher’s or textbook’s conception of what is “right”?

Much of the authors’ research done on students’ thinking has been on students working in groups of three to four, on *Case Studies for Kids* activities, which were designed to address these questions (Lesh & Zawojewski, 2006). In a very real sense, students working on these activities are conducting their own design experiments in which they are designing a mathematical model to explain a meaningful situation by quantifying, dimensioning, coordinating, systematizing, or (in general) mathematizing objects, relations, operations, transformations, patterns, regularities, or other systemic characteristics of learning or problem-solving situations. Over the course of the problem-solving session, their mathematical model goes through multiple design cycles as the students notice more relevant information about their problem and re-organize their model.

Two principles for developing *Case Studies for Kids* ensure that students’ models go through multiple design cycles, without the teacher or the textbook guiding them
directly (Lesh et al., 2000). The self-evaluation principle requires that activities have built into them ways for students to assess realistically the quality of their own modes of thinking, without predetermining what their final solution should look like. For example, in the quilt problem (see Appendix), students know that they have finished it when their quilt templates produce a quilt built to scale where all the pieces fit together, but they are not told what form their solution should take. Consequently, students working on these activities are able to develop their ways of thinking into powerful constructs and conceptual systems without being told which path to take.

The model documentation principle demands that the activities provide realistic conditions under which students are required to express and document their thinking in forms that are reusable, shareable, and modifiable. Instead of statements asking students to explain their solution, which often are added onto many word problems as an afterthought, these activities provide a realistic “client” who asks for a general solution to a specific problem, where the process is the product. For example, in the quilt problem, students are asked to describe how to make the templates for the pieces of a quilt, not only for the quilt pictured, but also for quilts of any design. Therefore, their final answer is a documentation of their problem-solving process.

The results of the preceding kind of problem-solving experiences show that, when students work on a series of such problems over the course of a semester, the power and the range of usefulness of their underlying ways of thinking tend to increase significantly. This is because every time they design a new, thought-revealing tool, they are extending and revising the underlying ways of thinking that the tools embody. As a result, the development of the tools involves significant forms of learning, and, as learning occurs, the tools produce auditable trails of documentation that reveal important information about the constructs and conceptual systems that the students are developing. Hence, the activities contribute to both learning and assessment. Furthermore, a broader range of students working on these tasks naturally emerges as having extraordinary potential (Zawojewski et al., in press). Many of these students whose abilities were unrecognized previously come from populations that are highly under-represented in fields that emphasize mathematics, science, and technology.

Investigating the Interacting Development of the Three Tiers Involving Students, Teachers, Researchers, Developers, and Others

One advantage of using the kind of thought-revealing design activities that are described in the preceding examples is that thought-revealing activities for students often provide ideal contexts for equally thought-revealing activities for teachers, researchers, curricula developers, and other kinds of evolving experts. For example, in Purdue University’s TCCT Center, emphasis has been placed on multitiered design experiments in which three interacting levels (and three interacting types) of problem-solvers are engaged in semester-long sequences of experiences in which: (a) students may design shareable, reusable, conceptual tools to meet the needs of clients who are described in Case Studies for Kids that focus on understandings and abilities that are typical of those that may be needed for success beyond school in a technology-based age of information, (b) teachers may design shareable, reusable, conceptual tools for assisting, analyzing, and assessing students’ work in Case Studies for Kids, and (c) researchers, developers, and other evolving experts may develop shareable, reusable, conceptual tools for assisting, analyzing, and assessing both teachers’ and students’ work in Case Studies for Kids. In other words, students develop models for making sense of mathematical, problem-solving situations; teachers develop models for making sense of students’ models and modeling
activities; and researchers develop models for making sense of both teachers’ and students’ models and modeling activities. Therefore, it is possible for researchers to investigate the interacting development of students, teachers, curricular materials, and programs.

Principles for Preventing the Typical Design Flaws that Occur in Multitiered Design Experiments

This chapter concludes by addressing a question of research quality: When the subjects being investigated are so complex, how can one ensure that both the designs and the underlying conceptual systems are being developed scientifically at each level of a multitiered design experiment? In answering this question, it is important to note that, in many ways, the goal of science is to go beyond common sense toward uncommon sense (based on inferred systems, patterns, and relationships beneath the surface of things). Therefore, design processes need to include ways to tease out and test foundation-level assumptions that seem like common sense. Failure to do so is often the source of serious design flaws. In particular, two of the most common design flaws occur in the following types of design experiments:

1. Practice-driven design experiments, in which the primary goal is to develop a complex artifact (or conceptual tool) that needs precise specific design specifications (situations and purposes) that exist apart from any given theory or theories. For example, in mathematics or science education, relevant artifacts or conceptual tools may involve software development, instructional materials development, assessment materials development, or the development of programs for learning, assessment, or problem-solving. In these cases, design flaws occur sometimes when the design processes involve little more than merely unprincipled tinkering; that is, where practical problems are solved, but no attempt is made to generate knowledge or information that is shareable or reusable beyond the immediate situation.

2. Theory-driven design experiments, in which the primary goal is to revise, refine, or extend a theory or way of thinking about teaching, learning, or problem-solving. In these circumstances, the most significant design flaws occur because the design processes never allow underlying ways of thinking to be tested (with the possibility of rejection) because they never consider seriously alternatives to the current ways of thinking. Theory-driven design can degenerate into little more than ideology-driven flights of fancy, where Karl Popper’s (1963) principle of falsifiability is never allowed to occur. According to Popper, the main difference between a scientific theory and an ideology is whether there are possibilities for current ways of thinking to be rejected or alternative ways of thinking to be adopted. In design experiments, there are two main ways in which the possibility of rejection is avoided. First, some studies are designed so that the only kinds of evidence and arguments that are allowed into consideration are those that the theory sanctions; tunnel vision often takes over so that basic ways of thinking are never really challenged. Second, some studies are designed so that the responses to failures never involve more than tinkering with (or endlessly embellishing) basic ways of thinking, rather than considering alternatives.

The authors have found it useful to consider four design principles in order to prevent these design flaws from occurring and to facilitate scientific development. These principles were developed so that they apply similarly to each level of a multitiered
design experiment: students, teachers, curricula developers, program developers, software developers, and other types of researchers, developers, or practitioners. When implemented, these principles ensure that, at each level, the design in question involves developing an underlying conceptual system that is testable, shareable, reusable, generalizable, and can add to the body of research knowledge in the field.

The Externalization Principle or Documentation Principle
Conceptual systems that are being investigated should be expressed in the form of artifacts or tools that can be examined by both researchers and other relevant participants. Thus, the task of designing such artifacts tends to be a thought-revealing activity (Lesh et al., 2000). In this sense, the artifact embodies the conceptual system(s) that were used to produce it because important components of the underlying conceptual systems are apparent in the artifacts themselves.

This is necessary because, in the process of designing complex artifacts and conceptual tools, participants often externalize their current ways of thinking in forms that reveal the constructs and conceptual systems that are employed. This is what is meant by saying that the products are embodiments of the relevant conceptual systems. This tends to be true especially if the products are conceptual technologies in the sense that they include not only procedures for doing something but also conceptual systems for describing and explaining the situations in which the artifacts or conceptual tools are intended to be useful. That is, the objectives for which these tools are developed often focus on interpretation, description, explanation, or sense-making more than on data processing (which presupposes that the data have been interpreted meaningfully already).

The Construct Assessment Principle
Design specifications (or goals) should be specified that provide criteria that can be used to test and revise trial artifacts and conceptual tools (as well as underlying ways of thinking) by weeding out products that are unacceptable or less useful than others.

The design specifications should function as Dewey-style “ends-in-view” (Zawokeyski et al., 2003). That is, they should provide criteria so that formative feedback and consensus-building can be used to refine thinking in ways that are progressively “better” based on judgments that can be made by the participants themselves. In particular, the “ends-in-view” should enable the participants to make their own judgments about the need to go beyond their first primitive ways of thinking and the relative strengths and weaknesses of the alternative ways of thinking that emerge during the design process. In addition to emphasizing power and usefulness in specified situations, productive goals also should require participants to develop constructs and conceptual systems that are: (a) powerful (to meet the needs of the client in the specific situation at hand), (b) shareable, (c) reusable, and (d) transportable. In other words, both the tools and the underlying ways of thinking should be shareable and generalizable.

The Multiple Design Cycle Principle or Knowledge Accumulation Principle
Design processes should be used in which the participants understand clearly that a series of iterative design cycles probably will be needed in order to produce results that
are sufficiently powerful and useful. Furthermore, they also should understand that any current state is always merely one in a continuing series.

If design processes involve a series of iterative development, testing, and revision cycles and if intermediate results are expressed in forms that can be examined by outside observers as well as by the participants themselves, then auditable trails of documentation are generated automatically. This documentation should reveal important characteristics of the developments that occur. That is to say, the design processes should contribute to learning and to the documentation and assessment of learning.

**The Diversity and Triangulation Principle**

Design processes should promote interactions among participants who have diverse perspectives; they also should involve iterative consensus building to ensure that the knowledge, tools, and artifacts generated will be shareable and reusable, and so that knowledge accumulates in ways that build iteratively on what was learned during past experiences and previous design cycles.

In general, to develop complex artifacts and tools, it is productive for participants to work in small groups consisting of three to five individuals who have diverse understandings, abilities, experiences, and agendas. By working in such groups, communities of relevant constructs tend to emerge in which the participants need to communicate their current ways of thinking in forms that are accessible to others. Once diverse ways of thinking emerge, selection processes should include not only feedback based on how the tools and artifacts work according to the objectives that were specified, but also according to feedback based on peer review. In this way, consensus-building processes involve triangulation that is based on multiple perspectives and interpretations. Thus, the collective constructs that develop are designed to be shareable among members of the group and in ways that enable knowledge to accumulate.

**Notes**

1. In mathematics, complexity theory deals mainly with systems in which the agents within the systems obey relatively simple condition-action rules; so it is the interactions among the agents that lead to complexity. In mathematics education, however, even the agents are (or involve) complex systems. So, compared with the kinds of complex systems that are emphasized in mathematics, those that occur in mathematics education are deeply complex. If several layers of agents are involved (much like letters within words, words within phrases, phrases within sentences, sentences within paragraphs, paragraphs within chapters, etc.), the agents at each layer have properties associated with complex systems.

2. In each of the examples, it is important to distinguish between the generalizations about the agents themselves and generalizations about the conceptual systems that the agents use. For example, generalizations about students themselves (e.g., this student is a gifted student, or a concrete operational thinker, or a creative thinker) are quite different from generalizations about the nature and development of students’ concepts and conceptual systems. Whereas cognitive science focuses on generalizations about the nature of learners or learning, mathematics and science educators tend to focus on generalizations about the nature of ideas and problem-solving situations, in what might be called experimental genetic epistemology (Piaget, 1970).

**References**


