Handbook of Design Research Methods in Education
Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching
Anthony E. Kelly, Richard A. Lesh, John Y. Baek

The “Compleat” Design Experiment

Publication details
James Middleton, Stephen Gorard, Chris Taylor, Brenda Bannan-Ritland
Published online on: 19 Jun 2008


PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Part 2

Design Research and its Argumentative Grammar
2 The “Compleat” Design Experiment
From Soup to Nuts

James Middleton
Arizona State University

Stephen Gorard
University of York

Chris Taylor
Cardiff University

Brenda Bannan-Ritland
George Mason University

Introduction

In this chapter, we articulate a methodology for design experiments and situate it within the larger cycle of design in educational settings, beginning with the initial conceptualization of design problems to the dissemination of educational products; hence, the impetus for the second part of our title: From Soup to Nuts. The term design is used by a variety of fields, ranging from art to engineering. In education, the type of research associated with the development of curricular products, intervention strategies, or software tools has been designated “design research” (e.g., Brown, 1992; Collins, 1992). In recent years, the number of investigators associating their work with this genre has become more prevalent in the literature and the focus of special issues in prominent journals about education, among them Educational Researcher (Kelly, 2003) and Journal of the Learning Sciences (Barab and Squire, 2004). This chapter contributes to the collective attempt of the field of education to clarify further methods of and perspectives on design research. Evidenced by the diversity of views incorporated in this book, ranging from product development to children as designers, it is clear that design research is a complex endeavor.

To begin our contribution to the conversation, we employ the theory of design from the engineering sciences as a useful analogy. Education research has been associated previously with an engineering process (as opposed to the sciences) by a number of prominent researchers (e.g., Cobb et al., 2003; Confrey, 2003). These research programs warrant revisiting the metaphors and analogies used to guide their work in design research. To a certain degree, a large portion of mainstream education research is engineering if what we mean by engineering is to design products and systems that solve problems of import to society.

This chapter describes the concept of design from an engineering perspective, touching on the distinction between design and artifact. The meat of the chapter (following the soup to nuts reference) presents a cyclic model of the overall design process and provides a detailed analysis of design experimentation and the role it assumes in the comprehensive cycle. Finally, we present an extended example of a program of research that conforms to our notion of design experimentation.
What is Design?

In the field of engineering, design is considered the essence of the profession. Engineers design. That is what they do. The emphasis of the word in this context is two-fold—relating to design as the practices of engineering, and designs, the intellectual products that those practices engender. As such, in the grammar that we are attempting to articulate, design functions as both a verb and a noun.

Design (the verb) is the creative process by which designer(s), considering a problem in their field of expertise, generate a hypothetical solution to that problem. Design (the noun) constitutes that general hypothetical solution, often embodied in material form (e.g., a blueprint or a physical model).

Design activity consists of a subtle but complex interaction between the designer and contextual constraints and is accomplished by proposing the form of an artifact, system, or process, which in turn drives its behavior, which in turn can be compared with its desired function. Because artifact, system, or process is cumbersome when repeated throughout a text, we will use henceforth the general term system to denote the ostensible outcome of a design except where a specific example is given. Both the form (proposed and at some point realized) and the desired function are hypothetical entities that are linked by a theoretical model (see Figure 2.1). In mechanical engineering, this model is called the function structure. The behavior of the system constitutes observable data. Therefore, the degree to which an artifact’s behavior embodies its function can be considered a partial index of both: (a) the quality of the design, and (b) the degree to which the form of the system produced or served as a conduit for the behavior (Shooter et al., 2000). The iterative, design-and-test cycle that engineers use is critical for the transformation of both the form of the system and of its (intended) function to conform to the pragmatic demands of utility and market.

If education research were merely the development of teaching tools and techniques, or the progressive empirical movement of students’ achievement toward the meeting of standards, there might be little or no emphasis on determining the adequacy and veracity of particular educational designs across contexts. The design cycle would consist of an entirely self-contained system, begun when a problem is encountered and finished when a product is refined to the extent that it solves a pragmatic problem and meets market demands. To a great extent, this approach epitomizes some product design cycles and traditional instructional systems design perspectives, at least in their implementation (Dick & Carey, 1990). Alternatively, many product development processes do incorporate significant testing throughout a systematic process of design and development (Urlich & Eppinger, 2000), and instructional design processes do support

![Figure 2.1 General Model of Design Research.](image-url)
conducting evaluation of learner preferences, behaviors, organizational results, and return on investment in designed training or education systems (Kirkpatrick, 1998). However, in many cases, the focus of the design process is primarily on the development of an instructional innovation (e.g., software or an interactive multimedia system), and the design is rarely evaluated beyond learner preferences and surface-level reactions to the innovation (Tessmer & Wedman, 1995).

More current perspectives on the intersection of design research and instructional systems design attempt to extend the traditional systematic approach to emphasize further learning processes, data collection, and diffusion processes across a program of research that can inform the adequacy, veracity, and viability of the instructional strategies and the theoretical rationale embodied by a design (see Bannan-Ritland, 2003, for example). The distinction between the traditional instructional design process, product development, and design research, then, lies in the overt emphasis on programmatic generation, development, and testing of a theory of learning, instruction, or human factors as it relates specifically to design research (cf., Shavelson et al., 2003). This shift in emphasis is even more crucial given the perspective that much of the research related to instructional systems design as well as educational research in general has been criticized for the subjects’ limited exposure to treatment materials, a lack of longitudinal analyses, and inadequate theory-building (Reeves, 2000).

For design to constitute design research generally, or a design experiment in particular, it must conform to standards of scholarship usually recognized by the scientific and educational communities as necessary conditions for programmatic study (Shavelson & Towne, 2002). Among these necessary conditions are the provision of a coherent and an explicit chain of reasoning, the attempt to yield findings that replicate and transport across studies, and the disclosure of methods and analyses for critique by the broader community (Gorard, 2002a). Although the structure of a complex instructional tool may be considered an embodiment of a local theory, unless that structure is made explicit and the propositional framework upon which the design rests laid bare, it does not constitute a test of that theory and therefore contributes little to the broader body of disciplined knowledge about teaching, learning, or human factors.

This is not to say that local theories are neither important nor transformative in a broader sense. In the iterative process of design and test, a design can be transformed to enact a chosen function better, but it also can change the way in which the initial problem is perceived, or it can generate as yet unseen problems such that new functions become possible. For example, the invention of Bakelite (phenol formaldehyde) in the early twentieth century was initiated by the need to coat bowling alleys with a hard, lacquer-like surface. Once the initial problem was solved, the designed properties of Bakelite afforded transportation to other problems of the day such as the mass production of billiard balls (before the invention of Bakelite, billiard balls had been manufactured by carving ivory on a lathe).

In part, due to its transportability across critical problems of the late industrial age, Bakelite became a common material for the production of everyday household products such as containers, picture frames, and heat-resistant handles for cookware. Theoretically, it stimulated the most pervasive material revolution of the twentieth century: the development of complex polymers such as polyethylene and nylon-plastics. We use the word transportation here as opposed to generalization or transfer to refer to the efficacy that a design provides beyond its initial conditions of development. We consider generalization to be a probabilistic argument (i.e., generalizability) as opposed to transportation as a fit between design and function. On the other hand, transfer is
a psychological construct relating to the degree to which something learned in one context can be performed in a different context. Transportation, as we are defining it, relates to the physical or applicational movement of a thing, a design, to a new applicational context (even if the details of the design have to be altered somewhat to fit the parameters of the new context).2

In education, also, there have been advances that have afforded widespread adoption and adaptation, leading to a similar revolution. The invention of the intelligence test and associated aptitude and achievement instruments has generated both new specialties in statistics (e.g., item response theory and multidimensional scaling) and intricate procedures for the adaptation of test theory to the assessment of attitudes, beliefs, and other indirect survey methods (e.g., the Likert scale and multiple-choice formats). It also has sparked a new political landscape focused on accountability in education—in essence, changing the very goals of education by changing the markers by which society gauges educational outcomes.3

Both of these examples illustrate the fact that a design is not an ostensible product. Rather, each of these designs can be thought of as the theory that specifies the parameters of an ostensible product and explicates the necessary and sufficient conditions under which a product (if it embodies the design) can be implemented successfully—in other words, its form, function, and behavior (Horowitz & Maimon, 1997).

For a design to become an actual thing (e.g., a structure, material, or program), it must go through cycles of modeling and testing, which may alter the original conceptual entity based on empirical evidence of how each successive enacted design performs under conditions of use. This distinction between a design and a product is crucial to our argument. To synthesize our notion of design with our earlier discussion of product design and instructional systems design, we see the development of an educational product or a software program as a substage in a larger set of coordinated theoretical activity. Imagine how a psychologist might create a series of tasks that uncover ways in which a student thinks about missing addend problems in arithmetic to test a model of how children process the semantics of story problems (see Carpenter & Moser, 1982). Similarly, an instructional designer might create a series of tasks or set of tools that, if enacted, would test a model of how children move through a hypothetical terrain that embodies the complexities of story problems. In both cases, the design of the tasks embodies a larger theory of cognition or situated learning. The tasks themselves (and the software or multimedia environment in which they are embedded) are tools that both uncover important information that helps build the theory (for instance, finding that children solve missing addend problems through counting up from the first addend before counting up from the larger) and that tests the theory (i.e., showing where the model of the development of arithmetic knowledge breaks down or where it is downright wrong). Lastly, the tasks and the environments embodying the theory eventually become enacted as plausible solutions to problems of education.

As in engineering, education design can be thought of as this process of generating plausible solutions to problems of teaching and learning that can be turned into ostensible products whose form enables particular behaviors, which in turn can be compared to the desired function (see Figure 2.1). However, just as in engineering, there are sources of variation that make the larger process of design problematic in moving from initial concept to product with appropriate fidelity and adaptation. These sources of variation have immense impact on the coherence of the methods of design research and on the subsequent claims that can be made about the theoretical implications of a data set.
Design Theory and Sources of Variation

All current theories of design assume a temporal process flowing roughly from conceptualization to realization. In this process, the flow of information moves downstream through various channels (i.e., departments, people with divergent expertise, design team members, different interested parties, target audience members, etc.) as a product is refined continually for market. However, design theories differ in the number of concurrent channels through which the design activity may flow and in whether information may flow back uphill in larger cycles. We synthesize these ideas later in our description of the Compleat Design Cycle into a general descriptive model that may serve the education community as first principles.

A point of fact: information flow between and among designers varies both upstream and downstream in the design process, as do materials and local conditions. As will be seen later on, a number of the conditions that have a potentially fatal impact on the enactment of a design are out of the control of the designer and, instead, are contingent upon the political and situational features of a potential application, including the decisions made by the technician building the designed product. Other conditions are under the control of the designer, but they can have a fatal impact if the design process involves a large team, or if the flow of information downstream from the conceptual designers to the manufacturers is inconsistent or prevented. In a similar manner, if the conditions under which the manufacture of the design is to be accomplished are not known by the upstream personnel, the parameterization of the design may not be realizable. For example, many technology-based design products are constrained severely by the designer’s limited understanding of the capabilities of the software creation tools and the programmer’s lack of knowledge of learning theory, thereby limiting potential interaction between these individuals and the possible flow of information both upstream and downstream. Under concurrent engineering methods (Ullman, 1992), whereby designers in all areas of product development work simultaneously, the distributed activity, by its very nature, is untenable without clear channels of information flow. The National Academy of Engineering (2004: 7) promotes an elegant definition of engineering (and other design fields by association) as being “design under constraint.”

Related to the idea of conducting design under constraints, one of the newest innovations in design theory is the idea of generating a range of satisfactory design parameters in upstream phases of design activity. By specifying the tolerances of deviation in function, a conceptual designer provides a wider range of possible local adaptations to the downstream designers (Kalsi et al., 2001). These approaches have revolutionized manufacturing by enabling local producers to use cheaper, more readily available materials or procedures that get the job done (e.g., a satisficing threshold), as opposed to conforming to overly rigid standards that might push the cost of production beyond a reasonable limit. In addition, they have fostered the design of products with diverse functions that use interoperable components (Dahmus et al., 2001).

In education, our analogous situation might be the development of modular curriculum materials (e.g., Connected Mathematics, Lappan et al., 2002; Mathematics in Context, Romberg, 2000). The designs of these curricula are flexible, allowing school districts and teachers to modify the order in which booklets are used and to choose, from among alternatives, the tasks and tools that enable them to accomplish the desired function of the curriculum. This analogy breaks down in the extent that the curriculum designers both understood and made explicit at the outset the range of parameters that could be altered and still ensure the fidelity of the product. Moreover, the channels of
communication of the design parameters from upstream (the curriculum developers) to downstream (the teachers and students) seem to be applied differentially across implementation sites as the market for these materials gets broader and reliance on publishing houses to support the professional development of teachers grows stronger. It is still not standard practice for teachers and school staff to be factored into product design from the outset (for alternative depictions of interactive social science, see Gibbons, 2000; Stokes, 1997).

What Makes Design Rigorous?

In education, it is often easier to make changes in learning or instruction than to gather information about what made those changes happen. This is the crucial difference between the design aspects and the scientific aspects of design research. Syllogistically, if design in education is about engineering particular forms of learning and instruction and if experimentation is about generating knowledge about those forms and subjecting that knowledge to empirical verification, then the theoretical model underlying particular designs must be made explicit and programmatic for any program of study to be termed a design experiment. Moreover, methods of model testing and revision must be generated that do test the comprehensiveness, accuracy, and utility of the theoretical model. These additional requirements are fundamental and also promote an approach to educational innovation that is coherent, logical, and disciplined on the one hand, but also pragmatic. In the classical view of design, products are seen to evolve through a continuous comparison of the design state and the desired function. This is possible only when the parameters of the function are capable of being specified beforehand (Ullman, 1992). In relatively simple, robust implementations of education research, this kind of parameterization may be possible.

The other, more modern, view of design tackles parameterization differently. Indeed, it has to primarily because the problems it attempts to solve involve too many variables to specify beforehand. Instead, a design problem is seen as the successive development and application of constraints that narrow down the subset of all possible product designs until only one remains (Ullman, 1992). This emergent solution should not be taken in any sense for an ultimate answer to the problem it attempts to solve. Despite these efforts, even the finest of designed architectural objects may fail ultimately (even the greatest bridges collapse eventually), and, at some point in the process, a decision will need to be made based on cost or other external constraints. After all, it is a fair question for policy-makers and interested parties to ask whether the enactment of an innovation is worth the material and human costs given the relatively small effect sizes (say, less than 0.25 standard deviation units) generally reported in education studies. However, should these considerations endure, the winning design holds a “good enoughness” in its behavior relevant to its desired function to justify its implementation and holds a cost relative to its benefit that makes implementation tenable. It is likely that most education problems are of this type—constrained by pragmatic and fiscal considerations in addition to those of scientific merit (Shavelson & Towne, 2002).

This is not to say that the classical design sequence is neither useful nor important in education research; it is and, in fact, is often used to develop solutions to small subsets of a larger design project. However, the economic and societal necessity for continuous improvement in education dictates that researchers and reformers engage in the design of tools, environments, and systems without knowing beforehand either: (a) all of the relevant parameters that impact their eventual success, or (b) the universe of potential designs from which their final design will emerge. It is likely that, in some instances, the
reformer may not be completely clear what the real problem is for which he or she is designing solutions. Cobb et al. (2003) refer to this as the “humble” nature of design experiments.

In attempting to articulate the design process as it applies to education research and development, we adopt this latter view. Our model is not prescriptive because our collective wisdom is limited to our own areas of scientific expertise, and we do not purport to know (let alone understand) many of the potential problems of education. As both researchers and designers, we present our model as descriptive, articulating the larger process and intersections of research and design of which design experiments, or design studies, are but a part. The role that experimentation plays in the design cycle ensures that the rigor and disciplined nature of the empirical facets of design, testing, and theory building are built in and not considered merely in the summative portion of project evaluation. In this manner also, design research aligns with the engineering (manufacturing) concept of continuous improvement (Bisgaard, 1997), which makes heavy use of iterative design cycles to generate pragmatic solutions with attention to time urgency. Our model may be useful for and applicable to the development and marketing of educational innovations by others who share our concern for the impact and rigor of scientific work.

The Compleat Design Cycle

The “classic” research model or cyclic research sequence promoted by advocates of scientific research, in education typically comprises four phases (see Figure 2.2). The first phase often (although not always) establishes the research problem or pedagogic issue that is to be addressed in the design of a particular artifact or intervention. In scientific terms, the aim is to produce a researchable hypothesis or question and a grounded theoretical model from which to develop an artifact that can be tested. This is commonly based upon a systematic approach, involving a review of the relevant literature and knowledge base, drawing upon existing theoretical developments and perhaps a desk-based analysis of existing data sets (both quantitative and qualitative). As the extent of our knowledge and understanding increases, much greater attention is being paid to the importance of these systematic approaches in identifying the pedagogic issue. Therefore, the review of existing literature has to become more systematic, and the availability of relatively large-scale secondary data sets means that a greater preliminary understanding of the problem can be ascertained before the research hypothesis or problem is finalized. However, this is not to ignore other, more nonsystematic inputs into this phase of the scientific research sequence, such as our own personal biographies, experiences, ideologies, and perhaps even serendipity. These types of inputs into developing the research hypothesis and a grounded theoretical framework are equally valuable and important to the initial phases (which might be the primordial “soup” from which clarity emerges).

Once a clear and researchable hypothesis has been established and the relevant grounded models identified, the second phase—to develop a testable artifact or intervention—can begin. This is often the most creative phase of the research sequence, where pedagogic experience can be of immense value. For scientific inquiry to continue, the resulting artifact and/or design of instructional sequences have to be testable against some predetermined criteria, such as pupil assessment, speed of delivery, pupil satisfaction, etc. Whatever educational criteria are chosen, the third phase of the scientific research sequence is to undertake a trial or intervention that will be able to isolate the systematic effects and begin to identify associated causal relationships. For a definitive
test, this trial would need an intervention and a control group, and, ideally, the groups or individuals would be allocated randomly. Such a randomized control trial would produce results indicating whether the artifact was more effective than an alternative or existing pedagogic practice. In some experimental research designs, it also may be possible to identify relevant contextual variables in this phase. These could provide useful information for the next phase of the research process.

The last phase of the classic scientific research model would be to take the results of the definitive trial and disseminate the findings to the rest of the teaching and learning community. This would show how effectively the artifact worked, providing practitioners and/or policy-makers with the necessary information to decide whether to implement the new system. However, this phase of the research sequence also should ensure that the artifact is transportable to other contexts or situations. It also should ensure that the new knowledge and understanding generated throughout the research sequence can contribute to further theoretical developments. So not only is the aim of the last phase to implement a “successful” system or pedagogic instrument more widely, but also it should be to develop a greater understanding of the original research problem or pedagogic issue, such that when the first phase of the research sequence resumes, the grounded theoretical models can be advanced.

In clinical medicine, as well as in education, the randomized controlled trial (RCT) is established as the best way of identifying the relative impact of alternative interventions
on predetermined outcomes (Shadish et al., 2002). The salience of this research design is largely because of the random allocation of participants to the alternative treatments, such that any difference in outcomes between the groups is due either to chance, which can be quantified, or to the difference in treatment. The RCT is applied most easily to the measurement of the efficacy of simple, well-defined interventions, such as a defined course of drug treatment, when delivered in an ideal research setting. Such four-phase studies are clearly useful in a range of other fields also and, perhaps especially, education (Torgerson & Torgerson, 2001), but they can be almost completely atheoretical in nature. For simple intervention studies, where we are concerned only with what works and not why, this lack of theory is not a problem, but it does limit the transportability of the results. Without an explanatory principle, it is not clear to what extent a successful intervention would be effective in a different context or for a different educational setting. And without in-depth data drawn from the same study, it is not clear what can be learned from an unsuccessful intervention (other than that it does not work).

But, by themselves, descriptive approaches to research, drawn from ethnographic and narrative genres, can provide considerable and rich detail about the processes of learning and can generate ideas about the improvement of learning. Yet, essentially passive approaches such as these cannot answer the probabilistic question of whether a suggested improvement works. For this, we have traditionally used the four-phase model (Figure 2.2).

The United Kingdom’s Medical Research Council (Medical Research Council, 2000) suggests that for more complex health education interventions, trials are most likely to be successful if they are based on sound theoretical concepts and involve both qualitative observation and quantitative testing (Campbell et al., 2000). Although many good quality RCTs have been undertaken in medicine and elsewhere, many of them generally have evaluated simple, almost naïve interventions, delivered in homeopathic doses, and inevitably have produced disappointing results. They have been less successful in identifying drug treatment interactions and other more complex clinical questions. On the other hand, several, high-quality, complex interventions generally have not been evaluated rigorously, and their effectiveness has not been demonstrated unequivocally.

Traditionally, trials have required that the interventions being tested are standardized and delivered uniformly to all participants. However, because educational interventions are so dependent on the quality of delivery, the value of trials predicated on “ideal” conditions can be limited. For example, some education interventions have been found to work well in efficacy trials when delivered by enthusiastic teachers with ample curriculum time. Yet, when implemented in practice, they have not been found to be effective, and the researchers have not known why necessarily (Nutbeam et al., 1993). Therefore, it is better to take a pragmatic approach, with the intervention delivered in the trial in a lifelike way. This approach sacrifices standardization for realism and means that the natural variability in delivery that occurs between practitioners must be recorded and monitored by in-depth means, perhaps by video-recording, as well as by more traditional outcome measures. This is not to imply that video-recording and analysis are simple tasks; in fact, educational researchers have called attention to the inherent problems and difficulties of this particular form of theoretical data collection (Hall, 2000).

However, video data have the potential to complement and inform RCTs. In summary, the “trial design ensures that an unbiased estimate of the average effect of the intervention is obtained, while the qualitative research provides useful further information on the external factors that support or attenuate this effect” (Moore, 2002: 5).

There is no doubt that it is easier to conduct RCTs of simple interventions. However, there is little value in compromising the likely effectiveness of the intervention by...
simplifying it merely to make it more amenable to evaluation in a trial. Indeed, RCTs are expensive both in monetary terms and, more particularly, in terms of their demands on research subjects and researchers. Hence, it is morally dubious to conduct a fully fledged trial until one is confident that the intervention is likely to be effective. Therefore, before conducting a RCT to demonstrate an intervention’s effectiveness, three earlier phases of investigation should have been completed. In effect, these three additional phases constitute what we term the *design experiment* (Figure 2.3). Meanwhile, the first two phases of the “what works?” model of educational research in essence remain the same (i.e., identifying the research problem and designing a testable solution).

The first new phase would involve the initial design of the intervention based on current theoretical understanding, ensuring that the intervention was grounded in theory and an explicit interpretation of the causal mechanism that it intended to promulgate. Furthermore, the transition between the second phase (the design of the artifact or intervention) and the (new) third phase (the feasibility study) would involve primarily qualitative methods in the formative evaluation of the intervention, using interviews, focus groups, observation, and case studies to identify how the intervention is working, barriers and facilitators to its implementation, and how it might be improved. Moving from the design to the feasibility stages might draw heavily on design and research processes in other fields such as product design and market research (Urlich & Eppinger, 2000). These more “explanatory” routes of inquiry complement powerfully the earlier use of secondary data analysis in identifying the research problem (Phase One).

In the third (new) phase (or the beginning of our design experiment), the intervention should be sufficiently well developed to be tested in a feasibility study, where it can be implemented in full and tested for acceptability by both the providers (health professionals, teachers, etc.) and the target audience (patients, pupils, etc.). The feasibility study is also an opportunity to test trial procedures, such as the definition of the alternative treatment, which may be the usual care, the control, or an alternative intervention, and to pilot and test outcome measures. It also may be used to provide a tentative estimate of the intervention effect, which then can be used to plan the size of the (main) (or definitive?) trial. The results of the feasibility study (Phase Three) will help to decide

---

**Figure 2.3** Three “New” Phases in the Design Experiment—Phases 3 to 5 of the Compleat Research Process.
whether the intervention should proceed to the next phase (teaching experiments) or whether it is necessary to return to the initial phases of identifying the research problem and developing the theoretical framework on which the intervention is based originally. Given the pragmatic and fiscal constraints of all scientific research discussed earlier, the results of the feasibility study may suggest that the entire research process should end.

The following fourth phase (prototyping and trials) begins a process of iteration between the testing and further modification of the intervention. Parallel to this is the potential to iterate the process between the laboratory (or other controlled environments) and the classroom (or real-life environments) (Brown, 1992). These iterative processes continue into the fifth phase (field study). However, Phase Four is characterized by piloting small-scale, multiple prototypes of the intervention. As the iterations between testing and further design become more sophisticated and the iterations between laboratory and classroom settings become more robust, advances are made in the intervention’s propositional framework and in outlining its plausible causal models. It is at this point that the research sequence enters the fifth phase (the field study), where it is implemented in full and tested for acceptability by both the providers (health professionals, teachers, etc.) and the target audience (patients, pupils, etc.). The field study is also an opportunity to test trial procedures, such as the definition of the alternative approaches, which may be the usual care, traditional instruction, the control, or an alternative intervention, and to pilot and test outcome measures. It, too, may be used to provide a tentative estimate of the intervention effect, which then can be used to plan the size of the main trial. The iterative process may continue, but the design of the instructional sequences becomes stronger, leading eventually to a robust model that aids the implementation of the artifact in many contexts. It is at this point that the documentation and recording of the process for implementing the artifact should be more systematic, because this develops the parameters for further transportability.

Once a series of “satisficing” objectives has been met, the compleat research process is ready to move on to the sixth phase, the definitive trial or evaluation. This phase is no different from the third phase of the original scientific model of educational research outlined earlier. However, the addition of three new phases to the research sequence ensures that the artifact or design to be tested should be effective. It also means that there is now a design of instructional sequences to aid the artifact’s transportability. Figure 2.4 shows the compleat research sequence with all seven phases combined. These additions to the simple scientific model of educational research are illustrated now in a real example of a mathematics design experiment.

An Extended Example From Mathematics Education

In this example, we analyze a current project, organized around the tenets we have established for design research, which has progressed far enough through the above phases to be a practical example. Note that, as a case, it specifies many of the necessary conditions for design experiments to be considered worthy contributions to the larger field of teaching and learning research. The case does not embody all possible, or probably all sufficient, conditions to be applicable generally to any study labeled design experiment. A broader survey of the current state of the art in the field is necessary to provide that background and that is beyond the scope of this chapter. As our case progresses through subsequent phases of the design cycle, we first provide the theoretical and methodological considerations that the researchers encountered and then follow them up with specific examples from practice.
Design Experiments: Phases Three through Six

For design to be iterative, progressive, and disciplined, a method of empirical analysis of the intended function → form → behavior proposition must exist and drive both the smaller iterative design cycles and the larger test of the overall theory under investigation. Usually, when a proposition (e.g., a hypothesis) is put to an empirical test, the general qualities of that portion of a line of inquiry is termed an experiment. Most experiments involve actual manipulation of the conditions that impact the parameters of the proposition, such that evidence is created that would either: (a) falsify the proposition by showing that the given conditions do not influence subsequent actions in the manner that the theory would suggest, or (b) provide confirmatory evidence that the proposition is functional for a given set of conditions. As stated earlier, design experiments, by which we mean the kind envisioned by Brown (1992), Collins (1992), and articulated more fully by Cobb et al. (2003), are a subprocess in the larger cycle of design. In particular, we see them as the articulation of Phases Three through Six in the model we present.

To illustrate the role of design experiments in this larger context, we describe a program of study embarked upon by one of the principal authors of this manuscript. In brief, this ongoing program attempts to create a theory of children’s development of understanding of quotients, within the larger structure of rational numbers and arithmetic operations (see the Rational Number Project publications, e.g. Behr et al., 1992;
Moreover, pertaining to broader sociocultural theory, this research attempts to generate methods to describe specifically how individual cognition contributes to the collective understanding of groups of individuals in instructional settings (de Silva-Lamberg & Middleton, 2002). We describe here the considerations for the feasibility of this program of research (Phase Three); the prototyping of tasks, trials, and the development of an initial plausible model of development through intensive individual teaching experiments (Phase Four); and the refinement of that model to be sufficiently explanatory so that it could be used as a hypothetical structure for the design and testing of instructional sequences (Phase Five).

Phase Three: Establishing the Feasibility of the Research Program

Assuming that there is some theoretical justification for a design to be created, an additional consideration for educational scholars is whether the social capital of their work warrants the perturbation of students’ educational experiences; namely, the relatively intrusive process of interviewing, video-taping, and more formal testing that might be interwoven over an extended period of time in the conduct of a study. Also, the nesting of any individual study within the larger program of research, particularly if the theory or model under examination is longitudinal in nature, must be justified. In the Children’s Understanding of Quotients project (CHUQ) (de Silva, 2001; Middleton et al., 2001; Toluk, 1999; Toluk & Middleton, 2001), we examined the curricular offerings in mathematics for middle-grade students and found that there was no systematic treatment of fractions as quotients, nor any integration of the instruction on division as an arithmetic operation, with the theory of fractions as rational numbers. Fractions were treated almost exclusively as part—whole quantities, and division was treated almost exclusively (but particularly so in the early portion of the middle grades) as resulting in whole numbers with remainder or decimal quantities. This established a curricular warrant, because the ability to understand fractions as numbers, or that quotients represent a division relationship (i.e., a multiplicative inverse), is critical for proportional reasoning and further understanding of algebraic fractions in general and rational expressions in particular—both central concepts of elementary algebra.

However, the negotiation of eight weeks of intense work with individual children and another five weeks with teachers and their classrooms for this portion of the program required the investigators to negotiate time, prerequisite experiences, and the involvement of the teacher and to establish a reasonably firm assurance of the benefits of participation in such a demanding set of procedures. This meant that the researchers had to develop a close relationship with the teacher, school, and school district, with the promise of long-term collaboration (i.e., beyond the study at hand and even beyond the overall program of research; see Middleton et al., 2002 for a description of the relationship of the author with this particular system).

The initial strategy of starting with individual interviews and building to the classroom study had two advantages. Theoretically, it allowed the researchers to articulate their model more precisely, providing initial examples and vignettes to help the classroom teacher understand the model of development they had created (see Figure 2.3). Pragmatically, it provided a means of withdrawing from the project for the teacher, school, and school district should the model prove untenable or inarticulate in the initial phases with individual children.

Lastly, the researchers had to articulate the initial theoretical justification for the study in terms that were both understandable and rigorous to social scientists (to pass the Human Subjects Institutional Review) and versed in the language and values of the
teacher, school, and school district. For proper informed consent to be made, there needed to be acceptance, conceptually and politically, by all the interested parties. That justification took the following form:

- Instruction on fractions consists almost entirely of part—whole treatment. Instruction on division consists almost entirely of whole numbers with remainder treatment. At no time in the elementary or middle grades are these two strands of instruction connected explicitly and their conceptual unity articulated.

We know this because we have analyzed both the larger national research literature and the particular curriculum and materials your teachers use.

- Having a connected understanding of fractions as division is critical for the success of your children.

We know this because so much of their future algebra experiences will use this knowledge. Again, we have analyzed both the national data and the curricular experiences of your children.

- Together (researchers, teacher, school, and school district), we can generate a model of the development of children’s understanding of quotients, use that model to create instructional sequences to teach in the real conditions of schools, and test and refine those sequences so that they can be transported across classrooms to maximize the benefit for all your students.

Here, the researchers negotiated the roles of the researchers, students, teacher, school, and school district, and articulated the individual study → classroom study structure.

This justification is a proposition—a hypothesis—stated in the form that design researchers often use: articulating a real problem that exists, a method of solution to that problem, and engaging the implementing community (i.e., the downstream designers and technicians) in the design process.

Phase Four: Prototyping, Trials, and Multiple Interventions

Following the establishment of the warrant, the articulation of the hypothetical structure to be investigated is critical for a design experiment to be truly an experiment. In particular, the researcher must pay attention to not only what is happening (i.e., the description), but also why it is happening (i.e., the causal model), and how it is happening (i.e., the cause-and-effect mechanism) (Gorard, 2002b; Shadish et al., 2002; Shavelson et al., 2003). In most design research, including our own (lest we point fingers), scholars have focused primarily on the descriptive portion of the data. The narrative of a child confronting fractions as division is compelling and useful as a case that can be compared and contrasted to other cases and used in professional development to illustrate certain processes of development. However, without the causal model articulating why the child is thinking the way he or she does, given the prompt and their particular learning history, the practitioner has little efficient information about how to impact thinking in ways that are mathematically sound. Moreover, without an articulated mechanism to effect appropriate learning (i.e., teaching methods, tools, and sequences), a practitioner’s subsequent course of action is left to a “best guess” kind of
strategy, as opposed to coherent and consistent designed experiences. This need under-
scores the requirement for iterative cycles of test and revision in the design process at
the prototyping phases to provide both the full complexity of a compelling case and
the theoretical model by which the case can be understood as a case of something
(Shulman, 1986).

The late Ann Brown (1992) presented her own view on design experimentation as
moving progressively from working with children in laboratory settings to more com-
plex and naturalistic interventions. Her characterization of design experiments does not
distinguish between the power of the laboratory or the classroom as settings for the
 generation of theoretical knowledge. Instead, she promotes each as providing a dif-
ferent lens on how children learn and on what systems and strategies can be envisioned
to promote quality learning. In particular, her presentation evokes a dual-method
approach whereby laboratory research assists in the building of detailed models of
learning, upon which instructional models can be based. The laboratory also can pro-
vide details of individual thinking and learning that the larger context of the classroom
cannot (see also McCandliss et al., 2003). The classroom, for its part, provides the
practical setting and the complexity that the laboratory cannot emulate properly. Any
compleat design experiment must take the instructional setting and epistemological
basis for the study into account when designing the methods of product design, data
collection, and analysis.

As presented in Phase Three, the researchers developed a general strategy for coming
to understand children’s thinking about quotients. In Phase Four, the details of design-
ing a method for researching children’s thinking and creating an initial plausible model
were undertaken. The researchers could have begun with prototypical problems in
small-group or classroom settings, or they could have used larger data sets and survey
instrumentation for the empirical data upon which a plausible model of development
was generated. Because the project is rooted in a psychological tradition and builds
upon a body of research that makes heavy use of teaching experiment methods, a
modification of microgenetic interview techniques (see Steffe & Thompson, 2000), it
was deemed appropriate to build from that methodological base. This decision is a
critical one, we think. As stated earlier, no design is any ultimate answer to the problem
it attempts to solve, but the epistemological basis upon which a design is built will
determine its form and function. In this case, the interest of the researchers was to
understand children’s thinking at the individual level and then build classroom
sequences upon that basis. Others would be more interested in the nature of classroom
discourse related to quotients and build their theory and instructional sequences on that
basis. It is unlikely that both approaches would yield the same solution design. The
coherence of theory to methodology, therefore, is of fundamental importance in the
evaluation of design experiments and is critical to explicate for any future scholar or
practitioner who attempts to replicate or implement the findings of a design study.

Following this example, in the CHUQ project, the researchers conducted four paral-
lel individual teaching experiments that lasted approximately 16 sessions per child (two
per week for eight weeks). The purpose of this initial research was to study fifth graders’
conceptualizations of the quotient under instructional conditions that expressly
required them to confront and connect the isomorphisms inherent in thinking about
fractions and in thinking about division. Specifically, the experimenters focused on the
transitional understandings that children construct as they move from the division of
whole numbers to the depiction of fractions as the quotient of division situations.
Baseline interviews were conducted with standard problems relating to fractions
and division to give the experimenter information on the initial understandings and
skills. From this information, the experimenter designed initial problems with two basic strategies in mind:

1. To provide isomorphic problems in close contiguity, one that promotes *fractional* interpretations, the other that promotes *whole number division* interpretations. A sample problem is shown in Figure 2.5(a).

2. To bridge the two conceptual strands by presenting fair sharing problems with partitionable remainders, then focus on the remainders. See Figure 2.5(b) for an example of this strategy.

By presenting problems that were isomorphic, but that had potentially fractional versus divisional interpretations, the authors attempted to confront the children with the notion that the process of partitioning quantities was the same conceptually. By focusing on the remainder in a division problem, the experimenters attempted to confront the children with division and fractions in the same problem—that is, that fractions were the result of division.

As the children constructed new schemes as new information was provided, the experimenter would devise additional problems to test whether children had developed

![Figure 2.5](image-url)

*Figure 2.5 Standard Fractions and Division Problems Given During Baseline Interviews.*
a different understanding and, if so, provide a description of that understanding. At the end of the data collection portion of the study, the authors had four parallel cases. The commonalities among the cases were analyzed and incorporated in the development of a first model—a plausible model—of children’s development of understanding of the quotient construct. At this point, the model was primarily descriptive, tracing the new schemes that the children had constructed to deal with successively more difficult depictions of fractions, division, and their confluence. As the research program moved to the larger scale design and implementation study in a classroom, it served as a causal model, addressing the trajectory along which the children were assumed to develop (e.g., a hypothetical learning trajectory; Cobb et al., 1997), and positing the key transitional conceptualizations the children needed to develop and positing the kinds of instructional strategies that would facilitate those transitions. At this point, it became both design specifications for a sequence of instruction and a hypothesis that could be tested empirically under the naturalistic conditions of classroom life.

Phase Five: Modification and Testing

As empirical evidence for a design grows, so does the complexity of the problem(s) it can address. Initial prototypes are likely to be flawed, both in their conceptualization and in their implementation. Coordinating the theoretical conceptualization and implementation (i.e., contextual) parameters as a design moves downstream in the process is by no means an easy task. For example, in the CHUQ project, a plausible model of the development of children’s understanding of quotient—telegraphic, simplistic, narrow in applicability—needed to be transformed into a sequence of instructional tasks that would move a whole class of children to a higher level of understanding. Where does one begin?

The CHUQ researchers began by using the story problems, tasks, and sequences of conversations—data they had gleaned from the significant amount of transcriptions generated in the teaching experiments. Many of their first attempts were inadequate at best, but the advantage of an iterative longitudinal design is that one gets better as one goes along (this actually came as a surprise). The project was influenced heavily by the classroom design studies described by Cobb et al. (1997), where they shared in some detail the theory of classroom teaching and learning as emergent phenomena and subsequently designed a program of research both to teach place value and to research students’ collective understandings (see also Cobb et al., 2003). Moreover, as the project moved from a model of individual development to an implementation situated in a real classroom, it had to expand its research base to include instructional theories. In our case, because it had a consistent epistemological worldview with the emergent perspective, the model of anchored instruction was employed to structure the initial curricular sequence of tasks and to situate classroom discourse in a common colloquial understanding: fair sharing through planning a party. It was assumed that as children encountered successively more formal and conceptually difficult notions of fractions as quotients, the party context would enable the teacher to identify and challenge their informal understanding better and also would serve as a source of problems for group discourse.

A lesson or sequence of instruction is a window into the long-term goals of a project. It is not merely a product, or an episode, but an enacted hypothesis about the nature of children’s understanding and how that plays out across an important mathematical or scientific concept. As an ostensible product, the instructional unit developed by CHUQ consisted of a workbook with sets of story problems, tasks, and thought questions.
that built on the original research into children’s thinking. It was structured to build conceptual linkages between children’s understandings of fractions and their understandings of division in the ways that the plausible model derived from the empirical evidence would predict. The unit began with a complex video anchor (e.g., Cognition and Technology Group at Vanderbilt, 1992) that exposed children to issues of unit and partitioning (e.g., a case of soda in the United States consists of 24 cans; a case then is partitionable into 24 (1) cans, 4 (6) cans, 2 (12) cans, etc.). In the process of planning a party for a large group of children, the notions of division, fractions, and remainders are encountered inevitably, and these notions are dependent upon the concepts of unit and partitioning. So, to summarize, the content of the unit: (a) embodied critical theoretical considerations from the research on rational numbers in general, (b) structured them based on the specific empirical evidence generated in the teaching experiments conducted earlier in the research program, (c) situated the sequences in a format that embodied the theory of teaching exemplified by anchored instruction, and (d) embedded the perspective on knowledge development in classroom settings in the emergent perspective from a sociocultural tradition. What then, of all of these things, do you test?

We admit that one cannot test all of these design considerations in a single study. For the purposes that the CHUQ researchers set out to accomplish with the teacher, school, and school district, the immediate test had to address the adequacy and veracity of the plausible model. Not only were these the simplest to test conceptually, but also they were the only considerations that could be tested given that they were the only real hypotheses of the lot. The first, third, and fourth considerations were assumptions upon which the instructional unit was based and, as such, were a backdrop through which the plausible model could be situated and tested under conditions of use. The researchers could have used a different set of assumptions to test the plausible model, and they may have seen different results. However, because all of the considerations were developed so that they held common assumptions about what constitutes knowledge, how teachers and students interact in the classroom discourse, and the roles of tasks and tools in facilitating knowledge, they represented one of a finite number of coherent configurations of perspectives at appropriate levels of detail and focus (Cobb et al., 2003). It is not that research uses a configuration of perspectives that is at issue from critics of the state of the art in design experimentation, but the mutuality of those perspectives and whether they each contribute at the appropriate level and scope to a coherent epistemological argument (Shavelson et al., 2003).

Following the initial drafting of the instructional sequence, the researchers (including the classroom teacher) developed conjectures—mini hypotheses—about how children would approach the tasks presented and what kinds of tools the teacher would use to build upon their reactions. They did this for the entire sequence, knowing full well it was likely that a large portion of the tasks would change either in order or in form as the study progressed. When the classroom instruction commenced, the researchers recorded whole-class conversations through a digital video camera, followed target students in a small group through their dialogue and individual seatwork with a second camera, and recorded all instances of all 24 children’s scribblings, journal entries, physical models, and other inscriptions with digital photographs for the five full weeks of the study. All the digital data had a coordinated time stamp so that the exact moment when a conjecture, drawing, or other exteriorized piece of evidence of thinking was generated, could be placed at its proper moment in the overall discourse.

The results of this study indicated that the practices of the classroom moved more or less along the hypothetical trajectory established by the plausible model and instantiated...
through the instructional unit. However, there were some fundamental differences in how the classroom as a collective progressed as opposed to the individual children in the study. In particular, the need to establish norms for representing fractions and division was more critical in the whole-class application. Also, the children in the classroom study had more difficulty seeing nested sets of units in fractions and had been socialized into seeing “improper” fractions as just that—improper. A huge watershed occurred when students saw improper fractions as division and as numbers greater than one without converting to the whole number plus a fraction less than one notation. Lastly, although children in whole and small groups appeared to understand and be able to deal with the concept of division as a number (e.g., a quotient as a fraction), many were not able to do so on their own. Much of these differences can be attributed to the difference between the conditions of individual teaching experiments and whole class instruction, but some (particularly the last) cannot. All indications are that the original model developed from studies of individual children is generally sound and is a useful structure with which to organize sequences of instruction. However, the ways in which the researchers originally characterized children’s understanding of division as number has to be reconceptualized and retested.

Additional benefits of situating studies of learning within a larger design experiment include the ability to generate new knowledge about several aspects of learning and instruction simultaneously. For example, although the use of coordinated time stamps on digital video and still photographs allowed the project to trace individual development within the larger complexity of the classroom, it also shed light on the ways in which inscriptions are used as media for communication and, more precisely, generate models of how information is propagated and how individuals contribute to a collective knowledge structure (Lamberg & Middleton, 2003). These data are helping the researchers articulate the emergent perspective on classroom discourse more carefully and are helping them generate a disciplined method for modeling classroom discourse from both the individual and the collective lenses simultaneously.

Phases Six and Seven: What Next?

If we go back to the beginning of this chapter and examine the general model of design research depicted in Figure 2.1, we see an endless cycle of theory-building, designing, and testing. The CHUQ project, as described here, exemplifies this academic side of design research. It has generated both a useful and verifiable theoretical model, developed that model into an embodiment with the form of an instructional sequence, and tested that model against conditions of use. At its current state, it is, in the lexicon of the mathematical community, an existence proof. We know that children can think about quotients in these ways under these conditions, and that, given the expertise of the researchers, the theoretical model can be transported from the individual interview to the classroom setting. At some point, however, the project has to disseminate some ostensible products that will be of use beyond the relatively small scale of the design test bed. Here is where issues of scope and scale loom eminently. We envision two avenues for definitive trials of the theory: first, by examining the qualities of students’ understanding of quotients following instruction using the instructional sequence or analogous materials developed; second, by examining the impact, if any, that such instruction has on the larger domain of rational numbers that currently constitutes the bulk of the US mathematics curricula in the middle grades.7

The first class of definitive trials would involve a nomothetic approach, using randomized or carefully stratified samples, with relatively large numbers of students.
Controlling for variation in the consistency of instruction and students’ prior experiences, the researchers will be able to determine the relative impact on students’ learning statistically through computation of effect sizes (reference) and repeated measures analyses. Pragmatically, this kind of trial gives an estimation of the immediate impact of an innovation in curriculum and instruction. However, not all impact is assessable in the short term. By nature, education is both cumulative and transformative (cognitively, both assimilative and accommodative). The consistency of experiences over several years of mathematics instruction also will determine the quality of students’ learning. These considerations predicate that other conceptualizations of rational numbers be considered simultaneously with that of the quotient.

The second class of definitive trials, assessing the impact of quotient understanding on the larger domain of rational numbers, is more difficult to assess. It may be that, viewed as a whole, focusing on one important subset of the rational numbers domain is insufficient for any long-term change to occur (e.g., any gains found in a short-term study are nullified by the overall impact of the curriculum that is taught for several years). Alternatively, it may be that quotients (fractions as division)—because they have both conceptual utility and operational applicability across the other conceptualizations of rational numbers—have a catalytic effect on long-term fluency. We do not know. Without some type of evaluative study in the form of a set of clinical trials or other nomothetic technique, we may never know.

Eventually, the unit designed for research may make it to market (Phase Seven). It is unlikely, however, that it will have much of an impact, in terms of scope and scale, because it is not attached to any large set of published curricular materials. In the literature on the diffusion of innovations, a determination of whether a clever design will become widely adopted by users is its observability; that is, the degree to which the results of the innovation are visible to potential users (Rogers, 1995). The curriculum adoption process in the United States is big business, with a small number of large publishing houses competing for huge profits. The development and marketing of coherent sets of materials and associated teaching tools virtually preclude the wide-scale adoption of our five-week, instructional sequence designed for 11-year-olds. Moreover, although the authors of this chapter strive to disseminate their work more broadly, the accepted venues for the communication of research in the university community—journal articles and other, expository-framed text—do not appeal to the masses of teachers we are hoping to reach.

Recently, however, the education research community has had a venue for wide-scale impact: working with publishing houses as partners for reform. With the assistance of the US National Science Foundation, several sets of curricular materials have been developed, based on (at the time) the latest research on learning, teaching, and technology. By and large, the authors of these curricula were researchers who had contributed to the body of work, but they certainly were not experts in all of the areas of cognition, instruction, and social psychology that they needed to incorporate into their materials development projects.

Nevertheless, the qualities of these curricula are fundamentally different from the materials that constituted the bulk of the published texts when they were published, and there is some large-scale evidence of their effectiveness under most conditions of schooling (Confrey & Stohl, 2004; Webb et al., 2001). In at least one of these projects much of the processes we are calling design research was employed with good measure (Romberg, 2000). Some projects are in the process of revision. Should the collective argument promoted by the authors of this chapter prove useful, we hope that our framework will serve these revision projects as principles for designs.
that will lead to more theoretically defensible products with even greater pragmatic value.

Conclusions
Throughout this chapter, we have emphasized the articulation of theory, method, and pragmatic considerations in the design cycle overall and in design experiments in particular. The examples we have provided illustrate both the conceptual and instrumental advantages of such an articulation, but also the tensions that arise when assumptions about theory, learning, policy, and social contexts are not coherent or do not reflect the conditions of the design process. These tensions are countermanded somewhat through the iterative structure of design whereby unanticipated difficulties can be dealt with expeditiously. However, this does not negate the imperative for careful consideration of epistemological coherence at the outset. In the cases we cited, success was predicated to a great extent upon the coherence of the theory and the attention to areas where the empirical evidence at different levels of cognitive or social complexity complemented each other and therefore could be merged into a comprehensive hypothesis about the impact of a design. This rigorous attention to complexity has not been a hallmark of education research in general.

Often, education problems are stated in terms of yes or no questions, that is: “Does a certain approach work or not?” The all-or-nothing nature of such questions yields outcomes of research, including theory, that take the form of particular prescriptions or nothing at all. A shift to research questions and associated methods that ask: “Why might a certain approach work?” (e.g., curricular sequences, social environments, human–computer interfaces, state-level policies) leads to the generation of potentially transportable models of teaching, learning, or policy that, in turn, facilitate the creation of products that embody these models (e.g., a shift to the creation of workable systems). Concurrently, attention to questions of “How a certain approach might work” can provide the theoretical basis for the mechanisms of innovation, their instantiation, and their adaptation. The use of both individual (i.e., laboratory) and group (i.e., field) studies allows the scholar to develop first plausible models of thinking and the kinds of tasks that facilitate that thinking, then to test those models subsequently to ascertain the mechanisms that make them transportable to analogous situations.

We offer the concept of transportation as an alternative to the traditional notions of generalizability and transfer. Transportation has many of the inductive features embodied by statistical and theoretical generalization, but it also carries the communicative and responsive features of analogical generalization (cf., Smaling, 2003). As such, it is complementary to both kinds of generalization but stands on its own, describing very different kinds of phenomena. In particular, transportation is concerned with the diffusion of innovation both broadly (i.e., scope) and situationally (i.e., value-added). Moreover, the term refers explicitly to designed systems as opposed to habits or mental constructs or population characteristics. After all, unlike these constructs, when a design is transported from one situation to another, like a good book, it can be shared, with corresponding benefits for all parties.

Concerning the diffusion of innovation, the issue of “research-to-practice” should not be problematic, if the research is practice. Even when the classroom teacher is not a member of the research team (as was the case in the CHUQ individual teaching experiments), the system of teacher/researcher to students’ interaction exhibited a remarkable self-similarity; that is, the ways in which problems were presented, students’ thinking
was recorded, and hypotheses were conjectured and tested were remarkably similar both in the individual teaching experiment conditions and in the whole-class conditions (but with added complexity). Lesh and Kelly’s (2000) description of multitiered teaching experiments has an analogous self-similarity, although their projects deal with the professional development of teachers related to the teachers’ own understanding of students’ thinking. Recent commentary on issues related to teachers’ professional development may provide the additional directions needed to address the critical importance of teachers’ learning and adoption of design research innovations (see Borko, 2004). A question for scholars who pursue design research in using laboratory and classroom settings might be: “To what extent do the differing conditions of the research project need to reflect a common methodology?” At present, this is unanswered, but our best guess is that the closer the methods are between the laboratory and the classroom, the more transportable the initial plausible model to be enacted will be.

This issue, then, changes the generalizability argument radically to one of scale and transportability; that is, “Are people able to take the key design aspects of the innovation and transport them to their own contexts in some useful fashion?” and “What aspects of the innovation are applicable to large numbers, where delivery, institutional context, and culture vary radically from those under which the innovation was designed?” It seems likely that these questions would require the development of design specifications that state in a clear way what the parameters are for transportability and scale and by what measures they can be assessed (Sloane & Gorard, 2003). In particular, we are intrigued by the discussion of the perceived relative advantage of an innovation provided by Rogers (1995), as well as the more current perspectives on this issue of Zaritsky et al. (2003). Some market research concerning the characteristics of educational innovations that are perceived as relative advantages over existing situations—with appropriate compatibility and complexity for successful adoption to occur—would be an excellent step in the direction of continuous educational improvement.

We have taken care to demonstrate that design experiments are valuable methodological additions to the standard procedures that already include randomized controlled trials and other traditional experimental studies, as well as descriptive studies from narrative and ethnographic traditions. The methods we describe assume a refreshingly uncomplicated combination of both qualitative and quantitative approaches that capitalize on their mutual strengths without sacrificing their mutual rigor. Each approach can be seen as necessary to the process of design, and neither can be seen as sufficient alone. Others have pointed out the complementary nature of methods and how they constitute lenses by which complex educational systems can be understood better (e.g., Behrens & Smith, 1996; Jaeger, 1997). We have attempted to add to this discussion by situating design research in this centrist position.

Perhaps the most important contribution we could provide is a set of references of good examples upon which to model design studies. After all, most of us learn by emulating others. The sources we have cited continue to inspire our own work, and each of those projects embodies high levels of rigor, openness to critique, and practicality. We think that the fears of the critics of design research would be assuaged somewhat if they examined the painstaking detail to which our examples have gone to ensure that their claims are warranted and that such claims have broader applicability to the education community.

The bulk of this chapter was devoted to the articulation of a model that would provide two contributions to the field by commenting on the theoretical considerations to be taken into account and by providing an extended example to illustrate the
complexities of doing so. The first contribution was to establish a model that researchers who are interested in design can use to help them plan and build research programs. We emphasize programs because the model is written at the program level and not at the project level. This highlights the issues of time and complexity (and money) that often keep design projects from leading to systematic and programmatic theory. Second, our model is provided as a first step in the establishment of standards by which design research can be evaluated. Different types of research and development can be mapped onto this model with fidelity, and the scholar outside the science and mathematics education fields, which represent the bulk of the research we have reviewed, can see his or her place in the larger genre of theory and methods.

Acknowledgments

The research reported in this chapter was supported in part by grants from the United States National Science Foundation, the United States Department of Education, and the United Kingdom’s Economic and Social Research Council. The opinions expressed are solely those of the authors and do not reflect the opinions of the United States National Science Foundation, the United States Department of Education, or the United Kingdom’s Economic and Social Research Council. All of the authors contributed equally to this chapter.

Notes

1 The archaic use of “compleat” in the title hearkens back to a classic text on angling by Izaak Walton, “The Compleat Angler, or the Contemplative Man’s Recreation,” originally published in 1653. Modern editions of the book are available.
2 See also the concept of transferability (Guba & Lincoln, 1989; Lincoln & Guba, 1985; Smaling, 2003), which is analogous to our use of transportation, but because the roots of the terms are identical, more easily confused with transfer.
3 The extent to which these consequences of design transportation are intended or even beneficial is beyond the scope of this chapter. Needless to say, because design research is, to a large extent, social engineering, the ethical and moral consequences of design creation and adoption must be considered critically.
4 Of course, the implicit assumption here is that there may be studies labeled design experiments that may not be design experiments according to this definition.
5 This approach to combining methods has been termed the new political arithmetic (Gorard, with Taylor, 2004).
6 We assume that the reader is familiar with the methods of establishing a theoretical warrant for the conduct of education research, so we do not treat it here. Suffice it to say that the authors did their literature review and found a significant gap in the knowledge base on rational number learning, right where the quotient should be located.
7 Although a number of mathematical strands are emphasized in both the standards and textbooks, rational numbers is a domain that transcends these boundaries, appearing in number, algebraic, geometric, and statistical contexts ubiquitously. It has been named a watershed domain in the field of mathematics (Kieren, 1976).

References


