4 Experimenting to Support and Understand Learning Processes

Paul Cobb
Vanderbilt University

Koeno Gravemeijer
Freudenthal Institute

Introduction
In this chapter, we describe an approach to design research that we have refined while conducting a series of design research projects in mathematics education over a ten-year period. Our intent in doing so is to highlight a number of issues that we believe it is essential to consider when conducting a design experiment regardless of the specific approach followed. For the purpose of this chapter, we define design research as a family of methodological approaches in which instructional design and research are interdependent. On the one hand, the design of learning environments serves as the context for research, and, on the other hand, ongoing and retrospective analyses are conducted in order to inform the improvement of the design. This type of research involves attempting to support the development of particular forms of learning and studying the learning that occurs in these designed settings. The learning of interest might be that of individual students who interact one-on-one with a researcher in a series of teaching sessions (Cobb & Steffe, 1983; Steffe & Thompson, 2000), a group of students in a classroom (Cobb, 2000a; Confrey & Lachance, 2000; Gravemeijer, 1994b), preservice teachers in a university course (Simon, 2000), or practicing teachers who collaborate with researchers as members of a professional teaching community (Kazemi & Franke, 2004; Stein et al., 1998). In each of these cases, design research enables us to investigate simultaneously both the process of learning and the means by which it is supported and organized. As we will argue later in the chapter, the potential contributions of the methodology become particularly apparent when the current research base is thin and provides only limited guidance for the design of learning environments.

We focus specifically on design experiments in classrooms in which a research team assumes responsibility for a group of students’ learning both because they are the most common type of design research and because most of our work has involved experiments in classrooms. However, the methodological issues on which we focus can be extended to experiments that attempt to support the learning of individual students and of preservice and practicing teachers. We introduce these issues by discussing the three phases of conducting a design experiment: preparing for the experiment, experimenting to support learning, and conducting retrospective analyses of the data generated during the course of the experiment. To ground the discussion, we draw on two classroom design experiments that focused on statistical data analysis as illustrative cases. The first experiment was conducted in a seventh-grade classroom and involved the analysis of univariate data: the second experiment conducted with some of the same students the following school year was focused on the analysis of bivariate data.
Preparing for the Experiment

The preparation phase for an experiment encompasses a range of issues that include clarifying the instructional goals to which the experiment aims and documenting the instructional starting points. Against this background, the immediate challenge is to delineate an envisioned learning trajectory that consists of conjectures about both a learning process that culminates with the prospective instructional goals and the specific means of supporting that learning process. Finally, the preparation phase also involves placing the planned experiment in a broader theoretical context by framing it as a case of a broader class of phenomena.

Clarifying the Instructional Goals

In our view, it is critical to scrutinize the current instructional goals in the mathematical domain of interest when preparing for a design experiment (Cobb, 2001; Gravemeijer, 1994a; Lehrer & Lesh, 2003). It is important to acknowledge that these goals have become institutionalized in the curriculum over an extended period of time and are the product of history, tradition, and assessment practices. Rather than accepting them unquestioned, we attempt to problematize the particular mathematical problem under consideration from a disciplinary perspective by identifying the central organizing ideas. As an illustration, the current goals for statistics instruction at the middle-school level typically consist of a relatively long list of separate topics such as mean, mode, and median, and conventions for drawing different types of graphs (e.g., histograms, box plots). The extensive literature review that we conducted when preparing for the seventh-grade design experiment revealed that there was a lack of consensus on what the central ideas of middle-school statistics should be (McGatha, 2000). However, by both synthesizing this literature and examining what statistical data analysis entails, we came to the conclusion that the notion of distribution plays a central role and could serve as an overarching idea in statistics instruction from elementary school through college. From this perspective, notions like “center,” “skewness,” “spread,” and “relative frequency” become ways of characterizing how the data are distributed, rather than separate topics or concepts that are to be mastered in isolation from each other. Further, various types of statistical graphs come to the fore as different ways of organizing data in order to detect relevant patterns and trends.

This illustration emphasizes that our intent when conducting a design experiment is not merely to develop more effective instructional approaches for addressing traditional instructional goals but also to influence what the goals could be by demonstrating what is possible for students’ mathematical learning. Consequently, the approach that we take to design research is interventionist in character. In the case of statistics, part of our agenda was to influence what the instructional goals might be by demonstrating what is possible for students’ learning of statistics.

Documenting the Instructional Starting Points

The intent in documenting the instructional starting points is to identify both the aspects of students’ current reasoning on which the designer might build and the initial instructional challenges involved in doing so. However, it is important to stress that we view ourselves not as documenting the level of reasoning that is typical of students of a particular age but as documenting the consequences of their prior instructional histories. The existing research literature can be useful, particularly interview studies
conducted from a psychological perspective, because they can give insights into what students typically learn in a particular mathematical domain in the context of standard instruction. Also, it is often necessary to conduct additional assessments when preparing for an experiment because the tasks used in many of the existing studies reflect the way in which the domain is institutionalized in school. Because such assessments involve the development of novel tasks, video-recorded interviews and whole-class performance assessments prove initially to be more effective than written instruments and can lead to the development of such instruments (Desimone & Le Floch, 2004).

In preparing for the two design experiments in teaching statistics, we relied primarily on whole-class performance assessments conducted with two, intact, seventh-grade classes. In conducting these assessments, the research team member who served as the teacher did not attempt to support the students’ learning but, instead, probed the students’ reasoning in order to understand why they used particular approaches. The tasks required the students either to analyze data to inform a pragmatic decision or to develop a way of representing data that would be useful for a particular audience. The whole-class assessments revealed that, for most of the students, data analysis involved trying to remember what they were supposed to do with the numbers (Bright et al., 2003; Garfield, 1988; McGatha et al., 2002; Shaughnessy, 1992). This indicated that, for these students, data were not numbers plus context, to use Moore’s (1997) pithy phrase. In other words, data were not measures of an attribute of a situation that was relevant for the problem or issue under investigation, and data analysis was not pertinent to gaining insights into that problem or issue. Instead, for these students, data were merely numerical values on which they either performed particular calculational procedures or graphed by following particular representational conventions. Therefore, our initial challenge in the design experiments was to support a change in what statistics meant for these students, so that they were analyzing data.

Delineating an Envisioned Learning Trajectory

When the instructional goals and starting points have been clarified, the next phase in preparing for a design experiment is to specify an envisioned or a hypothetical learning trajectory (Simon, 1995). In doing so, the research team formulates testable conjectures about both significant shifts in students’ reasoning and the means of supporting and organizing these shifts. Typically, these means of support include those considered by materials developers such as instructional tasks and associated resources (e.g., physical and computer-based tools). We also believe that it is essential to envision how tasks and tools might be enacted in the classroom. Therefore, additional but frequently overlooked means of support include the nature of classroom norms and the nature of classroom discourse. For example, we know that the sociomathematical norm of what counts as an acceptable mathematical argument can differ radically from one classroom to another and that this can make a profound difference in the nature and the quality of students’ mathematical learning even when the same tasks and resources are used (Hershkowitz & Schwarz, 1999; Lampert, 2001; McClain & Cobb, 2001; Simon & Blume, 1996; Voigt, 1996; Yackel & Cobb, 1996). The establishment of this and other norms, the ways in which tasks and tools are enacted in the classroom, and, indeed, the learning opportunities that arise for students depend crucially on the proactive role of the teacher. As a consequence, we do not view the instructional tasks and tools that we develop as supporting students’ mathematical learning directly but, instead, consider the teacher to be a codesigner of the classroom’s learning environment that constitutes the immediate social situation of their students’ mathematical development. Given the central
mediating role that we attribute to teachers, our immediate goal is to design tasks and tools that they can use as resources to support shifts in their students’ reasoning.

In many domains, the research literature provides only limited guidance when formulating a hypothetical learning trajectory. When preparing for the statistics experiments, for example, we were able to identify only six relevant articles (Biehler & Steinbring, 1991; deLange et al., 1993; Hancock et al., 1992; Konold et al., 1997; Lehrer & Romberg, 1996; Wilensky, 1996). The types of articles that proved helpful focus on instructional goals that are at least partially compatible with those delineated for the planned experiment and report both the process of students’ learning and the instructional settings, the tasks, and the tools that enabled or supported that learning. Given this limited research base, we considered the conjectures in our envisioned learning trajectory to be extremely speculative and anticipated that many of them would be refuted once we began experimenting in the classroom. The process of formulating the trajectory was nonetheless valuable in that it made it possible for us to improve our initial design by testing and revising the conjectures when we began experimenting in the classroom. Design research is therefore a bootstrapping methodology that is useful both when formulating and testing initial designs in domains where the current research base is thin and when adapting and improving already successful designs that are grounded in an established research base.

As an illustration, we focus on the first part of the envisioned trajectory that we formulated when preparing for the two experiments. As noted earlier, our initial challenge was to support a shift in the students’ reasoning so that they were analyzing data rather than merely manipulating numbers or following graphical conventions. Because of time constraints, it was not feasible for the students to collect the data that they were to analyze. Nevertheless, we conjectured that it would be essential for them to experience the process of generating data for the purpose of answering a pragmatic question if the data were to mean measures rather than mere numbers for them (Bakker, 2004; Lehrer & Romberg, 1996; Roth, 1996). We further conjectured that this could be accomplished if the teacher introduced each instructional task by reviewing the data generation process with the students. In one of the first tasks, the students were to compare data on the life spans of two brands of batteries. First, the teacher clarified the significance of the issue that the students were to investigate by asking them if they used batteries; most of the students indicated that they did (e.g., in portable CD players, tape-recorders, and so forth). Next, the teacher delineated the relevant attributes of batteries that should be measured by asking the students what factors they consider when buying batteries (e.g., cost, life span). Then, the teacher shifted the discussion to issues of measurement by asking the students how it would be possible to determine which of two brands of batteries lasted longer (e.g., by putting a number of batteries of each brand in identical appliances such as flashlights or clocks). Against the background of this discussion, the teacher introduced the data that the students were to analyze inscribed in the first of three computer tools that the students used over the course of the two design experiments, as shown in Figure 4.1. The life spans of ten batteries of two brands are each represented as a horizontal bar.

In designing this tool, we purposely chose situations that involved linearity (e.g., time) that would fit with this type of representation. We conjectured that this would enable the students to interpret the bars as signifying data values; thus they would use the options available on the computer tool to analyze data. Our conjectures about the means of supporting a shift in the students’ reasoning proved to be well founded. There were clear indications that, within the first week of the first experiment, doing statistics came to involve analyzing data (Cobb, 1999; McClain et al., 2000).
Placing the Experiment in a Theoretical Context

Ethically, design research has a strong pragmatic orientation in that any experiment involves supporting the learning of a particular group of people. However, the intent is not to develop a rich ethnographic account of such learning. Instead, the overriding goal is to produce knowledge that will be useful in providing guidance to others as they attempt to support learning processes (Brown, 1992; Cobb et al., 2003b; Collins, 1992; Design-Based Research Collaborative, 2003; Edelson, 2002; Gravemeijer, 1994a). Therefore, when preparing for an experiment, it is critical to frame it explicitly as a paradigmatic case of broader phenomena. For example, we initially viewed the two statistics experiments as a case of supporting middle-school students’ development of increasingly sophisticated forms of reasoning in a particular mathematical domain. This encompassed students learning about data generation (e.g., constructing representative samples, controlling extraneous variables) and developing and critiquing relatively sophisticated, data-based arguments (Cobb & Tzou, 2000; Cobb et al., 2003b). In addition, the experiments became a case in which a teacher became increasingly effective in supporting students’ learning by building on their mathematical reasoning (McClain, 2002). They also became a case of the design and use of tools to support students’ mathematical learning and, more generally, of semiotic processes in mathematical learning (Cobb, 2002; Gravemeijer, 1994b; Sfard, 2000b). Finally, we became aware during the experiments that the students were developing an interest in investigating real-world phenomena that they considered to be significant by analyzing data. As a consequence, it became a case of cultivating students’ mathematical interests, an issue that is related directly to teachers’ perennial concern about how they can motivate their students (Cobb & Hodge, 2002).

These illustrations do not exhaust the possibilities, of course. For example, a classroom experiment might be framed as a case of negotiating general classroom norms or sociomathematical norms; of orchestrating productive, whole-class discussions; and of supporting equity in students’ access to significant mathematical ideas. In addition, a series of experiments can be conducted and serve as the context for the development and refinement of interpretive frameworks that do useful work in generating, selecting,
and assessing design alternatives. Examples of such frameworks developed by design researchers include the theory of metarepresentational competence (diSessa, 1992, 2002), the theory of quantitative reasoning (Thompson & Thompson, 1996), the theory of actor-oriented abstraction (Lobato, 2003), and the design theory of realistic mathematics education (Gravemeijer, 1994a, 1999; Treffers, 1987). Such frameworks can function both as a source of guidance for instructional design and as interpretive structures for making sense of what is happening in the complex setting in which a design experiment is conducted (diSessa & Cobb, 2004).

We will return to this issue of framing an experiment as a case of a more encompassing phenomenon when we discuss the third phase in a design experiment—conducting retrospective analyses of the data generated during the course of the experiment. For the present, it suffices to emphasize that such framings are critical if the findings are to be potentially generalizable.

Experimenting to Support Learning

In our view, only after the preparation work has been completed, the instructional endpoints have been specified, the starting points have been documented, a conjectured instructional theory has been formulated, and the experiment has been located in a broader context should a research team begin experimenting to support a group of participants’ learning for an extended period of time. Because the term experiment may evoke associations with experimental or quasi-experimental research, it is important to clarify that the objective of the design experiment is not to demonstrate that the envisioned learning trajectory works. The primary goal is not even to assess whether it works; although the research team will do so, of course. Instead, when experimenting to support learning, the purpose is to improve the envisioned trajectory developed while preparing for the experiment by testing and revising conjectures about both the prospective learning process and the specific means of supporting it.

We begin our discussion of the process of experimenting to support learning by considering briefly the kinds of data that might be collected in the course of an experiment. Then, we address the need to explicate the frameworks used to interpret the participants’ activity, their learning, and the evolution of the learning environment in which they are situated. Next, we focus on the tightly integrated cycles of design and analysis that characterize design research. Finally, we clarify one of the primary products of a series of design experiments: a domain-specific, instructional theory.

Data Collection

Decisions about the types of data that need to be generated in the course of an experiment depend on the theoretical intent of the design experiment. These decisions are critical to the success of an experiment because when the researchers conduct retrospective analyses frequently, the data have to make it possible for them to address the broader theoretical issues of which the learning setting under investigation is a paradigmatic case. We will return to the issue of data generation when we discuss the argumentative grammar of design research that links data and analysis to final claims and assertions. For the present, we offer an illustration from the statistics design experiments. One of our broader goals in these experiments was to investigate the processes of supporting students’ development of increasingly sophisticated forms of data-based reasoning. Therefore, we needed to document the shifts in the students’ reasoning as well as the means by which these shifts were supported and organized in
the classroom. To this end, we conducted individual pre- and post-interviews with the students, video-recorded all classroom sessions, made copies of all of the students’ written work, and developed two independent sets of field notes. We also would have incorporated benchmark assessment items that focused on the central statistical idea of distribution had they been available. Generally, these data proved adequate for our purpose. However, in a prior experiment conducted in a first-grade classroom, we also investigated the relationship between the process of individual students’ learning and what might be termed the mathematical learning of the classroom community, as assessed by the evolution of mathematical practices in the classroom. In that case, we also found it essential to video-record the performance of target students during the individual and small-group activities in the lessons.

Because design research is a relatively new methodology, researchers often find that they have to develop new data generation techniques or instruments (Cobb et al., 2003a; Confrey & Lachance, 2000; Design-Based Research Collaborative, 2003; Drijvers, 2003; Lehrer & Schauble, 2004; Lobato, 2003). As an illustration, we have noted that the statistics experiments became a case of cultivating students’ mathematical interests. To investigate this issue, we had to document how the students understood both the general and the specifically mathematical obligations in the classroom and how they evaluated those obligations. A member of the research team conducted interviews with the students that focused on these concerns while the second of the two experiments was in progress. Although these interviews proved to be useful, our failure to conduct such interviews throughout the first experiment restricted the scope of our subsequent analyses. This example underscores the importance of thinking through the types of data that should be generated before experimenting in the classroom. Careful preparation of this type also ameliorates design researchers’ tendency to squander limited resources assembling vast collections of data, most of which are never analyzed (Dede, 2004; Kelly, 2004).

As we have indicated, when experimenting to support learning, the overall goal is to improve the learning trajectory envisioned by testing and revising conjectures about both the prospective learning process and the specific means of supporting it. This process of testing and revising conjectures constitutes the learning process of the research team (Gravemeijer, 1994b). In our view, it is critical to document this process by audio-recording all research group meetings and by compiling a log of ongoing interpretations, conjectures, decisions, and so forth (Edelson, 2002). This log and the audio-recordings are useful when conducting retrospective analyses because they make it possible to reconstruct the rationales for particular decisions about a design.

**Interpretive Frameworks**

In the process of experimenting to support learning, the research team makes ongoing interpretations of both the participants’ activity and the learning environment in which they are situated. These ongoing interpretations inform design and instructional decisions and thus shape the design effort profoundly. Unfortunately, design researchers often fail to articulate the key constructs of what they use when making these interpretations. This omission indicates strongly the status of design research as a fledgling methodology. Given the complexity and messiness of the settings in which design experiments typically are conducted, the ongoing interpretations are highly selective and involve implicit conjectures about the important aspects of the participants’ activity, the learning environment, and the relation between them. In our view, it is essential that design researchers make explicit the conjectures, suppositions, and assumptions.
that ground their interpretations so that they can be subjected to public debate and scrutiny. As Kelly (2004) observes, reports of design experiments will be dismissed as anecdotal by critics if they fail to differentiate between what is necessary and what is contingent in a design. We will return to this distinction when we discuss the process of conducting retrospective analyses of the data generated in the course of an experiment. For the present, it suffices to note that researchers inevitably make implicit claims about what is necessary in the process of interpreting the evolving learning environment and the participants’ developing activity.

The typical characterization of design research settings as complex and messy emphasizes further the importance of articulating, critiquing, and refining interpretive frameworks. These settings seem complex and messy because we have yet to develop adequate ways of understanding them and have difficulty in perceiving pattern and order. We have noted already that a series of experiments can develop and refine the interpretive frameworks that can guide the generation, selection, and assessment of design alternatives. For example, diSessa’s (2002) theory of metarepresentational competence and Thompson’s (1996) theory of quantitative reasoning both posit and account for previously unarticulated aspects of mathematical learning. In our own work in a series of classroom experiments, we attempted to develop an interpretive framework that enables us to account for students’ mathematical learning as it occurs in the social situation of the classroom (Cobb & Yackel, 1996). Our intent in doing so was to begin to see some order in the complexity and messiness of the classroom.

The details of the framework that we proposed do not concern us here. However, it is worth emphasizing that the framework emerged over a period of several years while we attempted to understand specific events in the classrooms in which we worked. On the one hand, the framework grew out of our efforts to support students’ mathematical learning. On the other hand, interpretations of classroom events organized in terms of the emerging framework fed back to inform the ongoing, instructional development effort. A central feature of this process is that the framework evolved in response to problems and issues encountered while experimenting to support learning. The frameworks proposed by the other researchers that we have referenced also were developed by means of a similar process. In each case, the proposed interpretive framework does not stand apart from the practice of experimenting to support learning but, instead, remains grounded in it. Therefore, each framework makes public a particular way of conceptualizing the learning process being supported. Furthermore, each one explicates suppositions and assumptions about what it is worth attempting to document when generating data. As a consequence, the frameworks guide the process of thinking through the types of data that should be generated.

Cycles of Design and Analysis

In focusing on the logistical issues involved in experimenting to support learning, we stress again that the overall goal is to test and improve the learning trajectory formulated during the preparation phase. Therefore, it is counterproductive at the outset to plan in finished form the means that might be used to support learning because, in all probability, they will be changed as the conjectures are revised. In the statistics experiments, for example, we outlined the types of instructional tasks that we anticipated using by developing the specific activities only a day or two in advance, as informed by our current conjectures. Therefore, each instructional task therefore embodied specific conjectures about the students’ future learning at that particular point in the instructional sequence. As a part of the process of testing and revising these conjectures, we
found it essential to have short debriefing meetings after each classroom session. In these meetings, members of the research team shared and debated their interpretations of events that had occurred in their classroom. When we reached consensus in our ongoing interpretations, we prepared for subsequent sessions by discussing our conjectures about possible developments in the students’ reasoning. It was in the context of this discussion that we designed specific instructional tasks and considered other means of support (e.g., the renegotiation of specifically mathematical norms). We call these daily cycles of design and analysis “design minicycles.”

Critics of design research have argued that this process of testing and revising conjectures based on single cases without controls is an inherent weakness of the methodology. Therefore, it is helpful to distinguish between two complementary treatments of causal explanation: the regularity type of causal description that is based on observed regularities across a number of cases and a process-oriented explanation “that sees causality as fundamentally referring to the actual causal mechanisms and processes that are involved in particular events and situations” (Maxwell, 2004: 4). Thus, process-oriented explanations are concerned with “the mechanisms through which and the conditions under which that causal relationship holds” (Shadish et al., 2002: 9, cited in Maxwell, 2004: 4). In contrast to the regularity conception of causality, in principle, viable explanations of this type can be discerned based on a single case (Maxwell, 2004), particularly if the research team is using a well-established, interpretive framework that has been honed during a series of prior experiments. When making ongoing interpretations and when conducting retrospective analyses, the intent is to develop explanations of this type that center on the relation between the learning processes and the means by which they are supported.

In addition to holding daily debriefing meetings during the statistics experiments, we found it valuable to have longer research team meetings each week in which we took stock of the continuous process of testing and revising conjectures in the classroom. In these meetings, we first clarified the overall goals of the design experiment in order to locate our ongoing work within the broader context of the entire experiment. Next, we outlined a revised learning trajectory for the entire experiment that took account of the adaptations we had made to the conjectures while experimenting in the classroom. It was only against the background of this broader discussion that we articulated possible learning goals and means of support for future classroom sessions. Our purpose in structuring the meetings in this way was to ensure that the relationship between the envisioned learning trajectory and the ongoing testing and revising of the conjectures was truly reflexive. On the one hand, the local design decisions that we made on a daily basis were guided by the envisioned learning trajectory. On the other hand, the envisioned learning trajectory evolved as a consequence of local interpretations and judgments that, ideally, should be grounded in a clearly articulated, interpretive framework.

In our view, organizing the research team’s activity so that there is a reflexive relationship between local judgments and the entire perspective should be a basic tenet of design research. Simon (1995) addresses this issue when reporting a design experiment in which he served as the teacher and attempted to support the mathematical learning of a group of preservice teachers. He clarifies that he had a pedagogical agenda and thus a sense of direction at any point in the experiment, but that this agenda is subject to continual modification in the act of teaching. Simon likens this process to that of undertaking a long journey such as sailing around the world.

You may initially plan the whole journey or only part of it. You set out sailing according to your plan. However, you must constantly adjust because of the conditions
that you encounter. You continue to acquire knowledge about sailing, about the current conditions, and about the areas that you wish to visit. You change your plans with respect to the order of your destinations. You modify the length and nature of your visits as a result of interactions with people along the way. You add destinations that prior to the trip were unknown to you. The path that you travel is your [actual] trajectory. The path that you anticipate at any point is your “hypothetical trajectory” (pp. 136–137).

As Simon emphasizes, this way of experimenting to support learning involves both a sense of purpose and an open-handed flexibility toward the participants’ activity and learning. It also brings the learning of the research team to the fore. The deviation of the actual learning trajectory from the learning trajectory envisioned at the outset provides a general summative record of the research team’s learning while experimenting to support the participants’ learning.

Thus far, we have considered both design and analysis minicycles and the relation between these minicycles and the encompassing learning trajectory. Stepping back still further, an entire experiment can be viewed as a single cycle of design and analysis when it is located within a series of experiments. In this macrocycle, the envisioned learning trajectory formulated when preparing for an experiment is tested and revised while experimenting in the classroom and while conducting retrospective analyses, resulting in a revised learning trajectory that can serve as the basis for a subsequent experiment. In most cases, when conducting a series of experiments and enacting a sequence of design and analysis macrocycles, a primary goal is to develop a domain-specific, instructional theory.

Domain-Specific, Instructional Theories

The products of a series of design experiments typically include sequences of activities and associated resources for supporting a particular form of learning, together with a domain-specific, instructional theory that underpins the instructional sequences and constitutes its rationale. A domain-specific, instructional theory consists of a substantiated learning process that culminates with the achievement of significant learning goals as well as the demonstrated means of supporting that learning process. We call such theories domain-specific to emphasize that their scope is restricted to significant learning goals in a particular domain (e.g., students’ development of sophisticated forms of reasoning in a specific mathematical or scientific domain, or mathematics or science teachers’ development of particular forms of instructional practice in particular content domains). Elsewhere, theories of this type have been called humble theories to acknowledge their domain specificity (Cobb et al., 2003a).

A domain-specific, instructional theory is useful because it enables other researchers to customize the sequence of activities and resources produced during a series of experiments to the setting in which they are working. If the activities and resources were justified solely with traditional experimental data, other researchers would know only that they had proved effective elsewhere but would not have an understanding of the underlying rationale that would enable them to adapt them. In contrast, the justification provided by a domain-specific, instructional theory offers the possibility that other researchers will be able to adapt, test, and modify the activities and resources as they work in different settings. In doing so, they can contribute to both the improvement of the activities and the development of the domain-specific, instructional theory, thereby making the production of design-based knowledge a cumulative activity. As an example, Bakker (2004) carried out a series of classroom experiments that focused on statistics at
the middle-school level in the course of which he tested, modified, and elaborated on the learning trajectory that had resulted from the two statistics design experiments we had conducted. Clearly, this process of building on and extending the findings of prior design experiments depends crucially on researchers distinguishing explicitly between what is necessary and what is contingent in their designs.

We conclude this discussion of experimenting to support learning by giving an overview of the learning trajectory that resulted from the two design experiments in statistics. In the context of the current discussion, this trajectory can be viewed as a prospective, domain-specific theory that could be elaborated in subsequent experiments, such as those conducted by Bakker (ibid.). Our intention in outlining this trajectory is to illustrate the level of specificity that we contend is essential when developing and revising designs for supporting learning. To make the presentation tractable, we focus on the three computer tools that the students used to analyze data with the understanding that they were but one of the means by which the students’ learning was supported. Thus, we omit a discussion of both the classroom norms and the role of the teacher. As background, we should clarify that the enactment of an instructional task often spanned two or more, 40-minute, classroom sessions and involved three phases: (a) a whole-class discussion of the data generation process, (b) an individual or a small-group activity in which the students typically worked at computers to analyze data, and (c) a whole-class discussion of the students’ analyses.

In describing the first of the three computer tools (see Figure 4.1), we clarified its role in supporting an initial shift in the students’ activity so that they were analyzing data. The students’ use of this tool also proved to be necessary for a second reason. The inscription of data values as horizontal bars and the options of organizing data oriented the students to compare collections of data values in terms of how they are spread out and bunched up, a precursor to distribution. For example, the students dragged a vertical value bar along the axis either to partition data sets or to find the value of specific data points. In addition, they used another option on the tool to isolate a particular interval and compare the number of data points in each data set that were in that interval (see Figure 4.2).

Against this background, the teacher introduced the second computer tool in which data points were inscribed as dots in a line plot (see Figure 4.3). As can be seen by comparing Figures 4.2 and 4.3, the dots at the end of the bars in the first tool have been collapsed down onto the horizontal axis in the second tool. The teacher introduced this new way of inscribing data first by showing a data set inscribed as horizontal bars, then by removing the bars to leave only the dots, and finally by sliding the dots down onto the horizontal axis. As we had conjectured, the students were able to use the second computer tool to analyze data with little additional guidance, and it was apparent that the line plot inscription signified a set of data values rather than merely a collection of dots scattered along a line. However, this development cannot be explained solely by the teacher’s careful introduction of the new tool. Instead, we also have to take account of a subtle but important aspect of the students’ learning when they used the first tool. The crucial point to note is that, in using the options on the first tool to partition data sets and to isolate the data points within a particular interval, the students focused on the location of the dots at the end of the bars with respect to the horizontal axis. In other words, a necessary shift occurred as the students used the first tool. Originally, the individual data values were represented by the lengths of the bars. However, in using the tool, these values came to be signified by the endpoints of the bars. As a result of this development, when they were presented with the second tool, they could understand readily the teacher’s explanation of collapsing the dots at the end of the bars down onto the axis.
The data shown in Figure 4.2 come from a task in which the students used the second computer tool to compare the effectiveness of two treatments for AIDS patients. The task was to assess whether a new experimental protocol in which 46 people had enrolled was more successful in raising patients’ T-cell counts than a standard protocol in which 186 people had enrolled. The options on this tool involved partitioning sets of up to 400 data values in various ways. For example, students could drag vertical bars along the axis in order to partition a data set into groups of points, groups with a specified interval width (i.e., a precursor of a histogram), and four groups that each contain the same number of data points (i.e., a precursor of a box plot). The students could use the second tool immediately because they had partitioned data sets routinely when they used the first tool. One of the more elementary analyses that the students produced involved partitioning the two data sets and the T-cell count of 550 (see Figure 4.4).

Figure 4.2 The Battery Data in the First Computer with the Options of Bounding an Interval and Partitioning at a Data Value Shown.

The data shown in Figure 4.2 come from a task in which the students used the second computer tool to compare the effectiveness of two treatments for AIDS patients. The task was to assess whether a new experimental protocol in which 46 people had enrolled was more successful in raising patients’ T-cell counts than a standard protocol in which 186 people had enrolled. The options on this tool involved partitioning sets of up to 400 data values in various ways. For example, students could drag vertical bars along the axis in order to partition a data set into groups of points, groups with a specified interval width (i.e., a precursor of a histogram), and four groups that each contain the same number of data points (i.e., a precursor of a box plot). The students could use the second tool immediately because they had partitioned data sets routinely when they used the first tool. One of the more elementary analyses that the students produced involved partitioning the two data sets and the T-cell count of 550 (see Figure 4.4).

Figure 4.3 Data on the T-Cell Counts of Two Groups of AIDS Patients in the Second Computer Tool.
They concluded that the experimental treatment was more effective because the majority of these data were above 550, whereas the majority of the standard treatment data were below it. In contrast, the most sophisticated type of analysis involved partitioning both data sets into four groups, each of which contained 25 percent of the data, and hiding the dots that showed the location of each individual data value (see Figure 4.5).

Figure 4.4 The AIDS Data Partitioned at the T-Cell Count of 550.

Figure 4.5 The AIDS Data Organized into Four Equal Groups with Data Hidden.
The students who produced this analysis argued that the experimental treatment was more effective because 75 percent of these data but only 25 percent of the standard treatment data are above the T-cell value of 550. In both solutions, the students compared the data sets in terms of how they were distributed by focusing on the proportion or relative frequency of the data points in particular intervals.

There was no regression in the sophistication of the analyses that the students produced between the end of the first experiment and the beginning of the second experiment despite the nine-month time lag. Further, within the first week of the second experiment, all the students were able to read the shape of a data set from a four-equal-groups display such as that shown in Figure 4.5 (i.e., they realized that the data were more bunched up or were distributed more densely when the bars were closer together and the interval was smaller). This development was necessary before the students could use the third computer tool productively to analyze bivariate data. In one instructional task, the students used the third tool to investigate possible inequities in the salaries of men and women who had the same number of years of education. Figure 4.6 shows the data for 300 women, 50 at each of the six education levels, inscribed in the third computer tool. It proved necessary to use displays of this type in which it was impossible to see all the data points because some of the dots were in the same location. One of the options on this tool allowed students to superimpose a frequency grid on the display and to hide the dots, as shown in Figure 4.7.

This option proved useful to the students because they could read how the data for each education level were distributed. For example, they could read how the salaries of the 50 women with 18 years of education were distributed from the last column of numbers in Figure 4.7 and noticed that they were skewed heavily toward the lower end of the range.

A second option on the tool allowed the students to partition the data at each of the six education levels into four groups, each of which contained the same number of

![Figure 4.6](image-url)
Figure 4.7 Salary and Education Data for Women Organized the Grids Option.

<table>
<thead>
<tr>
<th>Years of education</th>
<th>Annual salary (x 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>14</td>
<td>45</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>16</td>
<td>55</td>
</tr>
<tr>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>18</td>
<td>65</td>
</tr>
<tr>
<td>19</td>
<td>70</td>
</tr>
<tr>
<td>20</td>
<td>75</td>
</tr>
</tbody>
</table>

Figure 4.8 Salary and Education Data for Women Organized the Grids Option Four-Equal-Groups Option.
data points (see Figure 4.8). In essence, the four-equal-groups option on the second computer tool (see Figure 4.5) has been rotated 90 degrees and superimposed on each of the six “data stacks.” As a consequence of their activity with the second tool, the students could read easily how the women’s salaries at each of the six education levels were distributed. In particular, they could see again that although the women’s salaries increased with their years of education, the salaries at each education level were skewed toward the lower end of the range. When compared with the men’s salaries, the degree to which the women’s salaries were skewed indicated a possible inequity.

In addition to illustrating the level of detail that we contend is necessary when formulating, testing, and revising instructional designs, this overview of the learning trajectory resulting from the two experiments exemplifies what it means to problematize a particular domain. Typically, scatter plots are viewed as two-dimensional inscriptions that show how two sets of measures covary in a relatively straightforward manner. However, proficient statistical analysts’ imagery of covariation is no more two-dimensional than their imagery of univariate distributions is one-dimensional. This is clearer in the case of univariate data in that inscriptions such as line plots involve, for the proficient user, an additional second dimension that indicates relative frequency (see Figure 4.5). In the case of bivariate data, however, scatter plots do not provide such direct perceptual support for a third dimension corresponding to relative frequency. Instead, it appears that proficient analysts read this third dimension from the relative density of the data points. The third computer tool was designed to enable students to learn to read this missing dimension into scatterplots. This made it possible for them to analyze how bivariate data were distributed by investigating how the distribution of a set of measures (e.g., salary) changed as the second measure (e.g., years of education) increased.

Conducting Retrospective Analyses

Thus far, we have discussed the planning of a design experiment and the process of experimenting to support learning, which are central to design research. A further aspect of the methodology concerns the retrospective analyses that are conducted of the entire data set collected during the experiment. The ongoing analyses conducted while the experiment is in process usually relate directly to the goal of supporting the participants’ learning. In contrast, retrospective analyses seek to place this learning and the means by which it was supported in a broader theoretical context by framing it as a paradigmatic case of a more encompassing phenomenon. As a consequence, the goals of the retrospective analyses depend on the theoretical intent of the design experiment. For easy explication, we assume that one of the primary goals of a design experiment is to contribute to the development of a domain-specific, instructional theory. The issues that we discuss concern the argumentative grammar of design research analyses and the trustworthiness, repeatability, and generalizability of the findings.

Argumentative Grammar

As Kelly (2004) observes, methodologies are underpinned by distinct schemes of argumentation that link data to analysis and to final claims and assertions. He uses the term argumentative grammar to refer to the scheme of argumentation that characterizes a particular methodology. As he clarifies, an argumentative grammar comprises “the logic that guides the use of a method and that supports reasoning about its data” (p. 118). He goes on to note that “the argumentative grammar of randomized field trials can be described separately from its instantiation in any given study so that
the logic of a proposed study and its later claims can be criticized” (p. 118). This leads him to ask:

Where is the separable structure [in design research] that justifies collecting certain data and under what conditions? What guides the reasoning with these data to make a plausible argument? Until we can be clear about their argumentative grammar, design studies in education lack a basis for warrant for their claims.

(2004: 119)

Kelly pinpoints a weakness of design research that betrays its status as an emerging methodology. In the following paragraphs, we make an initial attempt to address this issue by restricting our focus to the development of domain-specific, instructional theories.

We begin by noting that design research differs from more established forms of research such as the randomized field trial in terms of the types of knowledge claims that the methodology aims to produce. In the case of a design experiment that concerns the development of a domain-specific, instructional theory, the goal is to develop an empirically grounded theory about both the process of students’ learning in that domain and the means by which this learning can be supported. In developing such a theory, it is essential for researchers to explicate what they conceive of as competence in the domain, together with their conjectures about the process of its development. We have indicated the importance of expressing as problems the goals of the instructional sequence and can speak similarly of expressing as problems the process of learning mathematics. We find Freudenthal’s (1973, 1983) analysis of mathematical learning to be valuable in this regard. Freudenthal argued that, in his view as a professional mathematician, mathematics is first and foremost an activity that he termed mathematizing and described as organizing subject matter (Freudenthal, 1971). Freudenthal’s conception, as it has been elaborated in the domain-specific, instructional theory of realistic mathematics education (Gravemeijer, 1994b; Treffers, 1987), offers points of reference for assessing mathematical learning processes in particular domains.

As far as the goals of the instructional sequence are concerned, it is clearly essential to determine the particular forms of domain-specific reasoning that indicate the students’ development of competence and to demonstrate that they would not have developed these forms of reasoning but for their participation in the design experiment. This obviously indicates the importance of using sound procedures for assessing the participants’ reasoning, especially at the conclusion of the experiment. Assuming that such procedures have been employed, the logic of argumentation is typically straightforward if the domain under investigation was problematized while preparing the experiment. In such cases, researchers can draw on the literature to demonstrate that the documented forms of reasoning would not have emerged by themselves. In the statistics experiments, for example, there were no reports in the research literature of either middle- or high-school students reasoning about either univariate or bivariate data in the relatively sophisticated ways that we documented. In experiments where the problem in the domain has not been described, it is essential to conducted postexperiment comparison assessments with nonparticipants. A researcher who served as the project evaluator conducted assessment interviews with participating and nonparticipating students at the conclusion of the two statistics experiments, in the process documenting significant differences in the ways that the two groups of students reasoned about the data (Konold et al., 2005).

The demonstration that participants developed relatively sophisticated forms of
reasoning is only the first element of an argumentative grammar for design research. To clarify the additional elements, we draw on Brown’s seminal article on design research:

The Hawthorne Effect, as presented in standard texts, refers to the fact that any intervention tends to have positive effects merely because of the attention of the experimental team to subjects’ welfare. The infamous Hawthorne Effect has been dogging my tail for a long time. . . . Everywhere I go I can predict that someone will tell me that my results are just a Hawthorne Effect.

(1992: 163)

Brown goes on to explain why she rejects such arguments:

I have never taken the Hawthorne criticism of my work seriously because of the very specific nature of the improvements obtained [in participants’ reasoning]. If I were creating a true Hawthorne Effect, I would not be able to predict which performance would improve. But in fact we can see a close coupling of the cognitive activities practiced and the improvements shown.

(pp. 163–164)

Brown’s observations further emphasize that the argumentative grammar of design research has to encompass the process by which the research team purposefully supported the participants’ development of particular forms of reasoning and not others. This gives rise to two challenges.

The first challenge is to document the major shifts that occurred in the participants’ reasoning in the course of the experiment. This requirement has obvious implications for the data collection procedures employed in that the goal is to document how each successive form of reasoning emerged as a reorganization of prior forms of reasoning. Explanations of this type are central to the argumentative grammar of design research in our view. Clearly, the use of an explicitly articulated, interpretive framework for understanding participants’ learning is crucial when developing such explanations. As an illustration, in presenting the learning trajectory that resulted from the two statistics design experiments, we explained how the students’ learning while using the first computer tool enabled them to use the second computer tool as soon as it was introduced. In particular, we described how, in using the first tool, the students learned to interpret the dots and the end of the bars as signifying data values. This analysis of the students’ learning was oriented by a distributed view of cognition (Dörfler, 1993; Kaput, 1994; Meira, 1998; Pea, 1993), which enabled us to tease out how the students’ use of the computer tools influenced the nature of their data analysis activity and thus their learning during the experiments. Elsewhere, we have described the interpretive framework that guided our analysis of the students’ learning during these experiments (Cobb, 2000b).

The second challenge concerns the close coupling to which Brown (1992) refers between the process of the participants’ learning and the means by which the emergence of successive forms of reasoning was supported and organized. This indicates the importance of being explicit about how the learning environment and its relationship to the participants’ activity and learning are conceptualized. In experimental and quasi-experimental studies, for example, the learning environment is conceptualized typically as consisting of independent features that the investigator can manipulate and control directly. The implicit ontology is that of environmental settings made up of separate independent variables and of students composed of collections of dependent psychological attributes. Together, these two theoretical suppositions ground
the investigations that seek to discern regularity types of causal relations between the manipulation of instructional conditions and the performance of a collective, statistically constructed subject. In our view, this conceptualization is highly appropriate when the goal is to produce a particular form of knowledge that concerns both the overall effectiveness of two or more instructional interventions and the conditions under which one is more effective than the other (cf. Slavin, 2004).

Design research aims to produce a different form of knowledge that involves creating and improving means of supporting learning and understanding how they work. In the context of this research enterprise, the conceptualization of the learning environment as composed of externally manipulable, independent variables proves inadequate. In the case of the two statistics experiments, for example, we noted that the means of support included both the classroom norms and the nature of the classroom discourse. Classroom norms and discourse cannot be manipulated directly by a researcher; instead, they are constituted jointly by the teacher and the students in the course of their ongoing interactions. As a second example, it is reasonable for the purposes of experimental studies to view the tools used in the classroom as components of the learning environment (cf. Shavelson et al., 2003). However, this conceptualization is inadequate for the purposes of design experiments because the same tools can be used very differently in different classrooms, and these differences can influence students’ learning profoundly. When conducting a design experiment, it is essential not merely to document the presence or absence of particular tools, but also to understand how they are used. The ways in which tools are used are established jointly by the teacher and the students and depend on the tasks as they are realized in the classroom as well as on the classroom norms and discourse. For the purposes of design research, it is reasonable to conceptualize the classroom learning environment as an evolving ecology that does not exist independently of the teacher’s and the students’ activity but is constituted in the course of classroom interactions. The relationship between the learning environment so conceptualized and the students’ activity is therefore one of reflexivity in that the students contribute to the constitution of the learning environment that both enables and constrains their learning.

Given this conceptualization, the second challenge involves specifying the aspects of the classroom learning ecology that are necessary, rather than contingent, in supporting the emergence of successive forms of reasoning. As an illustration, in discussing why the students who participated in the statistics experiments were able to use the second computer tool productively as soon as it was introduced, we identified their use of the first tool as a primary means of support. The way in which the students used the first tool depended on:

- The overall goal for doing statistics established in the classroom (i.e., to identify patterns in data that are relevant to the question or issue under investigation).
- The organization of classroom activities (e.g., talking through the data-generation process).
- The instructional tasks (e.g., comparing two data sets by analyzing data that students viewed as realistic for a purpose that they considered legitimate).
- The nature of the classroom discourse (e.g., engaging in discussion in which significant statistical issues emerge as topics of conversation).

This indicates that each of these aspects of the classroom learning ecology were also necessary in making it possible for the students to use the second computer tool productively.
In summary, the argumentative grammar for design research that we have outlined involves:

- Demonstrating that the participants would not have developed particular forms of reasoning but for their participation in the design experiment.
- Documenting how each successive form of reasoning emerged as a reorganization of prior forms of reasoning.
- Specifying the aspects of the learning ecology that were necessary, rather than contingent, in supporting the emergence of these successive forms of reasoning.

This argumentative grammar is grounded in what Maxwell (2004) terms process-oriented explanations, rather than the regularity type of causal descriptions. Taken together, the three components of this grammar specify how a particular form of learning that would not have occurred naturally was “engineered.”

**Trustworthiness**

Trustworthiness is concerned with the reasonableness and justifiability of inferences and assertions that result from a retrospective analysis. This notion of trustworthiness acknowledges that a range of retrospective analyses might be made of a given data set for a variety of purposes. The issue is the credibility of an analysis. The most important considerations are the extent to which the analysis of the longitudinal data set generated during an experiment is systematic and is open to monitoring and critique by other researchers.

It is critical to analyze systematically the entire data set generated during a design experiment systematically while simultaneously documenting the grounds for particular inferences. Furthermore, all phases of the analysis should be documented, including the refining and refuting of inferences. Only then can final claims and assertions be justified by backtracking through the various levels of the analysis, if necessary to the original data. It is this documentation that provides an empirical grounding for the analysis. Further, it provides a means of differentiating systematic analyses, in which sample episodes are used to illustrate general assertions, from questionable analyses, in which a few, possibly atypical episodes are used to support unsubstantiated claims. Additional criteria that enhance the trustworthiness of an analysis include both the extent to which the analysis has been critiqued by independent researchers and the extent to which it derives from a prolonged engagement with participants (Taylor & Bogdan, 1984). Typically, this latter criterion is satisfied in the case of classroom design experiments and constitutes a strength of the methodology.

As an illustration, the data generated during the two statistics experiments included video-recorded pre- and postinterviews, video-recordings of all classroom sessions, copies of all of the students’ written work, field notes, and audio-recordings of all research team meetings. The specific approach that we used to analyze these data is a variant of Glaser and Strauss’ (1967) constant comparative method (see also Cobb and Whitenack, 1996). First, we worked through the data chronologically, episode by episode, at each point testing our current inferences against a subsequent episode. This first phase of the retrospective analysis produced a sequence of inferences and refutations that were tied to specific episodes. In the second phase of the analysis, this sequence of conjectures and refutations became the data. It was while “meta-analyzing” these episode-specific inferences, confirmations, and refutations that particular episodes came to be seen as pivotal. And they were pivotal in the context of the analysis because they allowed us to
decide between two or more competing inferences. It was these episodes that we typi-
cally included in research reports.

In addition to analyzing the entire data set systematically, it is also important to
explicate the criteria for using key constructs of the interpretive framework so that
other researchers can monitor and critique the process of making inferences. This is a
nontrivial task in our experience, in that it is one thing to use an explanatory construct
while engaging in the activity of making sense of data and quite another to frame that
sense-making activity as an object of reflection and tease out criteria that capture how a
particular construct is being used. As an illustration, when characterizing the learning
environment established in a particular classroom, one of the primary constructs that
we use is a classroom norm. Generally, explicating the criteria for using this construct
involves specifying the types of evidence used to determine that a norm has been estab-
lished. A first, relatively robust type of evidence occurs when a particular way of reason-
ning or acting that initially requires a justification is used later to justify other ways of
reasoning or acting (Stephan & Rasmussen, 2002). In such cases, the role of the way
of reasoning or acting shifts from a claim that requires a warrant to a warrant that
substantiates subsequent claims. This shift provides direct evidence that a particular
way of reasoning or acting has become normative and beyond justification. A second,
equally robust type of evidence is indicated by Sfard’s (2000a) observation that norma-
tive ways of acting are not mere arbitrary conventions that can be modified at will.
Instead, they are value-laden and are constituted in the classroom as legitimate or
acceptable ways of acting. This observation indicates the importance of searching
for instances where a student appears to violate a proposed classroom norm in order
to check whether his or her activity is treated as legitimate or illegitimate by the teacher
and other students. In the former case, it would be necessary to revise the conjecture,
whereas, in the latter case, the observation that the student’s activity was viewed as
a breach of a norm provides evidence in support of the conjecture (cf. Cobb et al.,
2001). Finally, a third and even more direct type of evidence occurs when the teacher
and students talk explicitly about their respective obligations and expectations. Such
exchanges typically occur when one or more members perceive that a norm has been
violated.

In summary, retrospective analyses are trustworthy to the extent that:

- The method of analysis is systematic and involves refuting conjectures (Atkinson
  et al., 1988).
- The criteria for making claims are explicit, thus enabling other researchers to moni-
tor the analyses.
- Final claims and assertions can be justified by backtracking through the various
  phases of the analysis, if necessary to the original data sources.
- The analyses have been critiqued by other researchers, some but not all of whom
  are familiar with the settings from which the data were collected.

Repeatability

Accounts of design researchers have sometimes emphasized that each learning setting is
unique and have stressed the importance of developing “thick descriptions” of these
settings. In doing so, they appear to eschew a concern for the repeatability of designs.
In our view, it is critical that design researchers aim to develop designs or innovations
that can be used to support learning productively in other settings. This view implies
that, when conducting a retrospective analysis, one of the goals is to delineate the
aspects of the learning process that potentially can be repeated in different settings. The argumentative grammar that we have outlined indicates the general form that the specification of the repeatable aspects of a design experiment might take:

- The development of particular culminating forms of reasoning.
- Successive shifts or reorganizations of reasoning that specify the process of development of the culminating form of reasoning.
- The aspects of the learning ecology that are necessary to support the emergence of these successive forms of reasoning.

It is important to clarify that a specification of this type does not imply that a design should be repeated by ensuring that it is realized in precisely the same way in different settings. Instead, the intent is to inform others as they customize the design to the settings in which they are working by differentiating between the necessary and the contingent aspects of the design. In the case of the two statistics design experiments, the characterization of repeatability as complete fidelity clashes with the view of teachers as professionals who adjust their plans continually on the basis of ongoing assessments of their students’ mathematical reasoning (cf. Ball, 1993). However, the learning trajectory that resulted from the experiments can guide teachers as they use the instructional activities and computer tools as the basis for instruction. A detailed version of this trajectory specifies both successive reorganizations in students’ statistical reasoning and the means that are necessary to support those reorganizations. This type of rationale offers the possibility that teachers who have reconstructed the learning trajectory in the context of their professional development will be able to differentiate between the necessary and the contingent aspects of the instructional sequence as they adapt and modify it in their classes.

**Generalizability**

In the context of design research, generalizability is related closely to repeatability and implies that others will be able to use the products of a design experiment to inform their efforts to support learning in other settings. We indicated the importance of generalizability when we emphasized the value of framing an experiment as a paradigmatic case of a broader class of phenomena. It is this framing of activities and events in the learning setting as exemplars or prototypes that gives rise to generalizability. However, this is not generalization by means of a representative sample that is based on the regularity type of causal descriptions. Instead, it is generalization by means of an explanatory framework that is based on process explanations of causality (Steffe & Thompson, 2000). Therefore, the achievement of this type of generalizability depends on the development of domain-specific, instructional theories whose structure corresponds to the argumentative grammar that we have outlined.

As a part of the process of constructing a robust, domain-specific, instructional theory, we have found it important to conduct follow-up trials with a range of participants in a variety of settings. These trials are not full-scale design experiments that aim to refine and improve the design but, instead, might involve customizing the design while working in a new setting. In the case of the statistics experiments, for example, we conducted follow-up trials in which we worked with middle-school students; at-risk high-school students; prospective elementary-school teachers; and practicing middle-school teachers. We were surprised by the extent to which we have been able to document regularities in the development of the participants’ statistical reasoning.
across these various settings. In each trial, there was considerable diversity in how the participants reasoned at any time. However, we were able to predict the primary forms of reasoning in each group of participants at any point in each trial. This type of knowledge is useful because it enables teachers to anticipate their students’ types of reasoning on which they can build in order to achieve their instructional agenda at each point in the instructional sequence.

Conclusions

In this chapter, we have discussed a range of methodological issues that arise when preparing for a design experiment, experimenting to support learning, and conducting retrospective analyses of the data generated in the course of the experiment. The preparation phase is crucial to the success of the experiment but can be extensive, especially when there is little prior research on which to build. Key issues that need to be addressed include clarifying instructional goals by identifying the central ideas in a particular domain and documenting the instructional starting points both by drawing on the relevant literature and by conducting initial assessments as a part of the pilot work. In addition, it is essential to delineate an envisioned learning trajectory that consists of conjectures about both a learning process that culminates with the prospective instructional goals and the specific means of supporting that learning process. Although some of these conjectures might be highly speculative, formulation of the learning trajectory envisioned enables the research team to engage in the process of testing and revising its initial design as soon as it begins experimenting to support learning. Finally, the preparation phase also involves locating the planned experiment in a broader theoretical context by framing it as a case of a broader class of phenomena, thereby indicating the level of generalizability to which the experiment aims.

We clarified that the objective during the second phase of experimenting to support learning is not to demonstrate that the envisioned learning trajectory works but to improve the trajectory by testing and revising conjectures about both the prospective learning process and the specific means of supporting it. We indicated the importance of thinking through the types of data that will make it possible for the researchers to address the theoretical issues identified at the outset when they conduct retrospective analyses. In addition, we noted that in the process of experimenting to support learning, the research team makes ongoing interpretations that shape the design effort profoundly. Therefore, it is essential that design researchers make explicit the suppositions and assumptions that ground these interpretations, in the process explicating how they differentiate between what is necessary and what is contingent in a design. We also drew on the work of Maxwell (2004) to clarify that the intent both when making ongoing interpretations while experimenting to support learning and when conducting retrospective analyses is to develop process-oriented explanations that, in principle, can be discerned on the basis of a single case. Finally, we focused on the tightly integrated cycles of design and analysis that characterize design research at the level of both minicycles that are enacted in the course of a single experiment and macrocycles that span a series of experiments. In doing so, we noted that one of the primary products of a series of design experiments is typically a domain-specific, instructional theory that consists of a substantiated learning process together with the demonstrated means of supporting that process.

In discussing the final phase of a design experiment, conducting retrospective analysis, we followed Kelly (2004) in observing that design research does not have a clearly articulated, argumentative grammar that can be specified separately from its instantiation in
any given experiment. We made an initial attempt to address this pressing concern by outlining three components of such a grammar that corresponds to the structure of a domain-specific, instructional theory. Then, we considered the trustworthiness or credibility of retrospective analyses and emphasized the importance of analyzing systematically the large, longitudinal, data sets generated in the course of a design experiment so that the final claims and assertions can be justified by backtracking through the various levels of an analysis, if necessary to the original data. We also stressed the importance of explicating the criteria for using key constructs of the interpretive framework so that other researchers can monitor and critique the process of making inferences. Next, we discussed the issue of repeatability and argued that what must repeat across settings are a specified learning process and the means that have been identified as necessary to support that process. The final issue discussed was that of generalizability. We noted that the type of generalization that design research seeks is based on process explanations of causality and indicated the value of conducting follow-up trials with a range of participants in a variety of settings.

Acknowledgments

The analysis reported in this chapter was supported by the National Science Foundation under Grant No. REC 0231037. The opinions expressed do not reflect necessarily the views of the Foundation. The authors are grateful to Qing Zhao and Jana Visnovska for their constructive comments on a previous draft of the chapter.

Notes

1 In earlier publications, we used the term developmental research to denote this type of research. In this chapter, we use the terms design research and design experiment because they have become the more generally accepted terms.
2 The basic tenets of design research that serve to differentiate it from other methodologies have been discussed in some detail by Cobb et al. (2003) and the Design-Based Research Collaborative (2003).
3 The term learning trajectory can be used to describe either the rationale for a limited number of classrooms lessons or the rationale for an extended instructional sequence on a given topic. Here, we use the term in the latter sense.
4 The design experiments were conducted by Paul Cobb, Kay McClain, Koeno Gravemeijer, Jose Cortina, Lynn Hodge, Maggie McGatha, Beth Petty, Carla Richards, and Michelle Stephan. Erna Yackel and Cliff Konold served as long-term consultants.
5 The computer tools were developed by Koeno Gravemeijer, Paul Cobb, Michiel Doorman, and Janet Bowers.
6 Edelson and Joseph (2004) also have framed design experiments as cases in which to investigate the process of supporting the development of students’ domain-specific interests.
7 The term domain-specific, as we use it in this context, refers to particular mathematical domains. This usage should be differentiated from Treffers’ (1987) employment of the same term to refer to the domain of realistic mathematics education.
8 These aspects of the design are reported by Cobb (1999), Cobb et al. (2003), and McClain (2002).

References


