



The 'Compleat' Design Experiment: from soup to nuts

James Middleton
Stephen Gorard
Chris Taylor
Brenda Bannan-Ritland

Department of Educational Studies
Research Paper 2006/05

THE UNIVERSITY of *York*



THE 'COMPLEAT' DESIGN EXPERIMENT: FROM SOUP TO NUTS

James Middleton

Stephen Gorard

Chris Taylor

Brenda Bannan-Ritland

© University of York

ISBN: 1 85342 633 4

Department of Educational Studies
University of York
York
YO10 5DD

INTRODUCTION

In this paper, 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 as “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 in education including *Educational Researcher*, 2003 32(1) and *Journal of Learning Sciences*, 2004, 13(1). This paper contributes to the collective attempt of the field of education to further clarify methods 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 endeavour.

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 previously associated with an engineering process (as opposed to the sciences) by a number of prominent researchers (e.g., Confrey, 2003; Cobb, Confrey, diSessa, Lehrer, & Schauble, 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 paper describes the concept of design from an engineering perspective, touching on the distinction between design and artifact. The meat of the paper (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 twofold--relating to design as the practices of engineering, and designs, the intellectual products that

those practices engender. As such, in the grammar 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 within their field of expertise, generate a hypothetical solution to that problem, which solution can be termed a design. Design (the noun) constitutes that general hypothetical solution, often embodied in material form (e.g., a blueprint or 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, henceforth we will use 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 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 the quality of the design, and the degree to which the form of the system in fact produced or served as a conduit for the behavior (Shooter, Keirouz, Szykman, & Fenves, 2000). The iterative design-and-test cycle engineers utilize 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 towards 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 (ISD) 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 (Ulrich and Eppinger, 2000) and instructional design processes do support conducting evaluation of learner preferences, behaviors, organizational results and return on investment of 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 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 further emphasize 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 of *programmatically generation, development and testing of a theory of learning, instruction, or human factors* as it specifically relates to design research (cf., Shavelson, Phillips, Town, & Feuer, 2003). This shift in emphasis is even more crucial given perspectives that much of the research related to instructional systems design as well as educational research in general has been criticized for 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 generally 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 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). While 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 better enact a chosen function, but it can also change the way 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 20th 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 (prior to 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 20th 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 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. Transfer on the other hand, 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).¹

In education also, there have been advances which 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, 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, multiple choice formats). It has also 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.²

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 successfully implemented—in other words, its form, function, and behavior (Horowitz & Maimon, 1997).

For a design to become an actual thing (e.g., structure, material, 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

¹ See also the concept of *transferability* (Lincoln & Guba, 1985; Guba & Lincoln, 1989; Smaling, 2003), which is analogous to our use of *transportation*, but because the roots of the terms are identical, more easily confused with *transfer*.

² 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.

our notion of design with our earlier discussion of product design and instructional systems design, we see the development of an educational product or software program as a substage in a larger set of coordinated theoretical activity. Imagine the way 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). An instructional designer, similarly, 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 i.e., finding that children solve missing addend problems through counting up from the first addend prior to 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 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 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 regarding 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, stakeholders, target audience members, etc.) as a product is continually refined for market. However, design theories differ in the number of concurrent channels through which the design activity may flow, and in whether or not 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 which 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 designers, but 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 upstream personnel, the parameterization of the design may not be realizable. For example, many technology-based design products are severely constrained 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 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) promotes an elegant definition of engineering [and other design fields by association] as "...about design under constraint (p.7)."

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, Hacker, & Lewis, 2001). These approaches have revolutionized manufacturing by

enabling local producers to utilize 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, Gonzalez-Zugasti, & Otto, 2000).

In education, our analogous situation might be the development of modular curriculum materials (e.g., *Connected Mathematics*, <http://www.math.msu.edu/cmp/index.html>; *Mathematics in Context*, <http://mic.britannica.com/mic/common/home.asp>). The design 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 insure fidelity of the product. Moreover, the channels of communication of design parameters from upstream (curriculum developers) to downstream (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 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 really 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.”³ Moreover, methods of model testing and revision must be generated that actually do test the *comprehensiveness, accuracy, and utility* of the theoretical model. These additional requirements are fundamental and also promote an approach to

³ Of course, the implicit assumption here is that there *maybe* studies labeled “design experiments” that may not, in fact, be design experiments according to this definition.

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 *a priori* (Ullman, 1992). In relatively simple, robust implementations of education research, this kind of parameterization may be possible⁴.

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 *a priori*. 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 in any sense be taken for an ultimate answer to the problem it attempts to solve. Despite these efforts, even the finest of designed architectural objects may ultimately fail (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. It is, after all, a fair question for policymakers and stakeholders to ask whether or not the enactment of an innovation is worth the material and human cost given the relatively small effect sizes generally reported in education studies. However should these considerations hold up, 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 (National Research Council, 2001).

This is not to say that the classical design sequence is neither useful nor important in education research—it is, and in fact, 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, in fact, that in some instances, the reformer may not even be completely clear what is the real problem for which they are 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, as our collective

⁴ This approach to combining methods has been termed the “new” political arithmetic (Gorard with Taylor, 2004).

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 insures that the rigor and disciplined nature of the empirical facets of design, testing, and theory building are built-in and not just considered 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 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 cyclical research sequence promoted by advocates of scientific research, in education typically comprises four phases (see Figure 2). The first phase often (though 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 some 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 datasets (both quantitative and qualitative). Increasingly greater attention is being paid to the importance of these systematic approaches in identifying the pedagogic issue as the extent of our knowledge and understanding increases. Therefore the review of existing literature has to become more systematic, and the availability of relatively large-scale secondary datasets 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 non-systematic inputs into this phase of the scientific research sequence, such as our own personal biographies, experiences, and ideologies and perhaps even serendipity. These types of inputs into developing the research hypothesis and some 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 enquiry 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 randomly allocated. Such a randomized control trial would produce results indicating whether the artifact was more effective than some alternative or existing pedagogic practice. In some experimental research designs it may also 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 effective the artifact worked, providing practitioners and/or policy-makers with the necessary information to decide whether to implement the new system or not. However, this phase of the research sequence should also ensure that the artifact is *transportable* to other contexts or situations. It should also 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, it should also aim 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, Cook, & Campbell, 2002). The salience of this research design is largely due to 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 due to the treatment difference. The RCT is most easily applied 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 also useful in a range of other fields 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 actually works. For this, we have traditionally used the four-phase model (figure 2).

The UK 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, Fitzpatrick, Haines et al., 2000). While many good quality RCTs have been undertaken in medicine and elsewhere, many of these have generally evaluated simple, almost naïve interventions, delivered in homeopathic doses, and have inevitably 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 have generally not been rigorously evaluated, and their effectiveness has not been unequivocally demonstrated.

Traditionally, trials have required that the interventions being tested are standardized and uniformly delivered to all participants. However, since 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 actual practice they have not been found to be effective, and the researchers have not necessarily known why (Nutbeam, Macaskill, Smith et al., 1993). It is, therefore, 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 is a simple task, 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 has 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, p. 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 merely to make it more amenable to evaluation within a trial. Indeed, RCTs are expensive both in monetary terms, and more particularly in terms of their demands on research subjects and researchers. It is therefore morally dubious to conduct a fully fledged trial until one is confident that the intervention is likely to be effective. Therefore, prior to the conduct of 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 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 may be improved. Moving from design to feasibility stages might draw heavily on design and research processes in other fields such as product design and market research (Ulrich & Eppinger, 2000). These more “explanatory” routes of enquiry powerfully complement the earlier use of secondary data analysis in identifying the research problem (phase 1)⁵.

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 to both 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 usual care, control, or some alternative intervention, and to pilot and test outcome measures. It may also be used to provide a tentative estimate of the intervention effect, which

can then be used to plan the size of the (main) [or definitive?] trial. The results of the feasibility study (phase 3) will help to decide whether the intervention should proceed to the next phase (teaching experiments) or whether it is necessary to return to the initial phases in identifying the research problem and developing the theoretical framework from which the intervention is originally based. 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 trialing) 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 to both 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 usual care, traditional instruction, control, or some alternative intervention, and to pilot and test outcome measures. It may also be used to provide a tentative estimate of the intervention effect, which can then be used to plan the size of the main trial. The iterative process may continue but the design of instructional sequences become stronger leading, eventually, to a robust model that aides 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, as this develops the parameters for further transportability.

Once a series of "satisficing" objectives have been met the complete research process is ready to move on to the sixth phase, the definitive trial or evaluation. This phase is no different to 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 aide the artifact's transportability. Figure 4 now shows the complete research sequence with all seven phases combined. These additions to the simple

scientific model of educational research are now illustrated 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, and 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, nor 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 paper. As our case progresses through subsequent phases of the design cycle, we first provide theoretical and methodological considerations the researchers encountered and then follow them up with specific examples from practice.

Design Experiments: Phases 3 – 6.

For design to be iterative, progressive, and disciplined, some 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. Generally speaking, when a proposition (e.g., a hypothesis) is put to some empirical test, the general qualities of that portion of a line of inquiry is termed an “experiment.” Most experiments involve actual manipulation of conditions that impact the parameters of the proposition such that evidence is created that would either 1) falsify the proposition by showing that given conditions do not influence subsequent actions in the manner that the theory would suggest; or 2) provide confirmatory evidence that the proposition is functional for a given set of conditions. As said earlier, design experiments, by which we mean the kind envisioned by Brown (1992) and Collins (1992), and articulated more fully by Cobb, Confrey, diSessa, Lehrer, & Schauble (2003) are a subprocess in the larger cycle of design. In particular, we see them as the articulation of phases 3 through 6 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 (de Silva,

2001; Middleton, de Silva, Toluk, & Mitchell, 2001; Toluk, 1999; Toluk & Middleton, 2001). 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, Harel, Post, & Lesh, 1992; Kieren, 1976; 1993). 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 within instructional settings (de Silva-Lamberg & Middleton, 2002). We describe here the considerations for the feasibility of this program of research (Phase 3); the prototyping of tasks, trialing, and the development of an initial plausible model of development through intensive individual teaching experiments (Phase 4), and the refinement of that model to be sufficiently explanatory that it could be used as a hypothetical structure for the design and testing of instructional sequences (Phase 5).

CHUQ Example: Phase 3. Establishing the feasibility of the research program.

Assuming that there is some theoretical justification for a design⁶ to be created an additional consideration for educational scholars is whether the social capital of their work warrants the perturbation of students' educational experiences, the relatively intrusive process of interviewing, videotaping, 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), 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-number with remainder or decimal quantities. This established a curricular warrant, as the ability to understand fractions as numbers, or that quotients represent a division relationship (i.e., 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.

⁶ We assume 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 numbers right where the quotient should be located.

However, the negotiation of eight weeks of intense work with individual children and another five weeks with classrooms for this portion of the program required the investigators to negotiate time, prerequisite experiences, involvement of the teacher, and establish a pretty firm assurance of the outcome benefits of participation in such an intense 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, Sawada, Judson, Bloom, & Turley, 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 with which to help the classroom teacher understand the model of development they had created (see Figure 3). Pragmatically, it provided an “out” for the teacher, school, and 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 Human Subjects Institutional Review), and versed in the language and values of the teacher, school, and district. For proper informed consent to be made, there needed to be a buy-in, conceptually and politically by all the stakeholders. That justification took the following form:

1. Instruction on fractions consists almost entirely of part-whole treatment. Instruction on division consists almost entirely of whole number with remainder treatment. At no time in the elementary or middle grades are these two strands of instruction explicitly connected and their conceptual unity articulated.
 - a. We know this because we have analyzed both the larger national research literature, and the particular curriculum and materials your teachers use;
2. Having a connected understanding of fractions-as-division is critical for future success of your children.
 - a. We know this because so much of their future algebra experiences will utilize this knowledge.
 - b. Again, we have analyzed both the national data and the curricular experiences of your children.
3. Together (researchers, teacher, school, and 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 benefit for all your students.

- a. Here the researchers negotiated the roles of researchers, students, teacher, school, and district, and articulated the individual study → classroom study structure. This justification is a proposition—an hypothesis—stated in the form that design researchers often do. Articulating a real problem that exists, a method of solution to that problem, and engaging the community of implementation (i.e., the *downstream* designers and technicians) in the design process.

CHUQ Example: Phase 4. Prototyping, trialing, multiple interventions.

Following the establishment of the warrant the articulation of the hypothetical structure to be investigated is critical for a “design experiment” to truly be an *experiment*. In particular, the researcher must pay attention to not only what is happening (i.e., description), but also why it is happening (i.e., causal model), and how it is happening (i.e., cause and effect mechanism) (Gorard 2002b, Shadish, Cook & Campbell, 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 which 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 as to how to impact thinking in ways that are mathematically sound. Moreover, without an articulated mechanism to effect appropriate learning (i.e., teaching methods, tools, sequences), a practitioner’s subsequent course of action is left up to a best guess kind of strategy as opposed to coherent and consistent designed experiences. This need underscores the need 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 take on design experimentation as moving progressively from working with children in laboratory settings to more complex 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 different 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 provide details of individual thinking and learning that the larger context of the classroom cannot (see also McCandliss, Kalchman & Bryant, 2003). The classroom, for its part, provides the practical setting and the complexity that the laboratory cannot properly emulate. Any complete 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 3, the researchers developed a general strategy for coming to understand children's thinking about quotients. In Phase 4, the details of designing 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. Since 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 parallel individual teaching experiments that lasted approximately 16 sessions per child (2 per week for 8 weeks). The purpose of this initial research was to study 5th 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 children construct

as they move from 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 of the interviewed students. From this information, the experimenter designed initial problems with two basic strategies in mind:

1. Provide isomorphic problems in close contiguity, one which promotes *fractional* interpretations, the other which promotes *whole number division* interpretations. For example:

Problem that hypothetically promotes division interpretations:

Suzy had **12** brownies she wants to share with her friends. She puts them on **4** plates so that there is the same amount of brownies on each plate. How much does each person get?

Problem that hypothetically promotes fraction interpretations:

Suzy had **4** brownies she wants to share with her friends. She puts them on **12** plates so that there is the same amount of brownies on each plate. How much does each person get?

2. Bridge the two conceptual strands by presenting fair sharing problems with partitionable remainders, then focus on the remainders. For example:

Problem with partitionable remainder:

Suzy had **9** brownies she wants to share with her friends. She puts them on **2** plates so that there is the same amount of brownies on each plate. How much does each person get?

Focusing on the remainder:

Suzy had **1** brownies she wants to share with her friends. She puts them on **2** plates so that there is the same amount of brownies on each plate. How much does each person get?

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— id est., that fractions were the result of division.

As children constructed new schemes as new information was provided, the experimenter would devise additional problems to test whether or not children had developed 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 4 parallel cases. The commonalities among 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 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 children were assumed to develop along (e.g., a hypothetical learning trajectory; Cobb et al., 1997), and positing the key transitional conceptualizations 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.

CHUQ Example: Phase 5. 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 utilizing the story problems, tasks, and sequences of conversations—data they 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 (it actually came as a surprise) one gets better as one goes along. The project was heavily influenced by the classroom design studies described by Cobb et al. (1997), where they share in some detail, the theory of classroom teaching and learning as emergent phenomena and subsequently designed a program of research to both 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 better identify and challenge their informal understanding, and would also 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 just a product, or an episode, but is an enacted hypothesis about the nature of children's understanding, and how that plays out across an important mathematical or science concept. As an ostensible product, the instructional unit developed by CHUQ consisted of a workbook with sets of story problems, tasks, and thought questions (see above) that built on the original research on 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 inevitably encountered, and these notions are dependent upon the concepts of unit and partitioning. So, to summarize, the content of the unit: 1) embodied critical theoretical considerations from the research on rational numbers in general, 2) structured these based on the specific empirical evidence generated in the teaching experiments conducted earlier in the research program, 3) situated the sequences in a format that embodied the theory of teaching exemplified by anchored instruction, and 4) 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 the CHUQ researchers set out to accomplish with the teacher, school, and district, the immediate test had to address the adequacy and veracity of the plausible model. Not only was this simplest to test conceptually, but it was the *only* consideration that could be tested given that it was the only real hypothesis 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 in a way that they held common assumptions about what constitutes knowledge, how teachers and students interact in the classroom discourse, and the role of tasks and tools in facilitating knowledge, they represented one of a finite number of *coherent* configurations of perspectives at appropriate level of detail and focus (Cobb et al., 2003). It is not that research utilizes 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—regarding how children would approach the tasks as 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 that it was likely that a large portion of the tasks would change either order or form as the study progressed. When the classroom instruction commenced, the researchers recorded whole-class conversations through 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 scribbles, journal entries, physical models and other inscriptions with digital photos for the 5 full weeks of the study. All 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, it 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 the way 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, though children in whole and small group 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 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, while the use of coordinated time-stamps on digital video and still photos 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 individual's 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 collective lenses simultaneously.

CHUQ Example: Phases 6 and 7: What next?

If we go back to the beginning of this paper and examine the general model of design research as depicted in Figure 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 testbed. Here is where issues of scope and scale loom eminently. We envision two avenues for definitive trialing of the theory: first by examining the qualities of students' understanding of quotients following instruction utilizing the developed instructional sequence or analogous materials; second by examining the

impact, if any, such instruction has on the larger domain of rational numbers that currently constitutes the bulk of the US mathematics curriculum 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 student learning statistically through computation of effect sizes (reference), and repeated measures analyses. Pragmatically, this kind of trial gives a ball-park estimation of the immediate impact of an innovation in curriculum and instruction. However, not all impact is assessable in the short term. Education by nature is both cumulative and transformative (cognitively both assimilative and accommodative). The consistency of experiences over several years of mathematics instruction will also determine the quality of student 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 number, is more difficult to assess. It may be that overall, 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 taught curriculum 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 affect on long-term fluency. We just do not know. Without some form 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 7). 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 curriculum materials. In the literature on the diffusion of innovations, a determination of whether or not a clever design will become widely adopted by users is its *observability*: 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 precludes the wide-scale adoption of our five-week instructional sequence designed for 11 year olds. Moreover, although the authors of this paper strive to disseminate

⁷ While a number of mathematical strands are emphasized in both standards and textbooks, rational numbers is a domain that transcends these boundaries, appearing in Number, Algebraic, Geometric, and Statistical contexts ubiquitously. It has been christened a watershed domain in the field of mathematics (Kieren, 1976).

their work more broadly, the accepted venues for communication of research in the university community—journal articles and other expository-framed text, frankly, does 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, a number of sets of curriculum materials have been developed, based on (at the time) the latest research on learning, teaching, and technology. The authors of these curricula, by and large, were researchers who had contributed to the body of work, but 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 than the materials that constituted the bulk of published texts when they were published, and there is some large-scale evidence of their effectiveness under most conditions of schooling (reference Confrey & Stohl, 2004; Webb, Romberg, & Shafer, 2001). In at least one of these projects much of the processes we are calling design research were employed with good measure (Romberg, 1994). Some projects are currently in the process of revision. Should the collective argument promoted by the authors of this paper prove useful, we would hope that our framework would serve these revision projects as principles for design that would lead to more theoretically defensible products with even greater pragmatic value.

Final Discussion

Throughout this paper, 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 actual conditions of the design process. These tensions are countermanded somewhat, through the iterative structure of design—whereby unanticipated difficulties can be dealt with just-in-time. However, this does not negate the imperative for careful consideration of epistemological coherence at the outset. In the cases we cited, success was to a great extent predicated 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 could therefore be merged into a comprehensive hypothesis regarding 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, *id est*, “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, facilitates 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 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, and subsequently test those models 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 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 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 student interaction exhibited a remarkable self-similarity—that is, the ways in which problems were presented, student thinking was recorded, and hypotheses conjectured and tested were remarkably similar both in the individual teaching experiment conditions and in the whole class conditions (but with added complexity). Lesh & Kelly’s (2000) description of multi-tiered teaching experiments has an analogous self-similarity, though their projects deal with professional development of teachers related to the teachers’ own understanding of student thinking. Recent commentary on issues related to teacher professional development may provide additional needed directions to address the critical importance of teacher learning and adoption of design research innovations (see Borko,

2004). A question for scholars who pursue design research in utilizing 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 will be the initial plausible model to be enacted. This issue then, changes the generalizability argument radically to be one of scale and transportability— *id est*, “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 would seem likely that these questions would require the development of design specifications that state in some clear way, what are the parameters for transportability and scale, and by what measures can they 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 more current perspectives on this issue of Zaritsky, Kelly, Flowers, Rogers, & O’Neill (2003). Some market research concerning the characteristics of educational innovations that are perceived as relative advantages over the status quo—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 toolbox that already contains 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 better understood (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 embody 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 insure that their claims are warranted, and that such claims have broader applicability to the education community.

The bulk of this paper was devoted to the articulation of a model that would provide two contributions to the field by commenting on theoretical considerations to be taken into account, and also 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.

NOTES

The research reported in this manuscript was supported, in part, by grants from the National Science Foundation, the United States Department of Education, and the UK Economic and Social Research Council. The opinions expressed are solely those of the authors and do not reflect the opinions of the National Science Foundation, the United States Department of Education, or the Economic and Social Research Council. Authors contributed equally to this piece and are listed, by convention, alphabetically.

REFERENCES

- Bannon-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher*, 32(1), 21-24.
- Behrens, J. T., & Smith, M. L. (1996). Data and data analysis. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 945-989). New York: Simon & Schuster Macmillan.
- Behr, M.J., Harel, G., Post, T.R., & Lesh, R. (1992). Rational number, ratio and proportion. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). New York: Macmillan.
- Bisgaard, S. (1997). *The role of scientific problem solving and statistics in quality improvement: Some perspectives. Report No. 158*. University of Wisconsin-Madison: Center for Quality and Productivity Improvement.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3-15.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2, 141-178.
- Campbell, M., Fitzpatrick, R., Haines, A., et al. (2000) Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, 321, 694-6.
- Carpenter, T. P., and Moser, J. M. (1982). *Addition and subtraction: A cognitive perspective*. Hillsdale, NJ: Lawrence Erlbaum.
- Cobb, P., Gravemeijer, K., Yackel, E., McClain, K., & Whitenack, J. (1997). Symbolizing and mathematizing: The emergence of chains of signification in one first-grade classroom. In D. Kirshner & J. A. Whitson (Eds.), *Situated cognition theory: Social, semiotic, and neurological perspectives* (pp. 151-233). Mahwah, NJ: Lawrence Erlbaum.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design in educational research. *Educational Researcher*. 10(1), 9-13.
- Cognition and Technology Group at Vanderbilt (1992). The Jasper experiment: An exploration of issues in learning and instructional design. *Educational Technology Research and Development*, 40(1), 65-80.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. New York: Springer-Verlag.
- Confrey, J. & Stohl, V. (2004). On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations. National Research Council Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials.

Confrey, J. (2003, April). Scientific research in mathematics education: Answering the challenge and challenging the answer. Paper presented at the Research Preession of the National Council of Teachers of Mathematics Annual Meeting, San Antonio, TX.

Dahmus, J. B., Gonzalez-Zugasti, J. P., & Otto, K. N., (2000, September). Modular Product Architecture. *Proceedings of the ASME International Design Engineering Technical Conferences*, Baltimore, MD.

De Silva, T. (2001). Quotient construct, inscriptional practices and instructional design. Unpublished doctoral dissertation, Arizona State University.

De Silva Lamberg, T & Middleton, J.A. (2002). Inscriptional practices as indices of emergent understanding of the quotient. A paper presented at the American Educational Research Annual Meeting, New Orleans, Louisiana.

Dick, W., & Cary, L. (1990). *The Systematic Design of Instruction, Third Edition*. New York, NY: Harper Collins.

Gibbons, M. (2000). Mode 2 society and the emergence of context-sensitive science. *Science and Public Policy*, 27(3), 159-163.

Gorard, S. (2002a). Fostering scepticism: The importance of warranting claims. *Evaluation and Research in Education*, 16, 3, 136-149.

Gorard, S. (2002b). The role of causal models in education as a social science. *Evaluation and Research in Education*, 16, 1, 51-65.

Gorard, S. with Taylor, C. (2004). *Combining methods in educational and social research*, London: Open University Press.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Beverly Hills, CA: Sage.

Hall, R. (2000). Videorecording as data. In A.E. Kelly & R. A. Lesh (Eds.) *Handbook of resarch design in mathematics and science education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Horowitz, R., & Maimon, O. (1997, September). Creative design methodology and the SIT method. *Proceedings of the ASME International Design Engineering Technical Conferences*, Sacramento, CA.

Jaeger, R. M. (Ed.). (1997). *Complementary methods for research in education*. Washington: American Educational Research Association.

Kalsi, M., Hacker, K., & Lewis, K. (2001) A comprehensive robust design approach for decision tradeoffs in complex systems design. *Journal of Mechanical Design*. 123(1), 1-10.

Kirkpatrick, D.L. (1998). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler Publishers, Inc.

- Kieren, T. E. (1976). On the mathematical, cognitive, and instructional foundations of rational numbers. In R. Lesh (Ed.), *Number and Measurement*, (pp. 101-144). Columbus, Oh: Ohio State University, EEIC, SMEAC.
- Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. In T. Carpenter, E. Fennema, & T. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 49-84). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lesh, R., Kelly, A., (2000). Multi-tiered teaching experiments, In A. Kelly, R. Lesh (Eds.) *Research in science and math education* (pp. 197-230). Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- McCandliss, B. D., Kalchman, M., & Bryant, P. (2003). Design experiments and laboratory approaches to learning: Steps toward collaborative exchange. *Educational Researcher*, 32(1), 14-16.
- Medical Research Council. (2000). *A framework for development and evaluation of RCTs for complex interventions to improve health*. London: MRC.
- Middleton, J. A., de Silva, T., Toluk, Z., & Mitchell, W. (2001). The Emergence of Quotient Understandings in a Fifth-Grade Classroom: A Classroom Teaching Experiment. In R.S. Speiser & C. Maher (Eds.), *Proceedings of the 28th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Snowbird, UT: ERIC.
- Middleton, J. A., Sawada, D., Judson, E., Bloom, I., & Turley, J. (2002). Relationships Build Reform: Developing Partnerships for Research in Teacher Education. In L. English (Ed.), *Handbook of International Research in Mathematics Education* (pp. 409-431). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moore, L. (2002). Lessons from using randomised trials in health promotion, *Building Research Capacity*, 1, 1, 4-5.
- Nutbeam, D., Macaskill, P., Smith, C., et al. (1993). Evaluation of two school smoking programmes under normal classroom conditions. *British Medical Journal*, 306, 102-107.
- Reeves, T. C. (2000, April). Enhancing the worth of instructional technology research through "Design Experiments" and other development research strategies. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Rogers, E. M. (1995). *Diffusion of Innovations*. 4th Edition. New York: The Free Press.
- Romberg, T. A. (Ed.) (2000), *A Blueprint for Maths in Context: A Connected Curriculum for Grades 5-8*. Chicago, IL: Encyclopaedia Britannica Educational Corporation.
- Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher*, 32(1), 25-28.

- Shavelson, R. J., & Towne, L. (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.
- Shooter, S. B., Keirouz, W. T., Szykman, S., & Fenves, S. (2000, September). A model for information flow in design. *Proceedings of the ASME International Design Engineering Technical Conferences*, Baltimore MD.
- Shulman, L.S. (1986) Those Who Understand: Knowledge Growth in Teaching, *Educational Researcher*, 15(2), 4-14.
- Sloane, F. C., & Gorard, S. (2003). Exploring modeling aspects of design experiments. *Educational Researcher*, 32(1), 29-31.
- Smaling, A. (2003). Inductive, analogical, and communicative generalization. *International Journal of Qualitative Methods*. 2(1). Article available online: http://www.ualberta.ca/~iiqm/backissues/2_1/html/smaling.html.
- Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp.267-306). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Stokes, D. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington DC: Brookings Institution Press.
- Tessmer, M and Wedman, J. 1995. Context-sensitive instructional design models: A response to design research, studies and criticism. *Performance Improvement Quarterly*, 8(3), p38-54.
- Toluk, Z. (1999). Children's conceptualizations of the quotient subconstruct of rational numbers. Unpublished Doctoral dissertation, Arizona State University, Tempe, AZ. Unpublished doctoral dissertation, Arizona State University.
- Toluk, Z., & Middleton, J. A. (2001). The development of children's understanding of quotient: A Teaching experiment. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Annual Meeting of the International Group for the Psychology of Mathematics Education*. Utrecht, The Netherlands: Hogrefe.
- Torgerson, C. & Torgerson, D. (2001). The need for randomised controlled trials in educational research, *British Journal of Educational Studies*, 49, 3, 316-328.
- Ullman, D. G. (1992). *The Mechanical Design Process*. New York: McGraw Hill.
- Ulrich, K.T. & Eppinger, S.D. (2000). Product design and development. Boston: Irwin McGraw-Hill
- Webb, N. L., Romberg, T. A., & Shafer, M. C. (2001, April). Variations in Student Performance. American Educational Research Association, Seattle, WA.
- Zaritsky, R., Kelly, A. E., Flowers, W., Rogers, E., & O'Neill, P. (2003). Clinical design sciences: A view from sister design efforts. *Educational Researcher*, 32(1), 32-34.

FIGURES

Figure 1. General Model of Design Research

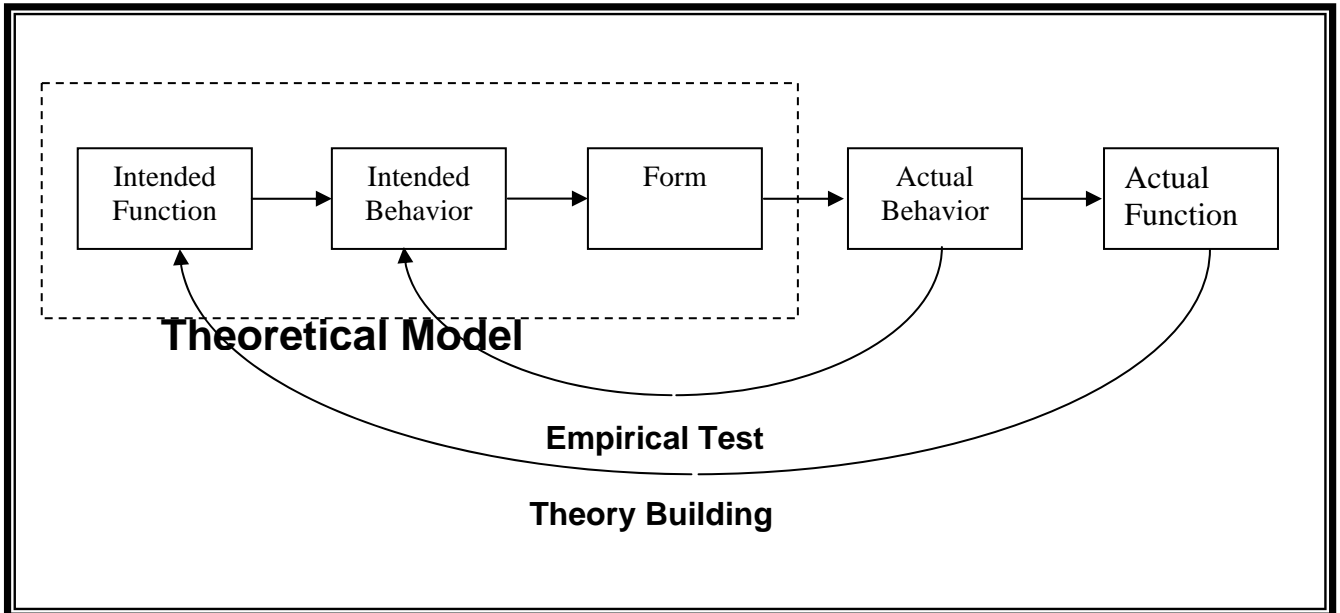
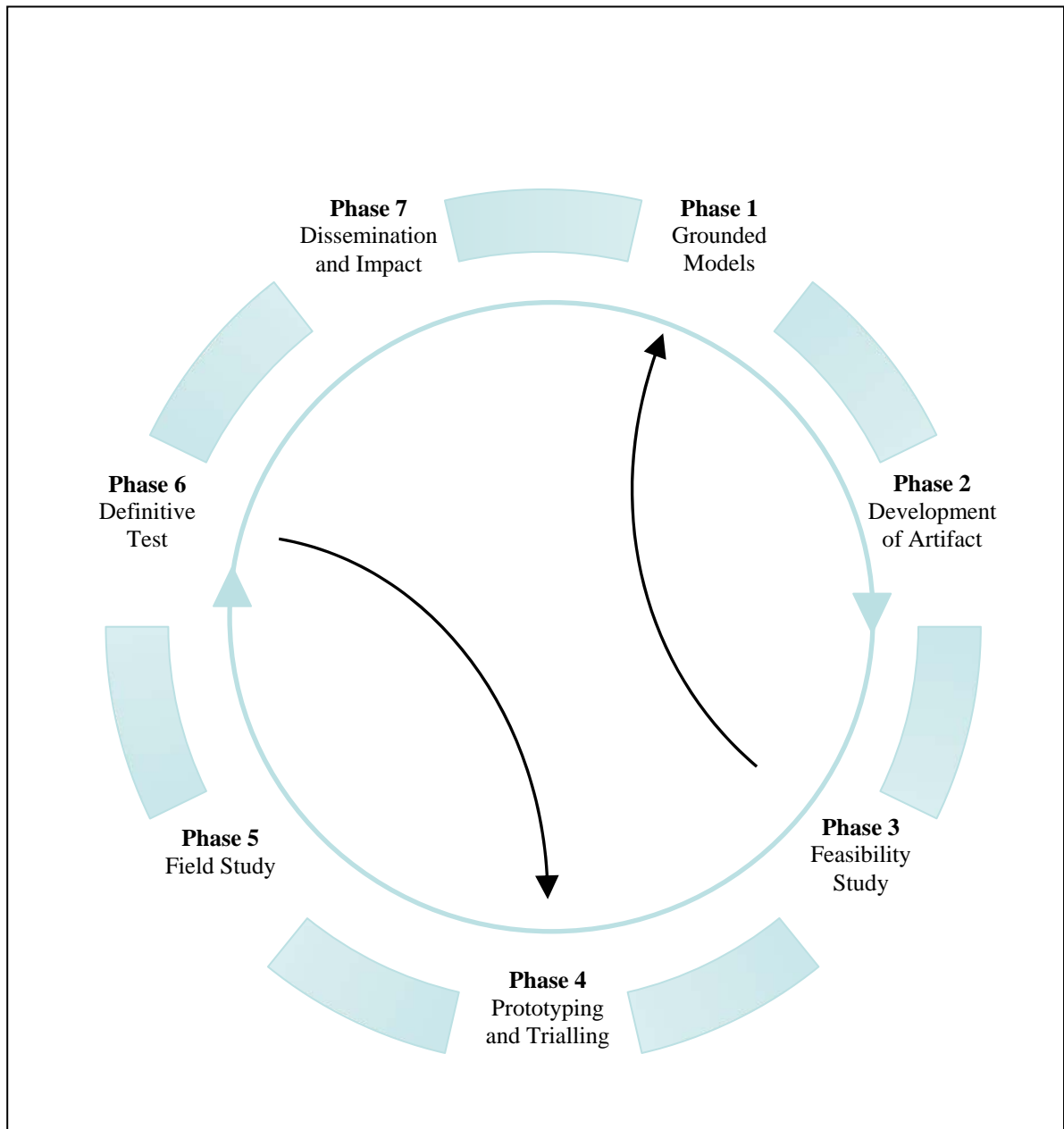


Figure 2. Basic model of scientific research in education



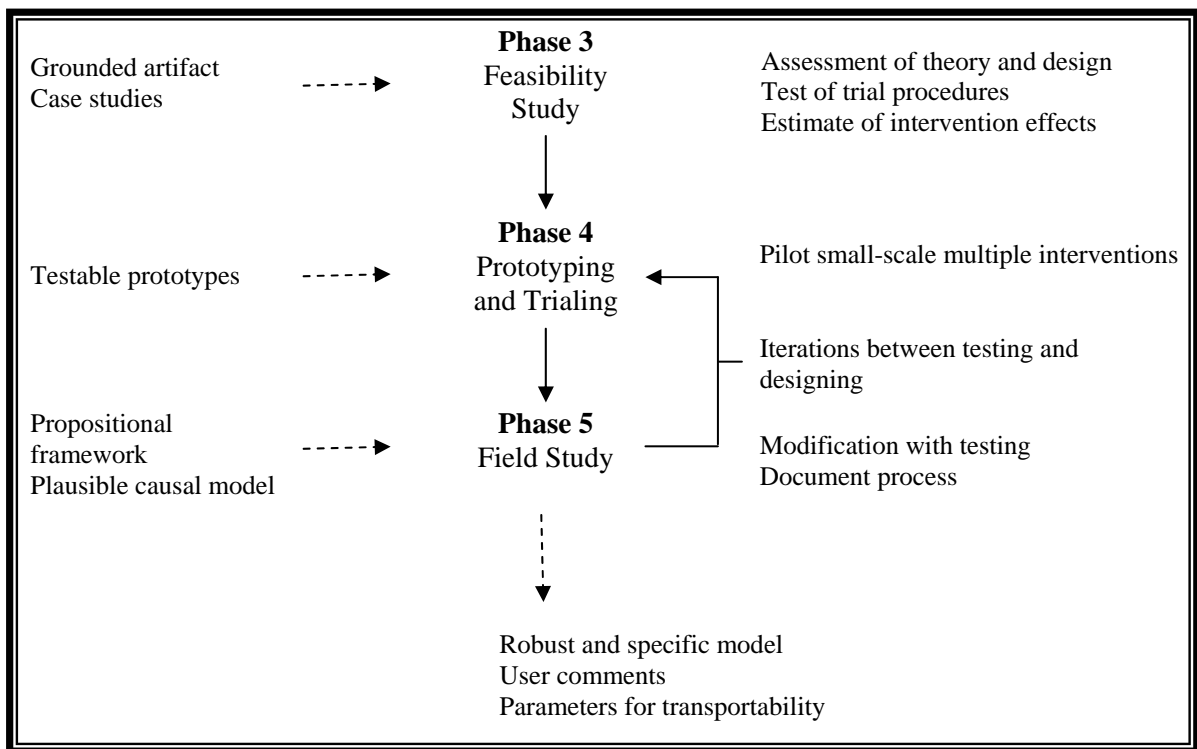
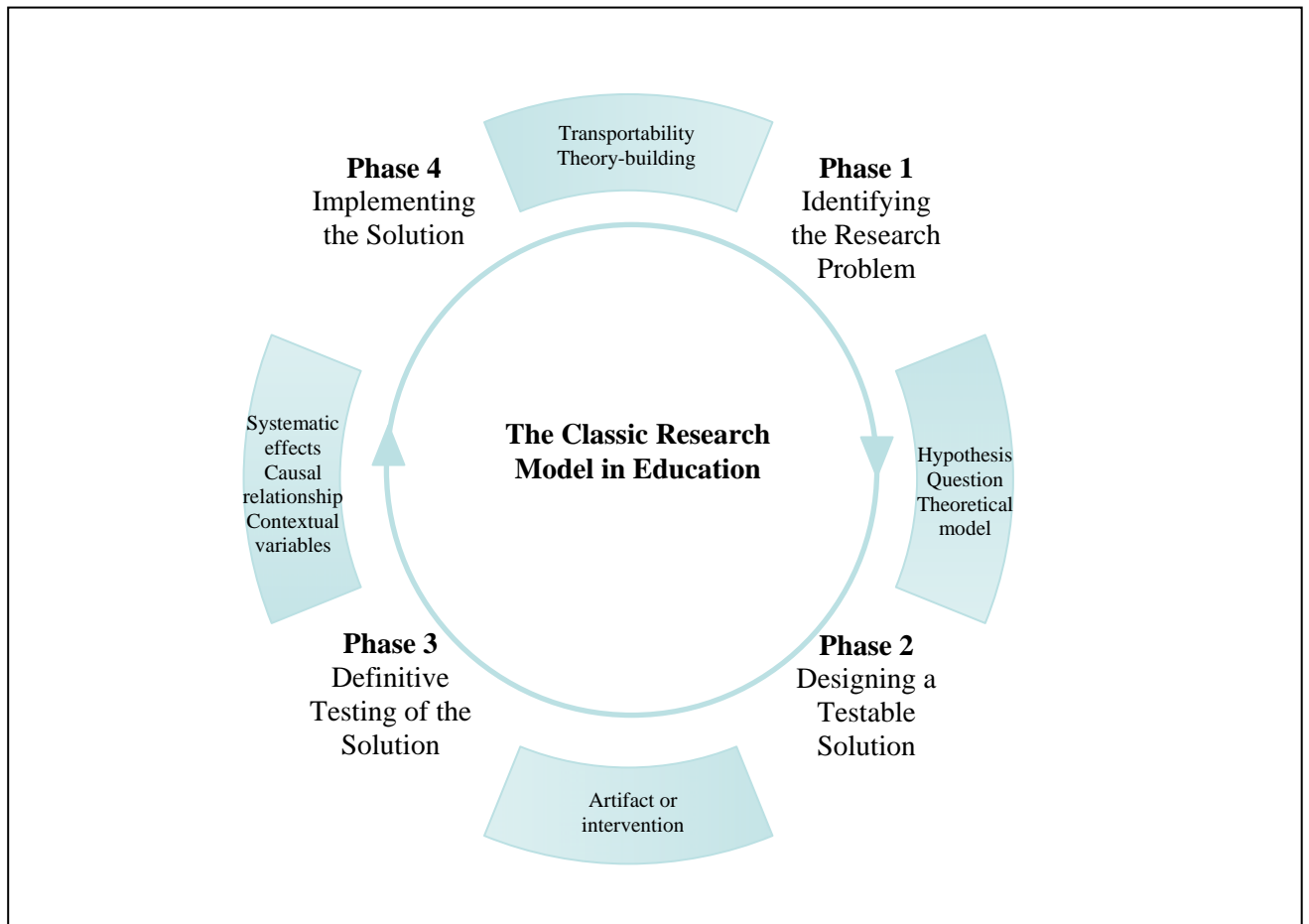


Figure 3. Three “new” phases in the design experiment – phases 3 to 6 of the “complete” research process

Figure 4. The Compleat Design Cycle



ISBN: 1 85342 633 4

THE UNIVERSITY *of York*

