a0005

Design Experiments

A Collins, Northwestern University, Evanston, IL, USA

Glossary

g0005 **Design experiment** – A formative experiment carried out to evaluate a particular design for a learning environment. It usually involves multiple iterations in order to identify problems with the design and to make refinements to improve the design.

g0010 **Randomized-control design** – The design of a comparative intervention experiment, where different groups of learners are assigned randomly to different conditions.

p0005 The evolving methodology of design experiments began as a reaction to traditional psychological experimentation (Brown, 1992; Collins, 1992) which had dominated educational research in teaching and learning. The methodology of psychological experimentation was based upon notions of controlling variables in order to be able to precisely determine what causes different effects. This led to conducting experiments under laboratory conditions according to carefully defined procedures. Design experiments, in contrast, attempt to carry experimentation into real-life settings, in order to refine the design so that it works in practice. This means giving up the notion of controlling variables, and therefore necessitates developing a new methodology for carrying out research.

p0010 While the initial work on this topic was carried out under the rubric of design experiments or design studies, more recently the terms design research and design-based research have come into common usage. Most of the research carried out in this vein concerns the design of computer-based learning environments. Designers of computer-based systems often go through multiple versions of a system until all the bugs are worked out, and in applying this approach to the design of learning environments, they bring a product-design mindset to the enterprise.

p0015 The novelty of the design-experiment methodology is seen most strikingly by comparing it to the experimental methodology used to study human learning in psychological literature. Learning research started before the turn of the century with the German psychologist Hermann Ebbinghaus, who invented the nonsense syllable in order to be able to study learning in its purest form. He identified many of the most important variables that affect learning, such as the similarity of stimuli to each other and the nature of the activity between learning and recall. This tradition of research on learning continues to this day and has evolved to address questions about how humans learn to solve problems and carry out complex tasks. It has produced many important findings about the conditions that affect both learning and transfer.

p0020 Seven major differences can characterize how design-experiment methodology that is currently evolving differs from the kind of psychological methodology that has dominated education research heretofore:

1. *Laboratory settings versus messy situations.* Experiments conducted in laboratories avoid contaminating effects. Learners concentrate on the task without any distractions or interruptions. The materials to be learned are well defined and are presented in a standardized manner, rather than the manner a particular teacher may choose at any given moment. In fact, the presentation is usually one directional, rather than relying on interactions between teachers and learners. In short, learning in a laboratory does not look anything like what goes on in a typical classroom, workplace, or home, where most learning occurs in life. Design experiments are set in the messy situations that characterize real-life learning, in order to avoid the distortions of laboratory experiments.
2. *A single dependent variable versus multiple dependent variables.* In most psychological experiments, there is one dependent variable, such as the number of items recalled or the percent correct on a test of some kind. In design experiments, there are many dependent variables that matter, though the experimenter may not pay attention to them all. They fall into three types of variables: (a) climate variables, such as engagement of the learners, cooperation among learners, and risk taking by learners, (b) outcome variables, including the learning of content, skills, strategies, and dispositions, and (c) system variables, such as spread of use, sustainability, and ease of adoption.
3. *Controlling variables versus characterizing the situation.* Psychological experiments use a methodology of controlling variables borrowed from early physics. The goal is to identify a few independent and dependent variables, and hold all the other variables in the situation constant. Therefore, for example, if the experimenter regards amount of learning as the dependent variable, the goal will be to hold motivation constant. But the goal of teachers in classrooms is to find ways to motivate students, so that they learn something. Thus, holding

motivation constant fundamentally undermines the usefulness of the results. In design experiments, there is no attempt to hold variables constant, but instead the goal is to identify all the variables, or characteristics of the situation, that affect any dependent variables of interest. The goal is not only to characterize what affects any dependent variable, but also to identify the nature and extent of the effect.

4. *Fixed procedures versus flexible design revision.* Psychological experiments follow a fixed procedure that is carefully documented, so that it can be replicated by other experimenters. Design experiments, in contrast, start with planned procedures and materials, which are not completely defined, and which are revised depending on their success in practice. For example, Brown (1992) developed a design called Fostering a Community of Learners, where elementary school children worked in groups to learn about ecology. In design experiments, the experimenter should characterize what happens as completely as possible, and document any changes made in the plans, together with the reasons for the changes. The goal is to start with teaching methods that are most likely to succeed, but to monitor how they are working and to modify them when appropriate. This progressive refinement is standard practice in the product-design community, as can be seen in the many refinements that are made in products over time. However, until recently, progressive refinement was not the approach taken with education innovations, because of the strictures for replicability on the experimental methods, inherited from psychology.
5. *Social isolation versus social interaction.* In most psychological experiments, the subjects are learning in isolation. There is no interaction with other learners and usually no interaction with a teacher or expert; the material to be learned is simply presented by text or video. By contrast, design experiments are set in complex social situations, such as a classroom where students may be working in groups (Brown, 1992). In consequence, students are sharing ideas, distracting and making fun of each other, being interrupted in their work, trying to make life difficult for the teacher, etc. Design experiments have to cope with all the noisy data that arise from such situations.
6. *Testing hypotheses versus developing a profile.* In psychological experiments, the experimenter has one or more hypotheses, which are being tested by systematically varying the conditions of learning. In design experiments, the goal is to see what conditions lead to different effects. Design experiments ideally are much more like what consumer reports do when they evaluate the quality of different automobiles. The goal is to look at many different aspects of the design and develop a qualitative and quantitative profile that characterizes the design in practice. There are a large number of contextual variables that determine the success of an innovation, such as the setting and professional development needed. It is best if evaluation is done with respect to a number of dimensions in a comparative fashion, as when consumer reports evaluate different products.
7. *Experimenter versus co-participant design and analysis.* In psychological experiments, the experimenter makes all decisions about the design and analysis of the data in order to maintain control of what happens and how it is analyzed. In design experiments, there is an effort to involve the different participants in the design, in order to bring their different expertise into producing and analyzing the design. Thus, teachers, curriculum designers, technology experts, cognitive psychologists, and anthropologists may all be involved in developing the design and evaluating its effects. Design experiments require many resources to be staged and hence it makes sense to bring to bear wide expertise in their design and evaluation.

Methodology of Design Research s0005

The design-research community is developing criteria for carrying out and reporting on design experiments. Not every design experiment embodies these criteria, but they characterize the elements the design-research community is responsible for. In an ideal world, design research will move in the direction of embodying many of the practices we outline here. But it will take teams of researchers and accessible archives documenting design experiments to make these practices possible. p0025

Implementing a Design s0010

Each implementation of an education design is different. Therefore it is important to identify the theory behind the design, the critical elements of the theory for the design, and how the elements fit together. In order to evaluate any implementation, one needs to analyze each particular case in terms of these key elements and their interactions. Some elements will be implemented more or less as the designers intended, some will be changed to fit the circumstances, and some will not be implemented at all. What is needed is a profile for each implementation as to how each of the critical elements were implemented and how well the elements in the implementation worked together toward the designers' goals. p0030

Modifying Designs s0015

A goal of design research is to improve the way a design operates in practice. The teacher or researchers may see that an element of the design is not working in the course of the experiment. It is important to analyze why it is not p0035

working, and take steps to fix whatever problems appear to be the reasons for failure. In this way we collect information about failures, plus information gathered from the attempted repairs to the design, and whether they succeed or fail. It is critical to document the failures and revisions, as well as the overall results of the experiment.

p0040 The experimental methods inherited from psychology that assume a fixed procedure are used throughout the experiment. Design research assumes continuous refinement. This difference has deep ramifications and requires changes in the way researchers analyze and report what is done. They should document their designs in detail, recording all major changes in design. These design changes mark the borders between phases. The goal then is to characterize the design elements that are in place in each phase and the reasons for the transitions from each phase to the next. Data relevant to research questions should be collected in each phase. For example, if there were four phases in a particular implementation, then it would be good if there were an intermediate assessment of learning outcomes between phases 2 and 3, as well as pretests and posttests. A detailed design history of this kind allows research audiences to evaluate the credibility of design decisions and the quality of lessons learned from the research.

s0020 Multiple Ways of Looking

p0045 Rogoff (1995) calls for analysts of learning environments to attend to three critical aspects: the personal layer (the experience of the individual), the interpersonal layer (one-on-one interactions), and the community layer. In the context of design experiments, researchers must additionally attend to interactions of learners with elements of the environment. There are many different aspects of what makes for an effective design, and so both designers and evaluators need to wear many hats in order to design and assess educational interventions. Consider some of the different aspects that are relevant to educational designs:

- *Cognitive level.* What do learners understand before they enter a particular learning environment, and how does that understanding change over time? Some of the tools for analysis at this level include observations of thinking through learners' representations and explanations. Through visual and verbal descriptions of ideas, researchers ask learners to expose their thinking. Are the explanations clear? Do representations capture important relationships?
- *Interpersonal level.* This viewpoint addresses how well teachers and students interact personally. Is there sharing of knowledge? Have the students bonded with each other so that they respect and help each other? Researchers use ethnographic techniques to observe these kinds of interactions.
- *Group or classroom level.* This viewpoint addresses issues of participant structure, group identity, and authority relationships. Is everyone participating? Is there a sense of the goals and identity of the group? Again, ethnography is an effective approach to analysis.
- *Resource level.* This level deals with what resources are available to learners and if they are easy to understand and use. How accessible are the resources? How well are they integrated into the activities?
- *Institutional or school level.* At this level, issues arise as to communication with outside parties and support from the entire institution. Are parents happy with the design? Do administrators support it strongly? What are the micro-political issues that impact the design?

p0050 These levels are very much intertwined. To design and assess these different issues require many different kinds of expertise: teachers, administrators, psychologists, anthropologists, media designers, etc. Conceivably one person can address all these different perspectives, but it helps to have them all represented explicitly.

Characterizing Dependent Variables s0025

p0055 Success or failure of an innovation cannot simply be evaluated in terms of how much students learn on some criterion measure. Different kinds of evaluation are necessary for addressing questions such as how sustainable the design is after the researchers leave, how much the design emphasizes reasoning as opposed to rote learning, how the design affects the attitudes of students, etc. To evaluate different variables, it is necessary to use a variety of evaluation techniques, including standardized pretests and posttests, survey and interview techniques, and systematic scoring of observations of the classrooms. Both qualitative and quantitative evaluations are essential parts of design-research methodology.

p0060 At least three types of dependent variables are important to assess: (1) climate variables, such as engagement, cooperation, risk taking, and student control; (2) outcome variables, such as content knowledge, skills, dispositions, metacognitive strategies, and learning strategies; and (3) systemic variables, such as sustainability, spread, scalability, ease of adoption, and costs.

p0065 Evaluating climate variables requires observational techniques, either by producing field notes while observing the intervention in practice, or collecting video records of the intervention and scoring those records subsequently. For example, these techniques might be used to evaluate three kinds of climate variables: the degree of engagement of students in learning in the classroom, the degree of cooperation among students in the classroom, and the degree of effort students are making to

understand the curriculum topic. To evaluate these variables, one might collect videos of different classes spread out over the time the teacher is carrying out the designed intervention. These videos can be scored systematically by multiple raters using a five-point scale for each specified interval in the lesson. Raters would be trained using benchmark lessons for which scores have been calibrated with experts.

p0070 Outcome variables are best assessed by collecting pretest and posttest measures. For example, pretests and posttests can be used to evaluate three kinds of learning variables: content, reasoning, and dispositions. To evaluate learning of content and reasoning, it is possible to use short-answer or essay questions, oral interviews, or multiple-choice items. By using items from standardized tests, it is possible to compare performance to national norms for the items. To evaluate learning of dispositions, one might apply instruments developed by Dweck (1986) to assess whether there are changes in students' beliefs reflecting a move from performance goals to learning goals. There have been such changes reported in a design experiment carried out by Scardamalia *et al.* (1994).

p0075 Systemic variables are best evaluated by interviews and surveys. For example, one might evaluate systemic variables, such as the ease of adoption of a design into the curriculum, the degree to which it is sustained in subsequent years, and the spread of use to other teachers and students. These can be measured by surveys and structured interviews with teachers and students. It is possible to develop a questionnaire that addresses the advantages and difficulties teachers encountered in adopting a design in their classroom. The other variables can be evaluated by surveys administered to both teachers and students at regular intervals. The surveys will ask about what aspects of the design are being sustained and are spreading, and which aspects are not.

s0030 Characterizing Independent Variables

p0080 In evaluating any design, there are a large number of independent variables that may affect the success of the design in practice. It is important to determine what general aspects of the situation researchers need to consider in order to decide what is affecting the success of the design. The contextual variables that can determine the success of an innovation include the following:

- *Setting.* The setting of the learning environment is a critical variable in how any design fares. The setting might vary over homes, workplaces, museums, schools, or colleges; elementary, middle or high schools; public or private schools; urban, suburban, or rural schools; elite or community colleges; etc. How broadly applicable an innovation is can be determined only by trying it out in many different settings.
- *Nature of the learners.* Critical variables about the learners include things such as their age, socioeconomic status, turnover rate, attendance rate, etc. For example, some innovations may work with weaker students and some with gifted students. So it is important to determine for which type of learners the design is effective, and in what ways.
- *Required resources and support for implementation.* In order to carry out any design, there will be a need for resources and supports of various kinds, including materials, technical support, administrative support, and parent support. If a design requires teachers to gather materials, spend time in preparation or other activities, enlist administrators or parents to make the design succeed, then these requirements need to be identified.
- *Professional development.* Often in order for a design to be successful, teachers (and perhaps others) need to be provided with professional development of various kinds. These can encompass workshops, design meetings, courses, videos of exemplary practice of the design, guided practice with expert practitioners, reflective meetings with colleagues, etc. Identifying what teachers need to implement the design successfully is an important aspect of designing an innovation.
- *Financial requirements.* Any intervention adds costs that need to be tracked, including equipment costs, service costs, professional support and development costs, replacement costs, etc. Very often substantial costs, such as technical support and replacement costs, are ignored when calculating the cost of a technological innovation.
- *Implementation path.* This term covers the variables involved in implementing a design, such as how the innovation is introduced, the time devoted to it, the difficulties teachers may face in introducing the design, etc. There is a structure to the introduction and evolution of a design that needs to be characterized in analyzing any implementation.

Reporting on Design Research s0035

The experimental literature developed a conventional p0085 structure for reporting on experiments that evolved over time. The structure consists of four parts: background to the problem, experimental method, results, and discussion. As design research reconceptualizes the experimental process, there needs to evolve a different structure for reporting, perhaps including five sections in reporting on design experiments:

- *Goals and elements of the design.* An important aspect of reporting on design experiments is to identify the theory behind the design, the critical elements of the theory, and how they fit together to accomplish the goals of the design. The critical elements of a design may be materials, activities, a set of principles, or some

combination of all these. It is equally important to describe the goals of the design and how all the elements are meant to work together to attain those goals. Goals, critical elements, and their interactions need to be described in enough detail, so that it is possible to evaluate how well the design was implemented in different settings.

- *Settings where implemented.* The description of the settings needs to include all the information relevant to the success of the design outlined in the section titled 'Characterizing independent variables' above. Differences between how the design was implemented in each setting should be detailed, so that readers can evaluate how faithfully the design was carried out in each setting.
- *Description of each phase.* The design is likely to go through a different evolution in each setting, so it is necessary to describe each phase in each setting. When changes are made in a setting, the reasons for the changes should be specified along with the effects of making the changes. It also makes sense to describe how the critical elements of the redesign accomplish the goals of the original design or how the goals have changed.
- *Outcomes found.* The outcomes should be reported in terms of a profile of values on the dependent variables in the different settings, just as qualitative and quantitative data are reported about different products in consumer reports. These should be included to the extent that intermediate data describing the different phases were collected.
- *Lessons learned.* Considering what happened in the different implementations, the report should attempt to pull together all the findings into a coherent picture of how the design evolved in the different settings. It is important to describe the limitations and failings of the design, as well as the successes, both in implementation and outcomes.

s0040

Implications for Summative Evaluation

p0090 While design experiments were conceived as a formative evaluation strategy, the principles involved do have implications for summative evaluation. Any such assessment of educational innovations must carry out both quantitative and qualitative assessments, using a randomized-control design as Cook (2002) advocates, and comparative analysis, which consumer reports use. For example, to compare how effective two reading programs are, one would need to carry out comparative analyses in a variety of different settings, such as urban, suburban, and rural schools, and perhaps even homes, workplaces, and military settings. In such studies there must be a fixed experimental procedure, unlike the flexible design revision necessary for formative evaluation. The assessment should produce a profile that shows the strengths and weaknesses of the designs being compared. Hence, different designs might be found to be more effective in some settings or with regard to some outcomes.

p0095 In order to have a sound assessment process, educational researchers as a community should develop a consensus process to determine what variables to look at and how to assess them. The assessment should address the multiple concerns of different stakeholders, including developers, and so they should be included in the consensus process. The design-research methodology argues that we need to look at multiple contextual and dependent variables, as described earlier.

p0100 To carry out such evaluations effectively, the country would need to invest in an independent agency, in the style of Consumer's Union, with the expertise to carry out comparative evaluation. Such an agency could develop the expertise and methods for looking in a cost-effective manner at innovations in use, in a way that best informs the many different stakeholders.

s0045

Conclusion

p0105 Brown (1992) felt that laboratory experiments, ethnographies, and large-scale studies are all valuable methodologies to study learning, but that design experiments fill a niche these methodologies do not address. Specifically, they allow researchers to evaluate and refine learning environments that are designed on particular principles and then revise the environment and the principles. Tharp and Gallimore (1982, 1988) have elegantly described how different methodologies can most effectively work together.

p0110 It is clear from the spread of these kinds of research methods (Barab and Kirshner, 2001; Barab, 2004, 2006; Edelson, 2001; Design-based Research Collective, 2003; Kelly, 2003; Sandoval and Bell, 2004) that design research is becoming an established practice. But design experiments often lead to the collection of large amounts of data that go unanalyzed. Hence, it makes sense for the design-research community to establish an infrastructure that would allow researchers at other institutions to analyze the data collected in design studies, in order to address their own questions about learning and teaching. This would require the community to honor such reanalysis of data with the same status as original research and it would require research journals and tenure committees to take such work seriously. Other fields, such as child language, have developed widely available archives of data, enabling researchers to discuss and analyze the same data from many different perspectives. The design-research community should strive to set up an infrastructure that can support researchers at different sites in analyzing the large data sets that design experiments are now producing.

Bibliography

Barab, S. A. (ed.) (2004). Special issue: Design-based research: clarifying the terms. *Journal of the Learning Sciences* **13**(1), 1–128.

Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 153–170. New York: Cambridge University Press.

Barab, S. A. and Kirshner, D. (eds.) (2001). Special issue: Rethinking methodology in the learning sciences. *Journal of the Learning Sciences* **10**(1 and 2), 1–222.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences* **2**(2), 141–178.

Collins, A. (1992). Toward a design science of education. In Scanlon, E. and O'Shea, T. (eds.) *New Directions in Educational Technology*, pp 123–130. Berlin: Springer.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis* **24**(3), 175–199.

Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher* **32**(1), 5–8.

Dweck, C. (1986). Motivational processes affecting leaning. *American Psychologist* **41**(10), 1040–1048.

Edelson, D. C. (2001). Design research: What we learn when we engage in design. *Journal of the Learning Sciences* **11**(1), 105–121.

Kelly, A. E. (ed.) (2003). Theme issue: The role of design in educational research. *Educational Researcher* **32**(1), 3–34.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In Wertsch, J. V., Rio, P. D., and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, p 252. Cambridge: Cambridge University Press.

Sandoval, W. A. and Bell, P. (eds.) (2004). Special issue: Design-based research methods for studying learning in context. *Educational Psychologist* **39**(4), 199–260.

Scardamalia, M., Bereiter, C., and Lamon, M. (1994). CSILE: Trying to bring students in world 3. In McGilley, K. (ed.) *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice*, pp 201–228. Cambridge, MA: MIT Press.

Tharp, R. G. and Gallimore, R. (1982). Inquiry processes in program development. *Journal of Community Psychology* **10**, 103–118.

Tharp, R. and Gallimore, R. (1988). *Rousing Minds to Life*. New York: Cambridge University Press.

Further Reading

Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society* **9**(3), 321–327. http://ikit.org/people/~bereiter.html (accessed September 2009).

Collins, A., Joseph, D., and Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences* **13**(1), 15–42.

Confrey, J. (2006). The evolution of design studies as methodology. In Sawyer, R. K. (ed.) *Handbook of the Learning Sciences*, pp 135–152. New York: Cambridge University Press.

diSessa, A. A. and Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences* **13**(1), 77–103.

Tharp, R. and Gallimore, R. (1982). Inquiry processes in program development. *Journal of Community Psychology* **10**, 103–118.

EDUC: 01296

Non-Print Items

Abstract:
The term design experiments was introduced in papers by Brown (1992) and Collins (1992). Design experiments were developed as a way to carry out formative research to test and refine educational designs based on principles derived from prior research. Then, based on an analysis of the learning environment, modifications and additions are made to the learning environment and, in some cases, to the underlying learning principles themselves. The article provides guidelines for how design research can best be carried out. In recent years, the terms design research and design-based research have been applied to this kind of work.

Keywords: Design-based research; Design experiments; Design research; Educational evaluation; Formative evaluation; Methodology; Program evaluation; Qualitative methods; Quantitative methods; Research methods; Summative evaluation

Author and Co-author Contact Information:

A Collins
135 Cedar St.
Lexington
MA 02421
USA
+1 781 861 8263
Collins@bbn.com

Biographical Sketch for Online Version

Dr. Allan Collins is professor emeritus of education and social policy at Northwestern University. He is a member of the National Academy of Education, a fellow of the American Association for Artificial Intelligence, the Cognitive Science Society, and the American Educational Research Association. He served as a founding editor of the journal *Cognitive Science* and as first chair of the Cognitive Science Society. He is best known in psychology for his work on semantic memory and mental models, in artificial intelligence for his work on plausible reasoning and intelligent tutoring systems, and in education for his work on inquiry teaching, cognitive apprenticeship, situated learning, design research, epistemic games, and systemic validity in educational testing. From 1991 to 1994 he was co-director of the US Department of Education's Center for Technology in Education centered at Bank Street College of Education.