Formative Research: A Methodology for Creating and Improving Design Theories

Charles M. Reigeluth Theodore W. Frick

Indiana University

In Chapter 1, Reigeluth described design theory as being different from descriptive theory in that it offers means to achieve goals. For an applied field like education, design theory is more useful and more easily applied than its descriptive counterpart, learning theory. But none of the 22 theories described in this book has yet been developed to a state of perfection; at very least they can all benefit from more detailed guidance for applying their methods to diverse situations. And more theories are sorely needed to provide guidance for additional kinds of learning and human development and for different kinds of situations, including the use of new information technologies as tools. This leads us to the important question, “What research methods are most helpful for creating and improving instructional design theories?” In this chapter, we offer a detailed description of one research methodology that holds much promise for generating the kind of knowledge that we believe is most useful to educators—a methodology that several theorists in this book have intuitively used to develop their theories.

We refer to this methodology as "formative research"—a kind of developmental research or action research that is intended to improve design theory for designing instructional practices or processes. Reigeluth (1989) and Romiszowski (1988) have recommended this approach to expand the knowledge base in instructional-design theory. Newman (1990) has suggested something similar for research on the organizational impact of computers in schools. And Greeno, Collins and Resnick (1996) have identified several groups of researchers who are conducting something similar that they call “design experiments,” in which “researchers and practitioners, particularly teachers, collaborate in the design, implementation, and analysis of changes in practice.” (p. 15) Formative research has also been used for generating knowledge in as broad an area as systemic change in education (Carr, 1993; Naugle, 1996).

We intend for this chapter to help guide educational researchers who are developing and refining instructional-design theories. Most researchers have not had the opportunity to learn formal research methodologies for developing design theories. Doctoral programs in universities tend to emphasize quantitative and qualitative research methodologies for creating *descriptive* knowledge of education. However, design theories are guidelines for practice, which tell us "how to do" education, not "what is."

We have found that traditional quantitative research methods (e.g., experiments, surveys, correlational analyses) are not particularly useful for improving instructional-design theory—especially in the early stages of development. Instead, we have drawn from formative evaluation and case-study research methodologies in the development of formative research methods. Researchers familiar with these qualitative methods should recognize them. However, they should keep in mind that the purpose is different here, and hence we must consider additional methodological concerns.

We first discuss three criteria for evaluating research which aims to create generalizable design knowledge: effectiveness, efficiency and appeal. Then we provide a detailed description of the formative research methodology, including designed cases, *in vivo* naturalistic cases, and *post facto* naturalistic cases. Finally, we address methodological issues of construct validity, data collection and analysis procedures, and generalizability to a design theory.

Criteria for Evaluating Research on Generalizable Design Knowledge

In research on descriptive theory, the major methodological concern is validity—how well the description matches the reality of "what is." In contrast, for a design theory (or a guideline, model, etc.), the major concern is **preferability**—the extent to which a method is "better" than other known methods for attaining the desired outcome. But what is "better"? What constitutes preferability? As discussed in Chapter 1, the criteria you use depend on your values, or more appropriately, they should depend on the values of all those who have a stake in the application of the design theory. Those values array themselves on at least three dimensions: effectiveness, efficiency, and appeal (cf. Frick & Reigeluth, 1992; Reigeluth, Volume 1—1983). Each of these is discussed next.

1. Effectiveness. Often the most important aspect of effectiveness is the extent or degree to which the application of the theory (or guideline or method) attained the goal in a given situation. This is usually measured on a numerical scale in either a norm-based or criterion-based manner (cf. Mager, 1984). Another aspect is the dependability with which it attained the goal over repeated trials. Dependability is measured by looking at probabilities. Analysis of patterns in time (APT) is a useful methodology for examining multiple cases and estimating probabilities (Frick, 1983; 1990). A third aspect is the breadth of contexts (or situations) in which it attains the goal. Different methods are often preferable for different situations, and, indeed, it is the provision of different methods for different situations that raises the design knowledge above the level of a method or model to that of a design theory.

2. Efficiency. This has to do with "bang for the buck," which includes two elements: a measure of the "bang" (effectiveness) and a measure of the "buck" (cost, either in money or time, or some other cost, or a combination of costs). For instructional-design theory, we must consider human time, effort, and energy required, as well as the cost of further resources needed, such as materials, equipment or other requirements of the setting needed for instruction.

3. Appeal. This is an issue of how enjoyable the resulting designs are for all people associated with them. For instructional-design theory, this includes teachers and students, support personnel, and perhaps even administrators and parents. Appeal is independent from effectiveness and efficiency.

These three criteria—effectiveness, efficiency and appeal—may be valued differently in different situations, because stakeholders' wants and needs are likely to differ. Therefore, all three criteria should be manifest in the research design for generating design knowledge. We need to look at how a particular design theory holds up on all three dimensions when continuing to refine it, and perhaps even generate different variations within the theory for different value weightings on the three criteria. For example, certain methods may be preferable when efficiency matters little compared with effectiveness, whereas other methods may be preferable when efficiency (e.g., low cost or short instructional time) is more important than effectiveness.

Finally, it should be patent that the development and testing of design theories for a set of cases is not a one-trial endeavor. It is a matter of successive approximation. Such theories continue to be improved and refined over many iterations. The Montessori system of education is a good example (Montessori, 1964; 1965). An educational-design theory can be useful early in its life, once initially substantiated as having instrumental value, and then continue to be refined and modified over many generations of educators who apply it.

Given these criteria for evaluating a research methodology for creating design theory—which include the preeminence of preferability over validity—how can one conduct research that meets these criteria? The remainder of this chapter is devoted to offering some guidance for using the formative research methodology, based on about a dozen studies that have used variations of this methodology.

Formative Research

Formative evaluation (sometimes called field testing or usability testing) is a methodology for improving instructional resources and curricula (Bloom, Hastings & Madaus, 1971; Cronbach, 1963; Scriven, 1967; Thiagarajan, Semmel & Semmel, 1974). It entails asking such questions as “What is working?”, “What needs to be improved?”, and “How can it be improved?” (Worthen & Sanders, 1987, p. 36). Using it as the basis for a developmental or "action" research methodology for improving instructional-design theories is a natural evolution from its use to improve particular instructional systems. It is also useful to develop and test design theory on other aspects of education, including curriculum development, counseling, administration, finance, and governance.

The underlying logic of formative research as discussed by Reigeluth (1989) is that, if you create an accurate application of an instructional-design theory (or model), then any weaknesses that are found in the application may reflect weaknesses in the theory, and any improvements identified for the application may reflect ways to improve the theory, at least for some subset of the situations for which the theory was intended. There are notable similarities to the logic of experimental design, in which one creates an instance of each parameter of an independent variable, one collects data on the instances, and one generalizes back to the independent-variable concepts. Replication with diverse students, content, and settings is necessary in both cases. However, for formative research the guiding questions are, "What methods worked well?" "What did not work well?" and "What improvements can be made to the theory?"

In the formative research methodology, an instance (or application) of a theory is created or identified. The design instance is based as exclusively as possible on the guidelines from that theory. For example, for an instructional-design theory, a course might be developed based solely on that theory, using as little intuition as possible. The application (the course in this case) is then formatively evaluated using one-to-one, small-group, and/or field-trial formative evaluation techniques (Dick & Carey, 1990; Thiagarajan, Semmel & Semmel, 1974). The data are analyzed for ways to improve the course, and generalizations are hypothesized for improving the theory.

Formative research has been used to improve existing instructional-design theories and models, including the Elaboration Theory (English, 1992; Kim, 1994), a theory to facilitate understanding (Roma, 1990; Simmons, 1991), a theory to foster awareness of ethical issues (Clonts, 1993), a theory for designing instruction for teams (Armstrong, 1993), and a theory for the design of computer-based simulations (Shon, 1996). It has also been used to improve instructional systems development (ISD) models, such as Keller's (1987) process for the motivational design of instruction (Farmer, 1989). Furthermore, it has been used to improve educational systems design (ESD) models for school systems engaging in systemic change (Carr, 1993; Naugle, 1996). The methodology has proven valuable for identifying ways to improve these theories and models, and it could also be used to improve theories and models in virtually all fields of education.

Methodological Procedures in Formative Research

Formative research follows a case study approach as outlined by Yin (1984). Specifically, the design is typically a holistic single case—one application of the theory. The study is exploratory in nature because there is "no clear, single set of outcomes" (Yin, 1984, p. 25). Yin believes that a single case study is appropriate when "a how or why question [has been] asked about a contemporary set of events" (p. 20), which includes how to improve a design theory. This type of methodology lends itself well to researcher-teacher collaboration.

Specifics of the research methodology vary depending on the kind of formative research study. Over the past seven years, we have gradually refined several methodologies for formative research, through the conduct of a dozen studies (Armstrong, 1993: Carr, 1993; Clonts, 1993; English, 1992; Farmer, 1989; Khan, 1994; Kim, 1994; Naugle, 1996; Roma, 1990; Shon, 1996; Simmons, 1991; Wang, 1992).

Case studies can be classified as designed cases or naturalistic cases, depending on whether the situation under investigation is manipulated in any way by the researcher. Formative research is a **designed case** if the researcher instantiates the theory (or model) and then formatively evaluates the instantiation. Alternatively, it is a **naturalistic case** if the researcher (a) picks an instance (or case) that was not specifically designed according to the theory but serves the same goals and contexts as the theory, (b) analyzes the instance to see in what ways it is consistent with the theory, what guidelines it fails to implement, and what valuable elements it has that are not present in the theory, and (c) formatively evaluates that instance to identify how each consistent element might be improved, whether each absent element might represent an improvement in the instance, and whether removing the elements unique to the instance might be detrimental. Furthermore, for naturalistic cases, the methodology varies depending on whether the observation is done during or after the practical application. This makes three major types of formative research studies:

- **designed cases**, in which the theory is intentionally instantiated (usually by the researcher for the research,
- ***in vivo*** **naturalistic cases**, in which the formative evaluation of the instantiation is done during its application, and
- ***post facto*** **naturalistic cases**, in which the formative evaluation of the instantiation is

done after its application.

And within each of these three types, the methodology also varies depending on whether the study is intended to develop a new design theory (one which does not yet exist) or to improve an existing theory. Table 1 shows these variations.

Insert Table 1 about here

For a designed case to improve an existing theory, the methodological concerns center within the following process:

1. Select a design theory.
2. Design an instance of the theory.
3. Collect and analyze formative data on the instance.
4. Revise the instance.
5. Repeat the data collection and revision cycle.
6. Offer tentative revisions for the theory.

For a designed case to develop a new theory, the process changes a bit:

1. (Not applicable.)
2. Create a case to help you generate the design theory.
3. (Same as for an existing theory.)
4. (Same as for an existing theory.)
5. (Same as for an existing theory.)
6. Fully develop your tentative theory.

For both *in vivo* and *post facto* naturalistic studies, the process is still different:

1. (Same as for a designed case, for either a new or existing theory.)
2. Select a case.
3. Collect and analyze formative data on the case.
4. (Not applicable.)

5. (Not applicable.)

6. (Same as for a designed case, for either a new or existing theory.)

Next is a description of each of these kinds of formative research, beginning with the most common one, a designed case to improve an existing theory.

Designed Case To Improve an Existing Theory

While there is often much variation from one such case study to another, the following is a fairly typical process for conducting this type of formative research study.

1. Select a design theory. You begin by selecting an existing design theory (or model) that you want to improve.

> For example, Robert English, a teacher at a university in Indiana, selected the Elaboration Theory of Instruction (Reigeluth & Stein, 1983) for his dissertation study (English, 1992).

2. Design an instance of the theory. Then you select a situation that fits within the general class of situations to which that design theory (or model) applies, and you design a specific application of the design theory (called a "design instance") . This instance may be a product or a process, or most likely both. It is important that the design instance be as pure an instance of the design theory as possible, avoiding both of the two types of weaknesses (omission: not faithfully including an element of the theory; and commission: including an element that is not called for by the theory). This is an issue of construct validity, and its counterpart in experimental design is ensuring that each of the treatments is a faithful representation of its corresponding independent-variable concept.

The design of the instance can be done either by the researcher (as participant) or by an expert in the theory (with the researcher as observer), preferably with the help of a subject-matter expert (usually the teacher for the course used in the instance). In either event, it is wise to get one or more additional experts in the theory to review the instance and ensure that it is a faithful instance of the theory. If you find yourself or the expert in the theory having to make decisions about which the theory offers no guidance, make special note of all such occurrences, as areas of guidance that should be added to the design theory later. It is also wise to get one, or preferably several, additional subject-matter experts to review the instance for content accuracy.

For example, Robert English picked a basic college course on electricity that he was regularly teaching. He took four chapters from the textbook for the course and re-sequenced them according to the Elaboration Theory. Then he had one of the authors of the theory (Reigeluth) review the sequence for validity of representing the Elaboration Theory's guidelines.

3. Collect and analyze formative data on the instance. Next, you begin data collection by conducting a formative evaluation of the design instance (see e.g., Dick & Carey, 1990). The intent is to identify and remove problems in the instance, particularly in the methods prescribed by the theory. In some situations, design and implementation of the instance occur simultaneously, in which case the data are collected during the design process (or alternatively design occurs during the data collection process). In other situations, design and development of an instance are completed before implementation begins, in which case data collection comes as a separate phase of activity. In still other situations, you can do a combination—some small-scale testing of parts as you design the instance, then larger-scale testing of the whole when it is completed.

In the case of English's study of the Elaboration Theory, design and development were completed before the implementation began, because it is hard to test a macro-level sequence before its design is completed.

First, you should prepare the participants, so that they will be more open in providing you with the data you need. This can be done by explaining that you are testing a new method, that you want them to be highly critical of it, and that any problems encountered will be due to weaknesses in the method, not to deficiencies in themselves. Try to establish rapport with them, and in one-to-one formative evaluations, try to get them to think aloud during the process (in this case, the instructional process).

For example, Robert English explained to the students that a new course design was

> being used and that they were being asked for their reactions to it. He told them that any mistakes they made or any misunderstandings encountered would be due to deficiencies in the course rather than to their learning ability. Before instruction actually began, he established rapport with the learners to increase their comfort level enough to interact and make frank comments, and he encourage them to be as critical as possible. He also asked the students to think aloud and to make notes on the material while proceeding.

Three techniques are useful for collecting the formative data: observations, documents, and interviews. **Observations** allow you to verify the presence of elements of the design theory and to see surface reactions of the participants to the elements. **Documents** on both elements (methods of instruction, in this case) and outcomes can help you to make judgments about the value of elements of the theory. For example, test results can help you to gauge how much learning occurred and what types of learning occurred. Newspaper reports of effects on the community can provide new insights about the value of certain elements or triangulation for elements on which you already have some outcome data, assuming the effects reported in the newspaper reflect the criteria you have established for assessing preferability, as discussed earlier.

But usually the most useful data come from **interviews** with the participants. Both individual and group interviews, or interactions, allow you to probe the reactions and thinking of the participants (such as teachers and students). They help you to identify strengths and weaknesses in the design instance, but they also allow you to explore improvements for elements in the design instance, to explore the likely consequences of removing elements from, or adding new elements to, the instance, and to explore possible situationalities (ways that methods should vary for different situations, such as kinds of learning, learners, learning environments, and development constraints for research on instructional-design theories). Although such data, as conjecture from the participants, are always suspect, they can also be highly insightful and useful. At a minimum they will likely provide some hypotheses worthy of testing with subsequent participants and situations. Interviews can be done during or after the implementation of the instance, or both.

Interactions with the participants ***during*** the implementation of the design instance should be guided by a set of questions that progress from very open-ended ones to very targeted ones. These questions should be tailored to the design theory under investigation, and should strive to collect data about how to improve the specific guidelines in the theory, including adding new guidelines that may better attain the goals targeted by the theory. Therefore, for instructional-design theory the questions should focus on identifying particular aspects of the implementation of the design instance that helped or hindered learning and finding ways to improve weak elements. The questions should be used flexibly and responsively, as they are prompted by such cues as facial expressions (e.g., a quizzical look), and used at break points in the implementation of the instance. If participants experience difficulties with certain elements of the instance, it is usually wise to help them overcome those difficulties before they proceed, so that future data will not be tainted by earlier weaknesses in the instance.

A different set of open-ended questions should be used ***after*** the implementation of the design instance. They should ask the participants such things as what they did and did not like about the various elements of the instance, what helped them, what did not help them, whether they felt that the materials and activities were appropriate for their needs, what changes they would make if they could, and whether they felt they attained the objectives. The purpose of the debriefing questions is to give the participants an opportunity to reflect on and evaluate the implementation of the design instance as a whole, to point out any strengths and weaknesses not mentioned before, and to make any additional comments. They should be strongly encouraged to point out weaknesses. Reliability or consistency across participants should be assessed so the point of saturation can be determined.

One additional point is worth mentioning here. Participants sometimes forget details about the design instance, and they have to be reminded where a particular element came in the overall process. Once shown, they usually have a lot to say. We suggest, then, after the first open-ended questions, to have the participants trace back through the process to specifically recall their impressions. It can be particularly helpful to show the participant a video tape of the process.

Usually, the most useful data come from one-to-one interviews with participants during the implementation of the design instance, because you avoid the memory-loss problem of interviews after the fact and you can overcome problems that might jeopardize data collection in the remainder of the implementation. But interviews during the implementation have less external validity because of their intrusiveness. As in formative evaluation, we recommend starting with the richer but less valid data collection technique (one-to-one interviews during the implementation of the design instance) and moving to progressively less rich but more representative techniques (small-group and field trials with interviews afterwards) to confirm the richer findings. It is usually helpful to record the interviews. And, in the more authentic trials for which the interviews are conducted afterwards, it is often helpful to video record the implementation of the design instance and have the participant comment about it while viewing the tape. Also, "member checking" (Guba & Lincoln, 1981) should be done with each participant as soon as possible after the information is recorded. One technique for member checking is to show each participant a typed summary of the information s/he contributed and discuss its accuracy.

> <u>English used all three techniques, but concentrated on interviews with students. He used the one-to-one interviews to explore ways to help each student with difficult content, and he recorded all comments made. There were two phases in his study: an "interactive" data collection phase, which entailed interacting with each student during the instructional process, and a "non-interactive" data collection phase, which entailed interacting with each student only in a debriefing session after the instruction. Phase 1 data were richer, and Phase 2 data were used to check the validity of the results from Phase 1.</u>

The data collection should always focus on how to improve the design theory. We have found it beneficial to focus on what should *not* be changed (strengths), as well as what should be changed (weaknesses). Wherever weaknesses are found, it is, of course, important to get the learners' (or users') suggestions for ways to overcome those weaknesses, or at very least their reactions to any ideas you have about how to overcome each weakness. Several iterations of data collection are strongly advised (equivalent to increasing the number of subjects in an experimental study), to assess dependability of results. In these iterations, it is wise to systematically vary the situation (types of people and conditions) as much as you can, within the limits of the class of situations for which the theory is intended. This enables you to identify situationalities (different methods for different contextual conditions) and enhances external validity (generalizability).

> English's data included strengths, weaknesses, and suggestions for improvement to the theory. Also, English interviewed a total of 10 students in Phase 1 before reaching saturation, and three students in Phase 2. The students in each phase were evenly distributed across the intelligence spectrum.

Data analysis should be conducted during the data collection process, if possible, to identify consistency of data across students. Of major concern is identifying the principal strengths and weaknesses in the instruction and what improvements could be made to the theory. Data analysis involves three activities: data reduction, data display, and conclusion drawing (Miles & Huberman, 1984). Data reduction is "selecting, focusing, simplifying, abstracting, and transforming the 'raw' data...." (Miles & Huberman, 1984, p. 21). The analytical procedure outlined by Miles and Huberman (1984) focuses on categorizing the data by the types of observations made during the implementation of the design instance or the types of answers to questions during debriefing. Summary information could be placed in a series of matrices (such as those developed by Roma, 1990) which specify relevant situational characteristics (e.g., the students, content, and context) and array categories of data (e.g., elements of the theory) across them. Each cell would then represent either a positive/negative or yes/no response, depending on the nature of the data. Specific recommendations for improvement could be keyed to each weakness identified in the matrix and described in detail apart from the matrix. Many of the matrix categories cannot be determined prior to the study, as the majority of questions are open-ended.

One potential problem with open-ended questions is that many of the cells you end up with in the matrices may not be filled, because some students might not offer any data on some categories. This would make it difficult to draw adequate conclusions for all categories across types of situations (e.g., students, content, and contexts). One way to eliminate this problem is to use a combination of both open-ended and directed questions during data collection. This mixture could contribute information about specific aspects of the design instance from all participants and would, therefore, increase the number of filled cells. But it would be impossible to predict all categories of information, so we do not recommend the use of *only* directed questions. Our suggestion is to start with open-ended questions, and then use directed questions for certain important issues you know of in advance of the study or that emerge very early during data collection.

4. Revise the instance. Next, you make revisions in the instance of the design theory, based on the data you collected. These revisions do not have to wait until you finish all the data collection and analysis. If you make the revisions as soon as you feel fairly confident in their value, then you can use them in your remaining data collection, perhaps even showing both versions of the design instance to the same student for comparative evaluation. You should also take note of the nature of the revisions, for they represent hypotheses as to ways in which the design theory itself might be improved.

5. Repeat the data collection and revision cycle. Several additional rounds of data collection, analysis, and revision are recommended, again systematically varying the situation (people and conditions) as much as you can from round to round, within the boundaries of the theory. This is a way of confirming the earlier findings, and it enhances external validity (generalizability) so essential for justifying changes in the design theory itself. During this process, you are likely to find that a method that works very well for some situations may not work as well as an alternative method for other situations. Such "situationalities" are important discoveries in a research effort to improve a design theory and better meet the needs of practitioners.

6. Offer tentative revisions for the theory. Finally, you should use your findings to hypothesize an improved design theory. Naturally, your suggestions will not become "knowledge" until they have been more thoroughly replicated and validated. Additional formative research studies will provide the needed replication, but experimental studies are a form of research well suited to validation (or refutation!).

Designed Case To Develop a New Theory

This kind of formative research differs from the previous one primarily in that you do not start with an existing design theory. This means that you must skip Step 1 above entirely. Second, you must greatly modify Step 2 so as to design the best case (counterpart to an instance of a design theory), without a design theory for guidance. The purpose of this is to be able to use a concrete case from which to build grounded design theory, based largely on experience and intuition. Several of the theories in this book seem to have been developed using basics of this kind of approach (e.g., Corno & Randi, Chapter 14; Nelson, Chapter 12; Perkins & Unger, Chapter 5; Pogrow, Chapter 15; Schwartz, Lin, Brophy & Bransford, Chapter 10), though the use of this approach would naturally have been intuitive rather than explicit. Steps 3, 4, and 5 remain virtually unchanged, but Step 6 must now be a process of inductively developing a new design theory, rather than modifying an existing theory. The new Steps 2 and 6 are as follows.

2. Create a case to help you generate the design theory. You begin by selecting a situation that fits within the general class of situations to which you want your new design theory (or model) to apply. Then you design the best case you can for that situation, using experience, intuition, and trial and error, often in combination with knowledge of related descriptive, scientific knowledge of education. This case may, of course, be a product and its implementation, or a process, or most likely both. The case must be created by the person who will be developing the design theory, because intimate familiarity with the case is essential for developing good grounded theory. As you develop the case, you should develop a tentative design theory in parallel. For each element you decide to include in your case, you should generate guidelines for selection and use of that element and incorporate them into your theory, so that your case will become an instance of the theory.

6. Fully develop your tentative theory. Finally, you should use your findings to revise and elaborate your tentative design theory. It is unlikely you will have been able to test your theory for the full range of situations for which the theory is intended, so there will likely be holes and other inadequacies in the theory. You should try to identify and describe all such inadequacies when you offer your theory to the public. And you should offer a research agenda that identifies the types of developmental studies you think would help most to further develop the theory.

Naturalistic Formative Research Studies

A naturalistic formative research study differs from a designed study in that you (the researcher) do not create a design instance or case. It is a naturalistic (in the general sense of the term) form of research, in that you observe a case that someone else has created. The case could be one that you study throughout its occurrence, which we call an ***<u>in vivo</u>*** naturalistic study, or it could be one that you study after the fact, which is called a "***<u>post facto</u>***" study. The major difference between <u>post facto</u> and <u>in vivo</u> studies is that observation is not possible in <u>post facto</u> studies (except perhaps through video segments), and interviews are more difficult to arrange and less likely to yield accurate and complete data, due to the passage of time. In either event, if it is a purely naturalistic study, there is no opportunity to revise the case based on your data (Step 4 above) nor to try out the revision (Step 5). A naturalistic case can be used either to improve an existing design theory or to create a new one (that doesn't exist yet). Here are the details for conducting a naturalistic study.

1. (Same as for a designed case, for either a new or existing theory.)

2. Select a case. Instead of creating an instance (for research on an existing theory) or case (for research to develop a new design theory), you select a case that is about to begin (for an in vivo study) or a case that has been completed (for a *post facto* study). For research on either an existing or new theory, the case should be in a situation that fits within the general class of situations to which the theory applies. This case may be a product or a process, or most likely a combination of the two. For testing an existing theory, it is helpful for the case to be as close as possible to being an instance of the theory—i.e., it will contain many of the elements that are called for by the theory. And it is wise to get one or more experts in the theory to observe the case to identify what elements are faithful instances of elements of the theory.

3. Collect and analyze formative data on the case. Next, you collect and analyze formative data on the case. For research on an existing theory, there are three major kinds of data you want to collect, based on presence and absence of elements in the theory and in the case. One is for elements that are present in both the theory and the case. You should identify the strengths, weaknesses, and possible improvements for elements of the theory that were used in the case. Can those elements be improved in the case (and by inference in the theory)? Another is for elements that are present in the theory but absent in the case. You should identify the consequences of the case not using specific elements of the theory. Should those elements have been used? The third is for elements that are absent in the theory but present in the case. You should identify the strengths, weaknesses, and possible improvements for elements in the case that are not in the theory, including situationalities as well as methods. Should those elements be added to the theory? For research to develop a new theory, only this third kind of data is relevant. You need to rely heavily on your intuition, experience, and knowledge of relevant descriptive theory to develop hypotheses as to what might generalize from this case and how far it might generalize, including situationalities.

These three kinds of data can be collected through the three techniques discussed earlier: observations, interviews, and documents. Observations help you to judge which elements of the theory are present in the case and which are missing, as well as to see surface reactions of the participants to those elements. However, for *post facto* studies, observation will likely only be possible if video tapes were made of the case. Interviews—both individual and group—allow you to probe the reactions and thinking of the participants and to explore improvements for the case. Although these are difficult for *post facto* studies, it is usually possible to find and interview some people who were involved in the case. Finally, documents on both methods and outcomes can help you to make judgments about the value of elements of the theory.

As with the designed cases described earlier, the data collection should always focus on how to improve the case, even though you cannot actually make any such improvements to it. This entails identifying strengths, so that you know what not to change, as well as identifying weaknesses, for which you should look for specific improvements. Several replications of data collection are strongly advised (equivalent to increasing the number of subjects in an experimental study), to assess reliability of results. Again, we recommend the technique of saturation. Also, by varying contextual factors in each round, you can look for situationalities and breadth of generalizability.

Steps 4 and 5 are not applicable, and **Step 6** is the same as for a designed case study, for either a new or existing theory.

Examples of *post facto* formative research include Khan (1994) and Wang (1992). Furthermore, Collins and Stevens' (Volume I—1983) Cognitive Theory of Inquiry Teaching is an example of a design theory developed through *post facto* naturalistic formative research. They obtained transcripts of a variety of “interactive” teachers—ones who used “some form of the case, inquiry, discovery, or Socratic method” (p. 250). And they “abstract[ed] common elements of those teachers’ teaching strategies, and ... show[ed] how these can be extended to different domains.” (p. 250).

These do not represent all the possible kinds of formative research methodologies, and they belie the fact that these are not mutually exclusive nor “hard and fast” categories. In fact, many formative research studies have elements of more than one of these kinds. Also, there is room for wide variation within each category. But there are methodological issues they all have in common, as well as some that are unique to each.

Methodological Issues for Formative Research

Case studies have been criticized in the past for their lack of rigor. However, this concern can be addressed by attending to three classes of methodological issues: A) construct validity, B) sound data collection and analysis procedures, and C) attention to generalizability to the theory.

A. Construct validity

Construct validity is concerned with "establishing correct operational measures for the concepts being studied" (Yin, 1984, p. 37). The concepts of interest in formative research are the *methods* offered by the design theory, any *situations* that influence the use of those methods, and the *indicators* of strengths and weaknesses (criteria for outcomes). The operationalization of the methods and analysis of relevant situations should be done by an expert in the theory, and preferably reviewed by one or two other experts in the theory, to assure their construct validity. As was mentioned under Step 2 above, there are two ways in which construct validity can be weakened: omission (not faithfully including an element of the theory) and comission (including an element that is not called for by the theory). The indicators of strengths and weaknesses should include the effectiveness, efficiency, and appeal of the methods, as discussed earlier. The indicators of effectiveness should be developed by an expert in measurement for the particular goals of the design instance or case, and reviewed by another. The indicators of efficiency should be developed by someone who is expert in measuring time and expense for both designing and using the methods, and those indicators should be reviewed by another expert. The indicators of appeal should be developed by an expert in motivational measurement and reviewed by another.

B. Sound Data Collection and Analysis Procedures

The soundness of the data collection and analysis procedures is influenced by two major factors: the thoroughness or completeness of the data, and the credibility or accuracy of the data. These two factors overlap to some degree, but it is helpful to think of them as two separate issues (cf., Rubin, 1994, on usability engineering).

Thoroughness of the data can be enhanced through a number of techniques, including advance preparation of participants, an emergent data-collection process, gradually decreasing obtrusivity, iteration until saturation, and identification of strengths as well as weaknesses.

First, participants often require *advance preparation* because they may have difficulty critiquing the design instance or case. For example, students tend to blame their learning problems on themselves rather than on their instruction. And even if they see problems with the instruction, students are often hesitant to criticize it in the presence of one who may have some ego investment in it. Therefore, it is important—before the implementation of the design instance or case begins—to prepare the participants to be critical. Furthermore, establishing rapport with the participants will tend to make them more open to sharing all their reactions.

Second, because you have little idea as to what weaknesses and areas of improvement you may find in the theory, it is important that your data-collection process be *emergent*, that your quest for data be flexible and responsive to your findings, starting with open-ended probes (e.g., questions, observations, documents) and gradually becoming more targeted in response to promising leads.

Third, it is helpful to start with fairly obtrusive probes (that interrupt the implementation of the design instantiation) in your early rounds of data collection (e.g., with your first students in one-to-one interactions) and gradually become *less obtrusive* to confirm the earlier findings under conditions that have greater external validity.

Fourth, it is generally wise to continue the rounds (or iterations) of obtrusive probes until you have reached the point of *saturation* (where new rounds of data collection merely confirm prior findings and yield no new findings) (cf. Merriam, 1988).

Finally, to be thorough, you should be sure to collect information about the *strengths* as well as the weaknesses of the design instance or case, and about what should not be changed as well as what should.

Credibility of the data can also be enhanced through a variety of techniques, including triangulation (Lincoln & Guba, 1985), chain of evidence, member checks (Guba & Lincoln, 1981), and clarification of the researcher's assumptions, biases, and theoretical orientation (Merriam, 1988).

First, *triangulation* entails using multiple sources of evidence (data) and cross-validating each source against the other sources. In a formative research study, the multiple sources of evidence are, first of all, multiple participants (e.g., students). Data should be collected in additional rounds (iterations) with different participants until considerable consistency of results (saturation) is obtained across participants. (This is a clear point of overlap with the thoroughness of the data.) And multiple sources of evidence should be collected for each participant (e.g., observations of a student during learning, interviews with the student, and the student's productions—tests, papers, project reports). We recommend that some objective measures be utilized for evaluating the design instance or case, to get some sense of the general acceptability of the outcomes (e.g., pre- and posttests for measuring the effectiveness of instruction, and similarly objective measures for efficiency and appeal).

Second, all data collection procedures should be clearly and precisely documented to establish a *chain of evidence*, and—as Yin suggests—the study should be performed as if someone is looking over the shoulder of the investigator.

Third, *member checks* usually entail taking data and interpretations back to the participants. Through further dialogue with participants errors or misconceptions by the researcher can be corrected, interpretations clarified, and emphases modified.

Finally, clarification of the *researcher's assumptions*, biases, and theoretical orientation should be done early in the research report, and every attempt should be made to make these views explicit.

C. Attention to Generalizability to the Theory

Finally, rigor in formative research is increased by enhancing ways that the results can be generalized to the theory. There are two major tools for doing this: recognizing situationality and replicating the study.

Situationality can be explored in at least two ways: (1) whenever you find different results in different rounds (iterations), look for differences in the situation (e.g., for a study on instructional design, the nature of: the learner, what is being learned, the learning environment, and the development constraints), and (2) purposely vary elements of the situation in your rounds of data collection to see if the results differ. These findings can allow you to hypothesize situationalities for the theory you are testing. When situationalities are incorporated into the theory, the theory becomes useful for a broader range of situations. At a very minimum, your research report should describe as completely as possible the situations under which the theory was applied in your study, so that others may draw conclusions about situationalities.

Replication is necessary to confirm the findings of any formative research study. With sufficient replications, hypotheses about improvements to the design theory gain sufficient evidence to warrant changes in the theory. Naturally, the replications should systematically vary all situational elements that may cause different methods to be preferable.

Concluding Remarks

Educational practitioners often find it difficult to utilize the kind of theoretical knowledge that is typically created by educational researchers. It is our observation that most educational research has resulted in descriptive knowledge. If this were true in each of the physical sciences, we would have primarily research in, say, physics or biology, with little research in mechanical engineering or medical practice.

Most of this chapter was devoted to describing various formative research methodologies for improving design theories—with our particular experience primarily in the area of instructional-design theories. Specifically, we described two methodologies for designed case studies: one for improving existing theories and one for developing new theories. And we described two methodologies for naturalistic case studies: one for *in vivo* studies and one for *post facto* studies, both of which can be used either for improving existing theories or developing new theories. These methodologies are based on knowledge gained from a dozen formative research studies that have been conducted over the past eight years.

It is our contention that to improve educational practice we need more—and betterquality—research on design theory. It is our hope that this chapter and volume will encourage more educational researchers to team with practitioners in conducting formative research, for this type of work shows great promise for advancing design theory that can better meet educators' needs and enhance our ability to reverse the status quo in which "The history of educational reform is one of consistent failure of major reforms to survive and become institutionalized." (Pogrow, 1996, p. 657).

It is not a simple matter to overcome the apparent biases towards descriptive knowledge of education to the exclusion of other types. Therefore, we also hope this chapter will encourage more foundations and agencies to fund this type of research and more professional journals to accept this kind of research for publication, for those organizations have great influence over the kinds of studies that researchers conduct.

References

Armstrong, R.B. (1993). The Team Instructional Prescriptions (TIP) Theory: A set of integrated models for prescribing instructional strategies for teams. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Bloom, B., Hastings, T., & Madaus, G. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

Carr, A.A. (1993). Selecting community participants for systemic educational restructuring: Who should serve? Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Clonts, J. (1993). Formative evaluation of an instructional theory for increasing awareness of ethical issues. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), New Directions in Educational Technology (pp. 15-22). Berlin: Springer Verlag.

Collins, A., & Stevens, A.L. (1983). A cognitive theory of inquiry teaching. In C.M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cronbach, L. (1963). Evaluation for course improvement. Teachers College Record, 64, 672-683.

Dick, W., & Carey, L. (1990). The systematic design of instruction. Glenview, IL: Scott, Foresman.

English, R. (1992). Formative research on the elaboration theory of instruction. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN. See also English, R.E., & Reigeluth, C.M. (1996). Formative research on sequencing instruction with the elaboration theory. *Educational Technology Research & Development, 44* (1), 23-42.

Farmer, T. (1989). A refinement of the ARCS motivational design procedure using a formative evaluation methodology. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Frick, T.W. (1983). Non-metric temporal path analysis: An alternative to the linear models approach for verification of stochastic educational relations. Unpublished doctoral dissertation, Indiana University Graduate School, Bloomington, IN.

Frick, T.W. (1990). Analysis of patterns in time: A method of recording and quantifying temporal relations in education. American Educational Research Journal, 27, 180-204.

Frick, T.W., & Reigeluth, C.M. (1992, February). Verifying instructional theory by analysis of patterns in time. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.

Greeno, J.G., Collins, A., & Resnick, L.B. (1996). Cognition and learning. In D.C. Berliner & R.C. Calfee (Eds.), Handbook of Educational Psychology (pp. 15-46). New York: Macmillan.

Guba, E.G., & Lincoln, Y.S. (1981). Effective evaluation. San Francisco, CA: Jossey-Bass.

Keller, J.M. (1987, November). The systematic process of motivational design. Performance & Instruction, 26, 1-8.

Khan, B. (1994). Post facto formative research on the educational systems design (ESD) process. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Kim, Y. (1994). Formative research on the Simplifying Conditions Method for task analysis and sequencing of instructional content. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Lincoln, Y.S., & Guba, E.G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage Publications.

Mager, R. (1984). Measuring instructional results. Belmont, CA: Lake Publishing.

Merriam, S.B. (1988). Case study research in education: A qualitative approach. San Francisco, CA: Jossey-Bass.

Miles, M.B., & Huberman, A.M. (1984). Analyzing qualitative data: A source book for new methods. Beverly Hills, CA: Sage Publications.

Montessori, M. (1964). The Montessori method. New York: Schocken Books.

Montessori, M. (1965). Spontaneous activity in education. New York, NY: Schocken Books.

Naugle, L. (1996). Formative research of the Reigeluth process model and an effort to initiate school restructuring. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Newman, D. (1990). Opportunities for research on the organizational impact of school computers. Educational Researcher, 19(3), 8-13.

Pogrow, S. (1996). Reforming the wannabe reformers: Why education reforms almost always end up making things worse. Phi Delta Kappan, 77, 656-663.

Reigeluth, C.M. (1983). Instructional design: What is it and why is it? In C.M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status (pp. 3-36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C.M. (1989). Educational technology at the crossroads: New mind sets and new directions. *Educational Technology Research & Development, 37* (1), 67-80.

Reigeluth, C.M., & Stein, F.S. (1983). The Elaboration Theory of Instruction. In C. M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status (pp. 335-382). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roma, C. (1990). Formative evaluation research on an instructional theory for understanding. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Romiszowski, A. (1988). Personal communication at AECT national conference in Dallas, TX.

Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York: Wiley.

Scriven, M. (1967). The methodology of evaluation. In AERA Monograph Series on Curriculum Evaluation, No. 1 (pp. 39-89). Chicago: Rand McNally.

Shon, M. (1996). Formative research on an instructional theory for the design of computer-based simulations for teaching causal principles. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Simmons, J. (1991). Formative evaluation research on an instructional theory for teaching causal principles. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Thiagarajan, S., Semmel, M.I., & Semmel, D.S. (1974). Instructional development for training teachers of exceptional children: A sourcebook. Minneapolis, MN: Leadership Training Institute/Special Education, University of Minnesota.

Wang, S.W. (1992). The process for creating and maintaining a Montessori public educational system in an urban community. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN.

Worthen, B., & Sanders, J. (1987). Educational evaluation: Alternative approaches and practical guidelines. New York: Longman.

Yin, R.K. (1984). Case study research design and methods. Beverly Hills, CA: Sage Publications.

Table 1

Kinds of Formative Research Studies

	... for an existing theory	**... for a new theory**
Designed case	Designed case for an existing theory	Designed case for a new theory
In vivo naturalistic case	In vivo naturalistic case for an existing theory	In vivo naturalistic case for a new theory
Post facto naturalistic case	Post facto naturalistic case for an existing theory	Post facto naturalistic case for a new theory