Design Experiments in Educational Research

by Paul Cobb, Jere Confrey, Andrea diSessa, Richard Lehrer, and Leona Schauble

In this article, the authors first indicate the range of purposes and the variety of settings in which design experiments have been conducted and then delineate five crosscutting features that collectively differentiate design experiments from other methodologies. Design experiments have both a pragmatic bent—"engineering" particular forms of learning—and a theoretical orientation—developing domain-specific theories by systematically studying those forms of learning and the means of supporting them. The authors clarify what is involved in preparing for and carrying out a design experiment, and in conducting a retrospective analysis of the extensive, longitudinal data sets generated during an experiment. Logistical issues, issues of measure, the importance of working through the data systematically, and the need to be explicit about the criteria for making inferences are discussed.

In this short article, we draw on our collective experience of conducting design experiments for a range of purposes in variety of settings in order to delineate prototypical characteristics of the methodology and to describe what is involved in conducting a design experiment. Although the term *design experiment* is most closely associated with Brown (1992) and Collins (1992), pedagogical design has informed the development of theories of instruction for well over a century. Prototypically, design experiments entail both "engineering" particular forms of learning and systematically studying those forms of learning within the context defined by the means of supporting them. This designed context is subject to test and revision, and the successive iterations that result play a role similar to that of systematic variation in experiment.

Design experiments are conducted to develop theories, not merely to empirically tune "what works." These theories are relatively humble in that they target domain-specific learning processes. For example, a number of research groups working in a domain such as geometry or statistics might collectively develop a design theory that is concerned with the students' learning of key disciplinary ideas in that domain. A theory of this type would specify successive patterns in students' reasoning together with the substantiated means by which the emergence of those successive patterns can be supported. This emphasis on theories reflects the view that the explanations and understandings inherent in them are essential if educational improvement is to be a long-term, generative process. Design experiments ideally result in greater understanding of a *learning ecology*—a complex, interacting system involving multiple elements of different types and levels—by designing its elements and by anticipating how these elements function together to support learning. Design experiments therefore constitute a means of addressing the complexity that is a hallmark of educational settings. Elements of a learning ecology typically include the tasks or problems that students are asked to solve, the kinds of discourse that are encouraged, the norms of participation that are established, the tools and related material means provided, and the practical means by which classroom teachers can orchestrate relations among these elements. We use the metaphor of an ecology to emphasize that designed contexts are conceptualized as interacting systems rather than as either a collection of activities or a list of separate factors that influence learning. Beyond just creating designs that are effective and that can sometimes be affected by "tinkering to perfection," a design theory explains why designs work and suggests how they may be adapted to new circumstances. Therefore, like other methodologies, design experiments are crucibles for the generation and testing of theory.

Design experiments are pragmatic as well as theoretical in orientation in that the study of function—both of the design and of the resulting ecology of learning—is at the heart of the methodology. This emphasis on function in a realized context holds for all design experiments even though they are conducted in a diverse range of settings that vary in both type and scope:

- One-on-one (teacher-experimenter and student) design experiments in which a research team conducts a series of teaching sessions with a small number of students. The aim is to create a small-scale version of a learning ecology so that it can be studied in depth and detail (Cobb & Steffe, 1983; Steffe & Thompson, 2000).
- Classroom experiments in which a research team collaborates with a teacher (who might be a research team member) to assume responsibility for instruction (Cobb, 2000; Confrey & Lachance, 2000; Gravemeijer, 1994).
- Preservice teacher development experiments in which a research team helps organize and study the education of prospective teachers (Simon, 2000).
- In-service teacher development studies in which researchers collaborate with teachers to support the development of a professional community (Lehrer & Schauble, 2000; Stein, Silver, & Smith, 1998).
- School and school district restructuring experiments in which a research team collaborates with teachers, school administrators, and other stakeholders to support organizational change (Confrey, Bell, & Carrejo, 2001).

Crosscutting Features of Design Experiments

We identify five crosscutting features that apply to these diverse types of design experiments. First, the purpose of design

Educational Researcher, Vol. 32, No. 1, pp. 9–13

experimentation is to develop a class of theories about both the process of learning and the means that are designed to support that learning, be it the learning of individual students, of a classroom community, of a professional teaching community, or of a school or school district viewed as an organization. We interpret processes of learning broadly to encompass what is typically thought of as knowledge, but also the evolution of learning-relevant social practices and even constructs such as identity and interest. When we look across these diverse types of design experiments, the means for supporting learning encompass the affordances and constraints of material artifacts, teaching and learning practices, and policy levers (e.g., performance-based pay), as well as other forms of mediation that might, for example, include the negotiation of domain-specific norms—such as what counts as a "good" scientific question in a classroom (Wertsch, 1998). It is apparent from this broad view of means of support that it is often necessary to document learning ecologies at multiple levels (Kelly & Lesh, 2000). In the case of an in-service teacher development experiment, for example, the research team might focus simultaneously on the norms and practices of a professional teaching community, the participating teachers' pedagogical reasoning and instructional practices, and their students' reasoning in a particular content domain. A challenge that arises in such cases is therefore that of coordinating multiple levels of analysis.

Although, as a practical matter, a design experiment is conducted in a limited number of settings, it is apparent from the concern for theory that the intent is not merely to investigate the process of supporting new forms of learning in those specific settings. Instead, the research team frames selected aspects of the envisioned learning and of the means of supporting it as paradigm cases of a broader class of phenomena. In the case of a one-on-one design experiment, for example, the broader theoretical goal might be to develop a psychological model of the process by which students develop a deep understanding of particular mathematical ideas, together with the types of tasks and teacher practices that can support that learning. In the case of a school district restructuring experiment, the theoretical goal might be to develop an interpretive framework that explicates the relations between teachers' instructional practices and the institutional settings in which teachers develop and refine their practices. In these and other types of design experiments, the initial design formulated when preparing for an experiment and the new form of learning it is designed to support are viewed as instances of broader classes of phenomena, thereby opening them to theoretical analysis.

The second crosscutting feature is the highly interventionist nature of the methodology. Design studies are typically test-beds for innovation. The intent is to investigate the possibilities for educational improvement by bringing about new forms of learning in order to study them. Consequently, there is frequently a significant discontinuity between typical forms of education (these could be studied naturalistically) and those that are the focus of a design experiment. The design developed while preparing for an experiment draws on prior research and attempts to cash in the empirical and theoretical results of that research. The process of engineering the forms of learning being studied provides the research team with a measure of control when compared with purely naturalistic investigation. Furthermore, in attempting to support a specified form of learning, the researcher is more likely to encounter relevant factors that contribute to the emergence of that form and to become aware of their interrelations.

By its very nature, the study of phenomena as complex as learning ecologies precludes complete specification of everything that happens. It is therefore all the more important to distinguish in the specification of the design between elements that are the target of investigation and those that may be ancillary, accidental, or assumed as background conditions. For example, in a study of children's mathematical development, classroom norms of justification might be assumed as background and the emphasis placed instead on conceptual development. Alternatively, the development of norms might serve as a primary target of investigation (e.g., Yackel & Cobb, 1996). The use of prior research to both specify a design and justify the differentiation of central and ancillary conditions is central to the methodology.

The third crosscutting feature builds on the first two: Design experiments create the conditions for developing theories yet must place these theories in harm's way. Thus, design experiments always have two faces: prospective and reflective. These two faces are familiar to all empirical scientists, but the forms they take in design experiments are somewhat specialized. On the prospective side, designs are implemented with a hypothesized learning process and the means of supporting it in mind in order to expose the details of that process to scrutiny. An equally important objective is to foster the emergence of other potential pathways for learning and development by capitalizing on contingencies that arise as the design unfolds.

On the reflective side, design experiments are conjecture-driven tests, often at several levels of analysis. The initial design is a conjecture about the means of supporting a particular form of learning that is to be tested. During the conduct of the design study, however, more specialized conjectures are typically framed and tested. For example, during a classroom design experiment, an initial conjecture about a prospective interaction between characteristics of tasks as they are realized in the classroom and student responses may be tested. If this conjecture is refuted, alternative conjectures can be generated and tested.

Together, the prospective and reflective aspects of design experiments result in a fourth characteristic, *iterative design*. As conjectures are generated and perhaps refuted, new conjectures are developed and subjected to test. The result is an iterative design process featuring cycles of invention and revision. Of course, to design iteratively demands systematic attention to evidence about learning and, as we later describe, this often involves the parallel development of measures sensitive to the changing ecology of learning. The intended outcome is an explanatory framework that specifies expectations that become the focus of investigation during the next cycle of inquiry.

The fifth feature of design experimentation again reflects its pragmatic roots: Theories developed during the process of experiment are humble not merely in the sense that they are concerned with domain-specific learning processes, but also because they are accountable to the activity of design. The theory must do real work. General philosophical orientations to educational matters—such as constructivism—are important to educational practice, but they often fail to provide detailed guidance in organizing instruction. The critical question that must be asked is whether the theory informs prospective design and, if so, in precisely what way? Rather than grand theories of learning that may be difficult to project into particular circumstances, design experiments tend to emphasize an intermediate theoretical scope (diSessa, 1991) that is located between a narrow account of a specific system (e.g., a particular school district, a particular classroom) and a broad account that does not orient design to particular contingencies. For example, the claim that invented representations are good for mathematics and science learning probably has some merit, but it specifies neither the circumstances in which these representations might be of value nor the learning processes involved and the manner in which they are supported. In contrast to most research methodologies, the theoretical products of design experiments have the potential for rapid pay-off because they are filtered in advance for instrumental effect. They also speak directly to the types of problems that practitioners address in the course of their work.

Preparing for a Design Experiment

As we have emphasized, a crucial issue to be addressed when one conducts any type of design experiment is that of clarifying its theoretical intent: What is the point of the study? For illustrative purposes, we will exemplify this aim for the case of classroom design experiments, although it applies equally to other kinds of design experiments, such as those that focus on school districts or larger educational systems, out-of-school learning contexts, workplaces, and the like.

Most classroom design experiments are conceptualized as cases of the process of supporting groups of students' learning in a particular content domain. The theoretical intent, therefore, is to identify and account for successive patterns in student thinking by relating these patterns to the means by which their development was supported and organized. However, different classroom design experiments may set their focus on different constellations of issues. For example, one might focus on the relation between classroom norms or standards for mathematical or scientific argumentation, and student learning. Another study might emphasize the ways in which diversity in students' prior experiences can be capitalized upon as a resource to ensure that all students have access to significant disciplinary ideas.

In addition to clarifying the theoretical intent of the experiment, the research team must also specify the significant disciplinary ideas and forms of reasoning that constitute the prospective goals or endpoints for student learning. This usually involves drawing on and synthesizing the prior research literature to identify central organizing ideas for a domain (e.g., the notion of distribution as a central idea for statistical analysis, Lehrer & Schauble, 2002; McClain, Cobb, & Gravemeijer, 2000). In the process of specifying instructional goals, a research team frequently proposes an alternative conception of a domain (e.g., typicality, center, variation, and relative frequency as characteristics of the single, overarching idea of distribution rather than as a set of discrete curriculum topics). Another source of discontinuity in curricular specification is that new resources, such as computer software, might be developed to support the envisioned form of learning. Yet another is that evolving theories of knowledge informed by analyses of how knowledge is used in complex settings may implicate different performances as indicative of deep understanding (diSessa, in press), such as the ability to innovate procedures in small-group design episodes in contrast to individual application of a given procedure.

As part of the process of preparing for a classroom design experiment, the research team also specifies its assumptions about the intellectual and social starting points for the envisioned forms of learning. To achieve the instructional agenda, the team identifies current student capabilities, current practices, and other resources on which it might be able to build. In relatively well-researched domains, the team can draw on the literature to develop conjectures about students' initial interpretations and understandings. However, in less researched areas, the team typically needs to conduct pilot work to document these understandings and, thus, the consequences of students' prior instructional histories. In the course of this pilot work, the team might also develop new methods for assessing aspects of student reasoning that need to be documented, given the purposes of the experiment.

When the conjectured starting points, elements of a trajectory, and prospective endpoints have been specified, the challenge is to formulate a design that embodies testable conjectures about both significant shifts in student reasoning and the specific means of supporting those shifts. In well-studied domains, the research team might have a reasonable level of confidence in some of their conjectures. However, in others, where knowledge is less developed, the team regards its conjectures as speculative and begins the experiment with the expectation that many will prove to be unviable. Even then, the advantage of explicating conjectures at the outset is that they orient the research team to identify and account for successive patterns in student thinking.

The means of supporting student learning are usually construed broadly, consistent with an acknowledgement of the complexity of teaching and learning. This relatively encompassing view of the means of support implies that the research team must generate multiple forms of data to adequately document the learning ecology. Because we have focused on classroom learning, it is important to emphasize that the focus and means of documentation vary with the institutional setting. For example, in a science museum, the built environment may constitute an important means for focusing visitor attention, communicating how to initiate the activity at hand, and framing reasonable interpretations of the outcome.

Conducting a Design Experiment

As we have indicated, a primary goal for a design experiment is to improve the initial design by testing and revising conjectures as informed by ongoing analysis of both the students' reasoning and the learning environment. The size of the research team and the expertise of the members vary depending on the type and purpose of the experiment. For example, it might be feasible for a single researcher who conducts the teaching sessions and a graduate assistant who records the sessions to carry out a one-on-one design experiment. In the case of a classroom design experiment conducted in collaboration with a teacher in a relatively well-researched domain, the team might include the teacher, a researcher, and two graduate assistants. The crucial determinant in any type of design experiment is that the team collectively has the expertise to accomplish the functions associated with developing an initial design, conducting the experiment, and carrying out a systematic retrospective analysis. Thus, in an experiment with a relatively broad scope that encompasses multiple classrooms and attends to the organizational setting at the school and district level, two or more researchers might be involved whose combined expertise includes the design and analysis of classroom learning environments, professional teaching communities, and schools as institutions.

Regardless of the type of experiment, strong involvement of the leaders of the research team is essential. The locus of that participation is again defined by the scope and purpose of the experiment. Accordingly, if the scope is district reform, the team leaders will need to be actively involved in nested levels of activity, extending from policy forums (such as school board or content standards meetings) to professional development settings to classrooms. If the scope is more constrained, for example, to a single classroom, the team leaders may be present in the classroom as the design unfolds.

There are at least four important functions that require ongoing direct engagement in the research setting and the associated planning and interpretive activities. These functions collectively compose researcher leadership in the conduct of design experiments. First, a clear view of the anticipated learning pathways and the potential means of support must be maintained and communicated within the research team, even while responding to contingency. Maintaining such an overview can be a daunting challenge, even for an experienced researcher. Second, the extended nature of most design experiments calls for the cultivation of ongoing relationships with practitioners. These relationships are sustained by the negotiation of a shared enterprise, which is typically developed over the long haul as lead researchers consistently demonstrate their personal commitment. Third, because of the reciprocal emphasis on learning and the means that support it, design researchers seek to develop a deep understanding of the ecology of learning—not simply to facilitate logistics, but because this understanding is a theoretical target for the research. As part of the process of refining conjectures, subtle and often unanticipated cues need to be recognized and drawn into a larger perspective. Fourth and finally, regular debriefing sessions are the forum in which past events are interpreted and prospective events are planned for. These sessions are the sites where the intelligence of the study is generated and communicated.

One of the distinctive characteristics of the design experiment methodology is that the research team deepens its understanding of the phenomenon under investigation while the experiment is in progress. It is therefore important that the team generates a comprehensive record of the ongoing design process. It is standard procedure in most engineering disciplines to keep records to support the retrospective analysis of the experiment (Edelson, 2002). Accordingly, the research team may employ audio records of meetings and logs to document the evolving conjectures, together with the observations that are viewed as either supporting or questioning a conjecture.

In addition to self-consciously building and documenting the design and its rationale, the team members, like all researchers, have a responsibility for communicating what they learn in ways that are open to public scrutiny. This implies a commitment to generate data that support the systematic analysis of the phenomenon under investigation. At a minimum, this entails the generation of data on both learning and the means by which that learning was generated and supported. In practice, achieving these aims frequently requires the collection and coordination of a complex array of data sources—for example, products of learning (such as student work); classroom discourse; body posture and gesture; tasks and activity structures; patterns of social interaction; inscriptions, notations, and other tools; and responses to interviews, tests, or other forms of assessment. Because the team often intends to use these data sources to track changes over time, the task is further complicated by the need to collect extended records of each type. Technological support for the generation of these forms of data (e.g., video cameras, sophisticated audio-recording systems, mass electronic storage devices) enables these efforts but also imposes its own challenges (e.g., the development of tools and procedures for managing and analyzing large quantities of data).

The team draws on a variety of data sources that may bear on the broader phenomena framing any particular design experiment. Consider, for example, an experiment in which the team has framed the process of cultivating students' interests in disciplinary ideas as an explicit focus of investigation. In this case, team members might document the nature of students' engagement not only in the target classroom but also in out-of-school activities. Multiple sources of data ensure that retrospective analyses conducted when the experiment has been completed will result in rigorous, empirically grounded claims and assertions. Of course, no data collection can be complete, and the revision of the data collection procedures may be a part of the iterative process. As with traditional experimental and quasi-experimental designs, the viability of the conclusions drawn from data depends on the soundness of the process that generated the data.

Attending to the process by which data are generated means attending to the problem of measure. Much of the cleverness of excellent design experiments resides in how the team handles issues of measurement. An obvious point, although one that is often overlooked, is that all measurements (even observations) are *indexes* to constructs of interest, not the constructs themselves. For example, consider all the decisions that must be made when using video as data, even though the surface impression is one of non-problematic capture (Hall, 2000). Measures are created, not found, and decisions about the creation of measures are among the most important made. An otherwise impeccable design will produce no useful information about the phenomena of interest if problems of construct validity are not successfully resolved. Measures that are feasible to administer, and that provide precise and reliable scores, may or may not adequately capture the phenomenon of interest. Because design experiments need to generate results that do work with respect to subsequent cycles of design, they focus on problems of construct validity.

Conducting Retrospective Analysis

An educational accomplishment is characterized by contingency in which earlier events open up, enable, and also constrain the events that follow. Accounting for this process requires an historical or retrospective explanation, one that provides a trustworthy account of the process whereby a series of events—each of which is local and contingent—can be seen as part of an emergent and potentially reproducible pattern. For example, consider a third-grade class working together to explore conjectures about whether the volume of a plant's canopy grows proportionally over the plant's life cycle. One might want to understand how such a capability came to be. Producing an explanation of this kind requires showing how the students' earlier histories of learning (e.g., about geometric similarity, rates, and plants) bear on the events under consideration. Doing so requires justifying both selection (among all events) and rational reconstruction that focuses on issues of cause and relative importance of events in the class's unfolding history. For this reason, it is methodologically advantageous to cultivate diverse points of view from members of the research team. Diversity of expertise and backgrounds among members of the research team can be an important resource for developing alternative interpretations, as can asking different team members to assume primary responsibility for representing particular perspectives during the analysis.

A central challenge in conducting retrospective analyses is to work systematically through the extensive, longitudinal data sets generated in the course of a design experiment so that the resulting claims are trustworthy. As part of this process, it is important to be explicit about the criteria and types of evidence used when making particular types of inferences so that other researchers can understand, monitor, and critique the analysis. A primary aim when conducting a retrospective analysis is to place the design experiment in a broader theoretical context, thereby framing it as a paradigm case of the more encompassing phenomena specified at the outset. In this regard, retrospective analyses can be contrasted with the analyses conducted while the experiment is in progress in that the latter are typically oriented toward the goal of supporting the learning of the participants. For example, in a classroom experiment, the research team may, under the pressure of time, intuitively and successfully modify aspects of its instructional design. Retrospective analysis attempts to generate a coherent framework that accounts for these effects, thus making it possible to anticipate outcomes in future designs. In sum, retrospective analyses results in situated accounts of learning that relate learning to the means by which it can be supported and organized.

The situated nature of retrospective analyses is a strength of the methodology, given the overall goal of engineering new forms of learning and the tendency of "high" theory to pass over what may be important details in an effort to paint phenomena in uniform terms. In particular, because the resulting accounts of learning are tied to the means by which it was generated, the design team is always in a position to develop testable conjectures about how those means of support and, thus, the instructional design might be improved. "What works" is underpinned by a concern for "how, when, and why" it works, and by a detailed specification of what, exactly, "it" is. This intimate relationship between the development of theory and the improvement of instructional design for bringing about new forms of learning is a hallmark of the design experiment methodology.

In summary, design experiments are extended (iterative), interventionist (innovative and design-based), and theory-oriented enterprises whose "theories" do real work in practical educational contexts. Although design experiments share many individual characteristics with other ways of conducting science in the service of education, the constellation of crosscutting themes we have identified distinguishes a genre of science with high promise but also with a host of characteristic difficulties that researchers need to manage effectively to achieve that promise.

NOTE

The authors contributed equally to the manuscript and are listed in alphabetical order.

AUTHORS

PAUL COBB is a professor of mathematics education at Vanderbilt University, Peabody College, Box 330, Nashville, TN 37203; paul.cobb@vanderbilt.edu. His research interests include classroom instructional design and analysis, the development of professional teaching communities, the institutional setting of teaching, and issues of diversity and equity as they play out in the mathematics classroom.

JERE CONFREY is a professor at University of Texas, Austin, Department of Curriculum and Instruction, SZB 518, Austin, TX 78712; jere@mail.utexas.edu. Her research interests include cognition and multiplicative relations, functions and trigonometry, technology design, and systemic reform.

ANDREA DISESSA is Chancellor's Professor at University of California, Berkeley, Graduate School of Education, 4647 Tolman Hall, Berkeley, CA 94720; disessa@soe.berkeley.edu. His research interests include conceptual and experiential knowledge in physics, and the design and use of flexible, comprehensible computer systems for learning.

RICHARD LEHRER is a professor at Vanderbilt University, Department of Teaching and Learning, Peabody College, Box 330, 166 Wyatt Center, Nashville, TN 37203; rich.lehrer@vanderbilt.edu. His research interests include the design of learning environments for developing an understanding of mathematics and science.

LEONA SCHAUBLE is a professor at Vanderbilt University, Department of Teaching and Learning, Peabody College, 1930 South Drive, Nashville, TN 37203; leona.schauble@vanderbilt.edu. Her research interests include cognitive development, especially the development of scientific thinking and model-based reasoning.

Manuscript received August 23, 2002
Revisions received November 6, 2002
Accepted November 7, 2002