



On the Science of Education Design Studies

by Richard J. Shavelson, D. C. Phillips, Lisa Towne, and Michael J. Feuer

The authors argue that design studies, like all scientific work, must comport with guiding scientific principles and provide adequate warrants for their knowledge claims. The issue is whether their knowledge claims can be warranted. By their very nature, design studies are complex, multivariate, multilevel, and interventionist, making warrants particularly difficult to establish. Moreover, many of these studies, intended or not, rely on narrative accounts to communicate and justify their findings. Although narratives often purport to be true, there is nothing in narrative form that guarantees veracity. The authors provide a framework that links design-study research questions as they evolve over time with corresponding research methods. In this way, an integration can be seen of research methods focused on discovery with methods focused on validation of claims.

Design studies, design experiments, and teaching experiments attempt to “engineer innovative educational environments” (e.g., Brown, 1992, p. 141).¹ Such research, based strongly on prior research and theory and carried out in educational settings, seeks to trace the evolution of learning² in complex, messy classrooms and schools, test and build theories of teaching and learning, and produce instructional tools that survive the challenges of everyday practice. Design researchers are part of a tightly knit community whose members share interests in new teaching and learning technologies, in how to study meaning developed by students in educational settings, and in new research methods. However, they have by no means reached consensus on terminology or the warrants for their work—a situation that is reflected in how we phrase some of our later points.

The community has been successful in garnering resources for their research (Suter & Frechtling, 2000) and is steadily gaining support among researchers in various subfields of education. Advocates and practitioners of design studies believe that they are contributing evidence-based knowledge to complex questions about the structure and output of learning environments. And yet, their quiet revolution has been taking place while an entirely different conceptualization of “evidence-based” education has captured the imagination of federal policymakers. Requirements that “scientifically based” research inform education funding decisions began appearing in federal law in the mid-1990s, and the watershed No Child Left Behind Act of 2002 includes 111 such references. Both trends espouse a fundamental commitment to harnessing research in the service of improving educational practice and outcomes; but the trends’ underlying assumptions and modes of operation differ considerably.

Scientists as well as policymakers now ask tough questions about the nature of inferences that derive from design studies: What is the basis of *knowing* in design studies? This question is no different from the fundamental challenge in all scientific research—to demonstrate a basis for knowledge claims, to demonstrate warrants. Should we believe the results of design experiments? A group working in the design study mode might eventually produce an impressive concrete product (a curriculum or instructional unit, for example), but the issue remains—does this practical success warrant whatever knowledge claims the group makes when it reaches the end of the complex (and often tortuous) design path? The design study community has engaged these questions, which go to the core of the validity of their enterprise; this special issue is evidence of sincere and critical introspection on matters of research practice, method, and theory.

The effort to better understand the nature and validity of design studies comes at a time when policymakers, increasingly, are defining high-quality science in terms of traditional cause-and-effect methods, such as randomized experiments (see, e.g., the U.S. Department of Education’s Strategic Plan at <http://www.ed.gov/pubs/stratplan2002-07/index.html>). We support the use of such designs, not as a blanket template for all rigorous work, but only when they are appropriate to the research question. Design studies, too, offer unique and important perspectives on learning and can be applied productively, often in tandem with other methods. The strengths of design studies lie in testing theories in the crucible of practice; in working collegially with practitioners, co-constructing knowledge; in confronting everyday classroom, school, and community problems that influence teaching and learning and adapting instruction to these conditions; in recognizing the limits of theory; and in capturing the specifics of practice and the potential advantages from iteratively adapting and sharpening theory in its context. Our comments here are made in the spirit of supporting, building, and improving this genre of research.

Our interest in design studies stems from our recent participation in a National Research Council (NRC) Committee charged with articulating the nature of scientific research in education and developing a framework for supporting high quality scientific work in a federal educational research agency (NRC, 2002; Feuer, Towne, & Shavelson, 2002).³ The Committee took a catholic view toward research methods and integrated both design studies, as well as studies (in education and other fields) that rely on more conventional empirical methods, into a common framework. In the context of our deliberations about research methods generally, we considered certain questions: How are design studies different from curriculum development projects? Are design studies intended primarily as exercises in “formative evaluation” or for generating hypotheses or innovations? In basic

terms the Committee concluded that (a) design studies are creative research endeavors that hold promise especially for developing and elaborating conjectures and addressing issues of research-based practice, and (b) principles of research that apply to all scientific work must apply also to knowledge claims from design studies. Our position echoes Steffe and Thompson (2000, p. 277), who underscored the scientific intent of their design-study work: “We use experiment in ‘teaching experiment’ in a scientific sense. . . . What is important is that the teaching experiments are done to test hypotheses as well as to generate them.”

In this article we focus on the scientific status of the conclusions reached via design studies. To this end, we review the Committee’s guiding principles for scientific research with emphasis on their applicability to design studies, as we understand design studies. Our main goal is to scrutinize the knowledge claims from design studies through the lens of the guiding principles. We next take up the key epistemological issue—that of warrants for knowledge. At issue is how we come to know about the effects of educational environments on students’ (and researchers’) learning and construction of meaning. Finally, we consider an important practical question about the conduct of design studies, namely whether specific designs are well matched to the research questions they are intended to elucidate.

Guiding Principles of Scientific Research

The Committee argued that all the sciences, including scientific educational research, shared a set of epistemological or fundamental guiding principles. The Committee argued that all scientific endeavors should

- pose significant questions that can be investigated empirically,
- link research to relevant theory,
- use methods that permit direct investigation of the questions,
- provide a coherent and explicit chain of reasoning,
- attempt to yield findings that replicate and generalize across studies, and
- disclose research data and methods to enable and encourage professional scrutiny and critique.

These principles do not constitute an algorithm, checklist, or “how-to” guide; no single study is likely to encompass them all (although a well-designed and coordinated program of scientific research would). Rather, the Committee viewed the principles as professionally internalized norms that reflect dedication to the primacy of evidence; to healthy skepticism about knowledge claims; to ruling out all alternative explanations (“competitive argumentation”; see Schoenfeld, Smith, & Arcavi, 1993); to elucidating and reducing biases that might affect the research process; to disciplined, creative, and open-minded thinking; and to the free flow of constructive criticism.

Although all scientists share (at least implicitly) these norms, how they are instantiated in different fields of science varies. Each field has features that influence the questions posed and the design, conduct, interpretation, and generalization of research. Values and democratic ideals in schools, volition and diversity of people (teachers, students, administrators), variability and organization of school processes (curriculum, instruction, governance) across educational settings, roles of tools and technologies, and the need for collaboration with practitioners and school communities all shape scientific research in education. These features,

though not individually unique from other fields, are singular in their combination and require close attention to powerful contextual factors in the research process. Scholars working in a particular area collectively—such as the community of design-study researchers—establish the scientific traditions and standards for how most appropriately to apply the guiding principles to their area of study.

On Design Studies and Methods

Design studies have been characterized, with varying emphasis depending on the study, as iterative, process focused, interventionist, collaborative, multileveled, utility oriented, and theory driven (e.g., Cobb, Confrey, diSessa, Lehrer, & Schauble, this issue). Design studies are *iterative* in that they involve tightly linked design-analysis-redesign cycles that move toward both learning and activity or artifact improvement. They are *process focused* in that they seek to trace both an individual’s (or group’s or school system’s; see Cobb et al., this issue) learning by understanding successive patterns in the reasoning and thinking displayed and the impact of instructional artifacts on that reasoning and learning. They are *interventionist* in testing theory and instructional artifacts by designing and modifying real-world settings. They are *collaborative* in that they depend on the knowledge and co-work of practitioners. They are often *multileveled* in that they link classroom practices to events or structures in the school, district, and community. They are *utility oriented* with the intent of improving the effectiveness of instructional tools to support learning. And they are *theory driven* in the sense of testing (“placing them in harm’s way”; cf. Cobb et al., this issue) and advancing theory through the design-analysis-redesign of instructional activities and artifacts.

Researchers conducting design studies (e.g., Brown, 1992; Collins, 1999) have contrasted their intense, iterative, ideographic methods with traditional psychological experimentation that involves control of variables and focus on causal effects: “Design experiments . . . attempt to carry experimentation into real-life settings in order to find out what works in practice. This means giving up the notion of controlling variables and necessitates the development of a new methodology to carry out research” (Collins, 1999, p. 290). Such contrasts are helpful in clarifying the nature of design studies, but we argue that in some instances capitalizing on the respective strengths of design studies and other methods—including traditional experiments—in an integrated, comprehensive design would enhance inquiries (cf. McCandliss, Kalchman, & Bryant, this issue).

Because of their emersion in complex, multivariable settings and their focus on the changes or learning that takes place in participants (be they students, teachers, or administrators), and the development of instructional activities and tools, design-study research is often distinguished by the generation of a comprehensive record of the evolving design process. This documentation typically involves extensive video recordings of student engagement, teacher activity, and the like supplemented by other textual material, periodic interviews, and questionnaires. Such documentation serves as the basis for a retrospective analysis of what happened in the design study. The central challenge for the retrospective analyses is to systematically work through this extensive longitudinal data to arrive at and communicate rigorous, empirically grounded

claims and assertions (Cobb et al., this issue) and to enhance the likelihood of replicability (Hoadley, 2002). The retrospective analysis produces a situated, narrative account (which is story-like, including actors, actions, intentions) of learning and how it can be supported and organized. And, to establish generalizability, the narrative analysis places the study in a broad theoretical context to show if and how the study is a paradigmatic case of the phenomenon under investigation. The narrative account, then, provides “a structure for conveying a series of related events, a plot. . . . Important agents, events, causes and results are relayed . . . [and thereby] can help make explicit some of the implicit knowledge the designer . . . used to understand and implement the intervention” (Hoadley, p. 2).

Design Studies, the Human Sciences, and Warrants for Knowledge

In recognition of the complex, multivariate, multilevel, iterative, and interventionist nature of design studies, those working in this research mode have argued for intensive, longitudinal studies that trace the design process and capture meaning constructed by individual subjects over time. Naturally enough, reports of this kind of work usually take the form of narratives, whether or not the researchers concerned recognize this; but in recent years there has been considerable explicit discussion of, and advocacy on behalf of, the use of narrative accounts. Moreover, many distinctions that design-study researchers make between traditional experimental methods and design-study methods (e.g., narrative accounts and interpretive frameworks) reflect the broader debate among psychologists and philosophers of science (e.g., Bruner, 1990; Donald, 2001; Toulmin, 2001; Polkinghorne, 1988) on positivist and “postpositivist” science (Phillips & Burbules, 2000) and the need for a new epistemology that meets the needs of “human sciences.”

The premise of those who take the narrative turn is that to understand human action, “a cultural psychology” is needed, one that “must venture beyond the conventional aims of positivist science with ideals of *reductionism*, *causal explanation* and *prediction*” (Bruner, 1990, p. xiii). Bruner adds rhetorically, “Are not plausible interpretations preferable to causal explanations, particularly when the achievement of a causal explanation forces us to artificialize [*sic*] what we are studying to a point almost beyond recognition as representative of life?” (p. xiii).

The point is that it can be argued that to understand phenomena such as student learning and to document how this develops during the course of a design study, it is necessary to take into account the desires, beliefs, goals, reasoning processes, and so forth of the students over time, and that this is best done in the form of a narrative. The narrative can also capture the socio-cultural setting in which the learning occurs and which helps to shape the students’ behaviors (Bruner, 1990, and for some cautions, see Olson, 1992).

Design studies, then, often rely on narrative accounts as data for modifying theory and the design of artifacts iteratively over time. They also rely on narrative to communicate their research findings (although the use of the narrative format is often not explicitly commented upon). The argument for the validity of such an account lies in the reasonableness of the argument (“explanatory framework”) as judged by, for example, expert practitioners

and researchers (“communities of practice”): “We know from our own experience in telling consequential stories about ourselves that there is an ineluctably ‘human’ side to making sense” (Bruner, 1990, p. 55). To the extent that the narrative evokes script and scene schema similarly among those well steeped in the phenomena under study, the narrative is claimed to be reasonable, credible, and generalizable.

For us the issue is that with so many confounding variables in a design study, can the knowledge claims be warranted? Our concern is not alleviated by the narrative turn suggested by some theorists and by some members of the design-study community. The NRC Committee’s guiding principles lead us to certain questions: To what extent can rival narrative accounts of the same action be ruled out? To what extent would another narrator replicate the account? To what extent does the narrative generalize to other times and places? There is nothing in the use of narrative form, by itself, that guarantees the veracity of the content of the account or which vitiates the need for the usual epistemic warrants used in science (see Phillips, 2000). As Olson (1992, p. 30) put it in his review of Bruner’s recent books on mind and narrative, “just because a doctrine is ‘full of comfort’ is not an argument for its truth.” In concrete terms, to what extent were decisions about the design of instructional artifacts based on information that ruled out other possible alternatives? To what extent would another narrative of the use of the artifact in learning replicate the one on which decisions were being made? How can it be determined that the narrative being used is complete, or does not misrepresent events? To what extent would the tool developed in one intensively engineered setting generalize to another setting?

The basis of narrativist knowledge claims—for example, that practitioners and researchers are able to recognize and understand the explanatory framework of the narrative based on the script and scene schema the narrative evokes—is provocative. But, who selects the practitioners and researchers? We could imagine a probability sampling basis for making some knowledge claims from narratives, selecting representative practitioners and researchers, and seeing if the narrative evokes the same understanding of human actions and meaning among them.⁴ However, we also stress that understanding an explanatory framework is one thing; whether the framework is well warranted or not is quite another matter. Even if practitioners and researchers agree in their interpretation of the narrative, on what grounds could the narrative be said to correspond to what actually transpired or would the narrative arise from the needs and imagination of the storyteller? For Bruner (1990, p. 44) this is not a problem; narrative “can be ‘real’ or ‘imaginary’ without loss of its power as a story.” For him (but not for us),⁵ plausibility rather than verifiability is the key validity issue.

There is more at stake here than understanding actions. Some narratives will inevitably (perhaps unknowingly) make causal claims (e.g., we did *x* and it produced *y*). But narratives lacking more traditional controls on extraneous variables will not be able to warrant the causal claim. As we have said before, we believe that the principles for scientific research apply to design studies. To ensure that they are scientifically rigorous, it is incumbent on those carrying out design studies to create a culture of science that includes ruling out competing hypotheses, that fosters skepticism about knowledge claims (including their own), and that

encourages powerful tests of rival conjectures. We now turn to question-driven methods that might apply to various phases of design studies.

Design Studies, Research Questions, and Methods

Our Committee was careful to point out that research questions should drive the choice of research methods. The questions driving design studies probably depend on what stage in the design evolution they are posed. In early stages, in the so-called context of discovery, open-ended exploration is common to design studies, just as it is in any other branch of science. This wide-ranging exploration turns into systematic descriptions and evolves into well-formulated questions, creating a context for verification. The design-study question—and, consequently, the research method—most likely depends upon where in the evolution of the design it is posed. The Committee identified three generic questions within which design studies might be appropriate given their developmental stage and linked various types of research designs to each:

- *What is happening?* This question invites description of various kinds including, for example, properly characterizing a population of students with a statistical sample; understanding the scope and severity of a problem through survey, ethnographic, or case study methods; or identifying changes over time from both an insider's and outsider's perspective by case study, interview, ethnographic, and other methods. In design studies, such questions most likely arise at the beginning of the evolutionary process. The rich descriptive information collected in design studies certainly can use both quantitative and qualitative methods for description, as well as qualitative narrative accounts based on warranted knowledge claims. These claims, if clearly identified, might then be combined with information about context and motivations into a narrative to communicate to practitioners. Such a narrative might suggest possible explanations and ways for redesigning learning environments or instructional artifacts.
- *Is there a systematic effect?* This question is about the intent to establish cause and effect. The Committee concluded that randomized experiments, when feasible, are the best approach to ferreting out cause-and-effect relationships. We believe that experiments, quasi-experiments, and causal models have a role to play in design studies. The role of these methods might differ depending on where in the design evolution they are applied. As well-formulated questions arise, for example, about which alternative activities or changes in an applet are most likely to lead to a desired outcome, a small, randomized trial might be used within a classroom. To be sure, experiments should not be brought out willy nilly anytime a design decision is made. Rather they should be reserved for choices among important design alternatives. Moreover, experiments are particularly apt at the scaling up stage of a design study. An artifact or program that has been shown to work locally would be transported to and tested in other locales. The use of experimental studies combined with case (and other) studies of implementation seems appropriate to test the generalizability (and limits) of effects.
- *Why or how is it happening?* This question seeks a causal agent or mechanism: How does x cause y ? It might be asked once a systematic effect between x and y has been established, or,

alternatively, underlying theory might drive the question. In design studies theory often drives the design of activities or artifacts with a tentative causal explanation or mechanism. Through iterative tryout-redesign-tryout, claims for understanding the mechanism are advanced, and the question of replicability and generalizability then comes into play.

Final Comments

We believe that the wide range of questions posed in educational research calls for a healthy diversity of scientific methods. The questions and methods may range from pre-science exploration to well-warranted descriptive, causal, and mechanism-driven studies. We believe it is possible for those doing design studies to incorporate our guiding principles into their enterprises and, indeed, many already have. And we believe that coupling scientifically warranted knowledge and rich contextual information in some narrative form might lead to increased understanding and use of scientific research in practice.

NOTES

We thank Maria Araceli Ruiz-Primo and two anonymous reviewers for their helpful comments on the article. Errors of commission and omission, of course, are ours.

¹ We use the term *design study* to refer to this genre of research recognizing reasons others prefer terms that include *experiment*. We avoid the term *experiment* so as not to confuse it with other writing on experiments as randomized trials in social and behavioral science.

² Design studies focus on the evolution of learning. The learning might be that of a student, a teacher, or an organization.

³ Although we draw on our experiences with the Committee, this article reflects the views of the authors and not other members of the Committee or the NRC.

⁴ Such a strategy might satisfy Toulmin's (2001) call for a middle ground between the conventional view of rationality and his preference for *reasonableness* in research.

⁵ Or for Olson (1992, p. 30).

AUTHORS

RICHARD J. SHAVELSON is a professor at Stanford University, School of Education, 485 Lasuen Mall, Stanford, CA 94305-3096; richs@stanford.edu. His research interests include social science research methods and science teaching, learning, and assessment.

D. C. PHILLIPS is a professor of education and (by courtesy) philosophy at Stanford University, 485 Lasuen Mall, Stanford, CA 94302-3096; dcpphd@stanford.edu. His research areas of interest include philosophical and methodological issues in social science and educational research and constructivist positions in psychology, sociology, and science (including science education).

LISA TOWNE is a senior program officer at the Center for Education of the National Research Council of the National Academies, 500 5th Street, NW, Washington, D.C. 20001; ltowne@nas.edu. Her current interests include the conduct, use, and infrastructure of education research.

MICHAEL J. FEUER is the executive director of the Division of Behavioral and Social Sciences and Education at the National Research Council of the National Academies, 500 5th Street, NW, Washington, D.C. 20001; mfeuer@nas.edu. His research interests include public policy, social science theory and practice and education.

Manuscript received June 3, 2002

Revisions received November 6, 2002

Accepted November 7, 2002