

Designing Qualitative Studies

5

The First Evaluation

The young people gathered around Halcolm. "Tell us again, Teacher of Many Things, about the first evaluation."

"The first evaluation," he began, "was conducted a long, long time ago, in ancient Babylon when Nebuchadnezzar was king. Nebuchadnezzar had just conquered Jerusalem in the third year of the reign of Jehoiakim, King of Judah. Now Nebuchadnezzar was a shrewd ruler. He decided to bring carefully selected children of Israel into the palace for special training so that they might be more easily integrated into Chaldean culture. This special program was the forerunner of the compensatory education programs that would become so popular in the 20th century. The three-year program was royally funded with special allocations and scholarships provided by Nebuchadnezzar. The ancient text from the Great Book records that

the king spoke unto Ashpenaz the master of his eunuchs that he should bring certain of the children of Israel, and of the king's seed, and of the princes; children in whom was no blemish, but well-favored and skillful in all wisdom, and cunning in knowledge, and understanding science, and such as had ability in them to stand in the king's palace, and whom they might teach the learning and the tongue of the Chaldeans.

And the king appointed them a daily provision of the king's meat, and of the wine which he drank; so nourishing them three years, that at the end thereof they might stand before the king. (Daniel 1:3-5)

Designing Qualitative Studies

5

The First Evaluation

The young people gathered around Halcolm. "Tell us again, Teacher of Many Things, about the first evaluation."

"The first evaluation," he began, "was conducted a long, long time ago, in ancient Babylon when Nebuchadnezzar was king. Nebuchadnezzar had just conquered Jerusalem in the third year of the reign of Jehoiakim, King of Judah. Now Nebuchadnezzar was a shrewd ruler. He decided to bring carefully selected children of Israel into the palace for special training so that they might be more easily integrated into Chaldean culture. This special program was the forerunner of the compensatory education programs that would become so popular in the 20th century. The three-year program was royally funded with special allocations and scholarships provided by Nebuchadnezzar. The ancient text from the Great Book records that

the king spake unto Ashpenaz the master of his eunuchs that he should bring certain of the children of Israel, and of the king's seed, and of the princes; children in whom was no blemish, but well-favored and skillful in all wisdom, and cunning in knowledge, and understanding science, and such as had ability in them to stand in the king's palace, and whom they might teach the learning and the tongue of the Chaldeans.

And the king appointed them a daily provision of the king's meat, and of the wine which he drank; so nourishing them three years, that at the end thereof they might stand before the king. (Daniel 1:3-5)

"Now this program had scarcely been established when the program director, Ashpenaz, who happened also to be prince of the eunuchs, found himself faced with a student rebellion led by a radical named Daniel, who decided for religious reasons that he would not consume the king's meat and wine. This created a serious problem for the director. If Daniel and his conspirators did not eat their dormitory food, they might fare poorly in the program and endanger not only future program funding but also the program director's head! The Great Book says:

But Daniel purposed in his heart that he would not defile himself with the portion of the king's meat, nor with the wine which he drank; therefore he requested of the prince of the eunuchs that he might not defile himself.
And the prince of the eunuchs said unto Daniel, I fear my lord the king, who hath appointed your meat and your drink; for why should he see your faces worse liking than the children which are of your sort? Then shall ye make me endanger my head to the king. (Daniel 1:8, 10)

"At this point, Daniel proposed history's first educational experiment and program evaluation. He and three friends (Hananiah, Mishael, and Azariah) would be placed on a strict vegetarian diet for ten days, while other students continued on the king's rich diet of meat and wine. At the end of ten days the program director would inspect the treatment group for any signs of physical deterioration and judge the productivity of Daniel's alternative diet plan. Daniel proposed the experiment thusly:

Prove thy servants, I beseech thee, ten days: and let them give us pulse to eat, and water to drink. Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou sees, deal with thy servants.
So he consented to them in this matter, and proved them ten days. (Daniel 1:12-14)

"During the ten days of waiting Ashpenaz had a terrible time. He couldn't sleep, he had no appetite, and he had trouble working because he was preoccupied worrying about how the evaluation would turn out. He had a lot at stake. Besides, in those days they hadn't quite worked out the proper division of labor so he had to play the roles of both program director and evaluator. You see. . . ."

The young listeners interrupted Halcolm. They sensed that he was about to launch into a sermon on the origins of the division of labor when they still wanted to hear the end of the story about the origins of evaluation. "How did it turn out?" they asked. "Did Daniel end up looking better or worse from the new diet? Did Ashpenaz lose his head?"

"Patience, patience," Halcolm pleaded. "Ashpenaz had no reason to worry. The results were quite amazing. The Great Book says that

at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat.

A Meta-Evaluation

A meta-evaluation is an evaluation of an evaluation. A great deal can be learned about evaluation designs by conducting a meta-evaluation of history's first program evaluation. Let us imagine a panel of experts conducting a rigorous critique of this evaluation of Babylon's compensatory education program for Israeli students:

1. Small sample size ($N = 4$).
2. Selectivity bias because recruitment into the program was done by "creaming," that is, only the best prospects among the children of Israel were brought into the program.
3. Sampling bias because students were self-selected into the treatment group (diet of pulse and water).
4. Failure to clearly specify and control the nature of the treatment, thus allowing for the possibility of treatment contamination because we don't know what nation because we don't know what
5. Possibility of interaction effects between the diet and the students' belief system (i.e., potential Hawthorne and halo effects).
6. Outcome criteria vague: Just what is "countenance"?
7. Outcome measurement poorly operationalized and nonstandardized.
8. Single observer with deep personal involvement in the program introduces possibility of selective perception and bias in the observations.
9. Validity and reliability data are not reported for the instruments used to measure the final, summative outcome ("he found them ten times better than all the magicians and astrologers").
10. Possible reactive effects from the students' knowledge that they were being evaluated.

other things, aside from a change in diet, either group was involved in that might have explained the outcomes observed.

—From Halcolm's *Evaluation Histories*



"And that, my children, is the story of the first evaluation. Those were the good old days when evaluations really got used. Made quite a difference to Ashpenaz and Daniel. Now off with you—and see if you can do as well."

As for these four children, God gave them knowledge and skill in all learning and wisdom; and Daniel had understanding in all visions and dreams. Now at the end of the days that the king had said he should bring them in, then the prince of the eunuchs brought them in before Nebuchadnezzar. And in all matters of wisdom and understanding, that the king inquired of them, he found them ten times better than all the magicians and astrologers that were in all his realm. (Daniel 1:15-18, 20)

Thus, Melzar took away the portion of their meat, and the wine that they should drink; and gave them pulse.

Despite all of these threats to internal validity, not to mention external validity, the information generated by the evaluations appears to have been used. The 10-day formative evaluation was used to make a major decision about the program, namely, to change the diet for Daniel and his friends. The end-of-program summative evaluation conducted by the king was used to judge the program a success. (Daniel did place first in his class.) Indeed, it would be difficult to find a more exemplary model for the uses of evaluation in making educational policy decisions than this "first evaluation" conducted under the auspices of Nebuchadnezzar so many years ago. This case study is an exemplar of evaluation research having an immediate, decisive, and lasting impact on an educational program. Modern evaluation researchers, falling away in seemingly futile efforts to affect contemporary government decisions, can be forgiven a certain nostalgia for the "good old days" in Babylon when evaluation really made a difference.

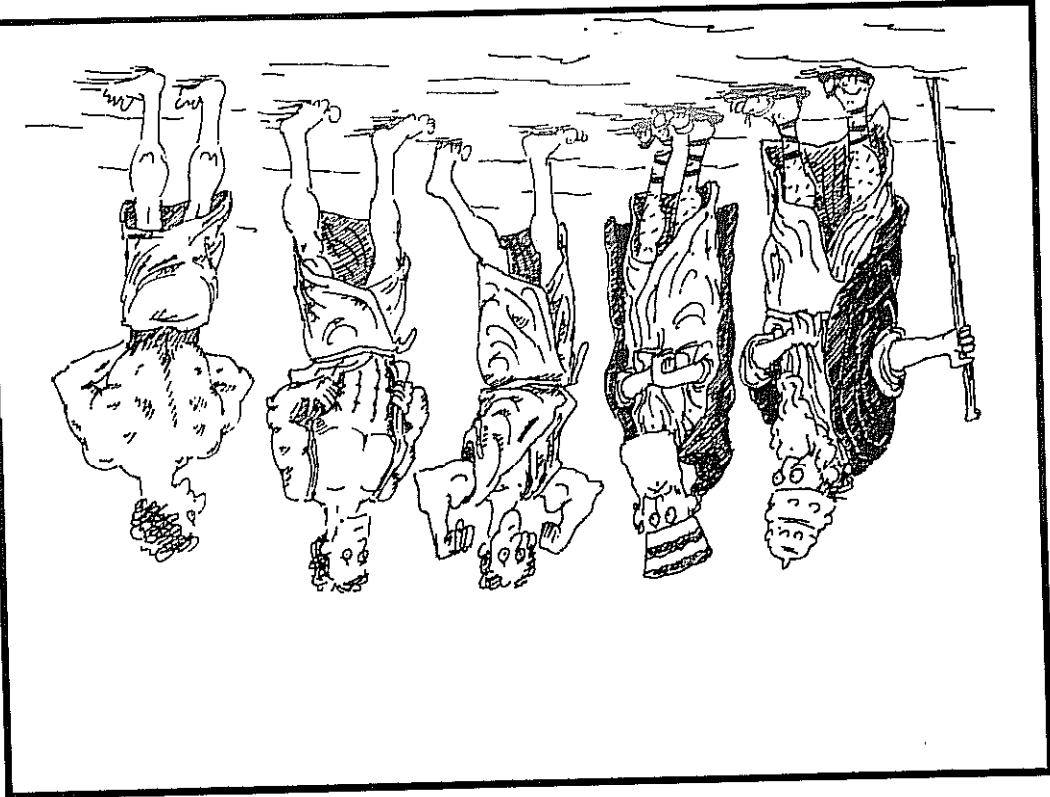
But should the results have been used? Given the apparent weakness of the evaluation design, was it appropriate to make a major program decision on the basis of data generated by such a seemingly weak research design?

I would argue that not only was use impressive in this case, it was also appropriate because the research design was exemplary. Yes, exemplary, because the study was set up in such a way as to provide precisely the information needed by the program director to make the decision he needed to make. Certainly, it is a poor research design to study the relationship between nutrition and educational achievement. It is even a poor design to decide if *all* students should be placed on a vegetarian diet. But those were not the issues. The question the director faced was whether to place four specific students on a special diet at their request.

The information he needed concerned the consequences of that specific change and *only* that specific change. He showed no interest in generalizing the results beyond those four students, and he showed no interest in convincing others that the measures he made were valid and reliable. Only he and Daniel had to trust the measures used, and so data collection (observation of countenance) was done in such a way as to be meaningful and credible to the primary intended evaluation users, namely, Ashpenaz and Daniel. If any bias existed in his observations, given what he had at stake, the bias would have operated against a demonstration of positive outcomes rather than in favor of such outcomes.

While there are hints of whimsy in the suggestion that this first evaluation was exemplary, I do not mean to be completely facetious. I am serious in suggesting that the Babylonian example is an exemplar of utilization-focused evaluation. It contains and illustrates all the factors modern evaluation researchers have verified as critical from studies of utilization (Patton 1997a). The decision makers who were to use findings generated by the evaluation were clearly identified and deeply involved in every stage of the evaluation process. The evaluation question was carefully focused on needed information that could be used in the making of a specific decision. The evaluation methods and design were appropriately matched to the evaluation question. The results were understandable, credible, and relevant. Feedback was immediate and utilization was decisive. Few modern evaluations can meet the high standards for evaluation set by Ashpenaz and Daniel more than 3,000 years ago.

This chapter discusses some ways in which research designs can be appropriately matched to evaluation questions in an attempt to emulate the exemplary match be-



© 2002 Michael Quinn Patton and Michael Cochran

Observing countenance

alternative purposes along a continuum from theory to action:

1. *Basic research:* To contribute to fundamental knowledge and theory
 2. *Applied research:* To illuminate a societal concern
 3. *Summative evaluation:* To determine program effectiveness
 4. *Formative evaluation:* To improve a program
 5. *Action research:* To solve a specific problem
- Basic and applied researchers publish in scholarly journals, where their audience is other researchers who will judge their contributions using disciplinary standards of

Purpose is the controlling force in research. Decisions about design, measurement, analysis, and reporting all flow from purpose. Therefore, the first step in a research process is getting clear about purpose. The centrality of purpose in making methods decisions becomes evident from examining

Clarity About Purpose: A Typology

between evaluation problem and research design achieved in the Babylonian evaluation. As with previous chapters, I shall emphasize the importance of being both strategic and practical in creating evaluation designs. Being strategic begins with being clear about the purpose of the intended research or evaluation.

Did it accomplish his objectives? Should it be continued? Could the outcomes be served be attributed to the program? This is the kind of research we have come to call *summative evaluation*—summing up judgments about a program to make a major decision about its value, whether it should be continued, and whether the demonstrated model can or should be generalized to and replicated for other participants or in other places.

Now imagine that researchers from the University of Babylon wanted to study the diet as a manifestation of culture in order to develop a theory about the role of diet in transmitting culture. Their sample, their data collection, their questions, the duration of fieldwork, and their presentation of results would all be quite different from the formative evaluation undertaken by Ashpenaz and Daniel. The university study would have taken much longer than 10 days and might have yielded empirical generalizations and contributions to theory, yet would not have helped Ashpenaz make his simple decision. On the other hand, we might surmise that University of Babylon scholars would have scoffed at an evaluation done in 10 days, even a formative one. Different purposes. Different criteria for judging the research contribution. Different methods. Different audiences. Different kinds of research.

These are examples of how purpose can vary. In the next section, I shall present a more formal framework for distinguishing these five different research purposes and examine in more depth the implications of varying purposes for making design decisions. Previous chapters have presented the nature and strategies of qualitative inquiry, philosophical and theoretical foundations, and practical applications. In effect, the reader has been presented with a large array of options, alternatives, and variations.

rigor, validity, and theoretical import. In contrast, evaluators and action researchers publish reports for specific stakeholders who will use the results to make decisions, improve programs, and solve problems.

Standards for judging quality vary among these five different types of research. Expectations and audiences are different. Reporting and dissemination approaches are different. Because of these differences, the researcher must be clear at the beginning about which purpose has priority. No single study can serve all of these different purposes and audiences equally well. With clarity about purpose and primary audience, the researcher can go on to make specific design, data-gathering, and analysis decisions to meet the priority purpose and address the intended audience.

In the Babylonian example, the purpose was simply to find out if a vegetarian diet would negatively affect the healthy appearances (countenances) of four participants—not *why* their countenances appeared healthy or not (a causal question), but *whether* the dietary change would affect countenance (a descriptive question). The design, therefore, was appropriately simple to yield descriptive data for the purpose of making a minor program adjustment. No contribution to general knowledge. No testing or development of theory. No generalizations. No scholarly publication. No elaborate report on methods, just find out what would happen to inform a single decision about a possible program change. The participants in the program were involved in the study; indeed, the idea of putting the diet to an empirical test originated with Daniel. In short, we have a very nice example of simple formative evaluation.

The king's examination of program participants at the end of three years was quite different. We might infer that the king was judging the overall value of the program.

work to generate new theories or test existing theories. Doctoral students are typically expected to make theoretical contributions in their dissertations. Theories encapsulate the knowledge of a discipline.

Basic researchers are interested in formulating and testing theoretical constructs and propositions that ideally generalize across time and space. The most powerful kinds of findings in basic science are those findings that are universal, such as Boyle's law in physics that the volume of a gas at constant temperature varies inversely with the pressure exerted on it. Basic researchers, then, are searching for fundamental patterns of the universe, the earth, nature, society, and human beings. For example, biologists have discovered that "changes in DNA are inherited, but changes in proteins (specifically, in their amino acid sequence) are not . . . , perhaps the only universal truth biologists have" (Smith 2000:43). Social science, to date, is markedly short of "universal truths." Nevertheless, generalizations across time and space remain the Holy Grail of basic research and theory.

The findings of basic research are published in scholarly books, journals, and dissertations. Each discipline has its own traditions, norms, and rules for deciding what constitutes valid research in that discipline. To be published in the major journals of any particular discipline, scientists must engage in the kind of research that is valued by the researchers in that disciplinary tradition.

Chapter 3 reviewed theoretical traditions closely associated with qualitative inquiry, for example, ethnography and phenomenology. Qualitative inquiry also contributes to basic research through inductive theory development, a prominent example being the "grounded theory" approach of Glaser and Strauss (1967), essentially an inductive strategy for generating and confirming theory that emerges from close involvement and di-

How does one sort it all out to decide what to do in a specific study? The answer is to get clear about purpose. The framework that follows is meant to facilitate achieving this necessary clarity about purpose while also illustrating how one can organize a mass of observations into some coherent typology—a major analytical tool of qualitative inquiry. The sections that follow examine each type of research: basic research, applied research, summative evaluation research, formative evaluation, and action research.

Basic Research

The purpose of basic research is knowledge for the sake of knowledge. Researchers engaged in basic research want to understand how the world operates. They are interested in investigating a phenomenon to get at the nature of reality with regard to that phenomenon. The basic researcher's purpose is to understand and explain.

Basic researchers typically work within specific disciplines, such as physics, biology, psychology, economics, geography, and sociology. The questions and problems they study emerge from traditions within those disciplines. Each discipline is organized around attention to basic questions, and the research within that discipline derives from concern about those basic questions. Exhibit 5.1 presents examples of fundamental questions in several disciplines.

The fundamental questions underlying each discipline flow from the basic concerns and traditions of that discipline. Researchers working within any specific disciplinary tradition strive to make a contribution to knowledge in that discipline and thereby contribute to answering the fundamental questions of the discipline. The most prestigious contribution to knowledge takes the form of a *theory* that explains the phenomenon under investigation. Basic researchers

on environmental studies often involves researchers from a number of disciplines. In agricultural research, the field of integrated pest management (IPM) includes research-ers from entomology, agronomy, agricul-tural economics, and horticulture. Fields of interdisciplinary research in the social sci-ences include gerontology, criminal justice studies, women's studies, and family re-search. Exhibit 5.2 offers examples of ap-plied *interdisciplinary* research questions for economic anthropology, social psychology, political geography, and educational and or-ganizational development. Notice the dif-ference between these questions and the kinds of questions asked by basic research-ers in Exhibit 5.1. Applied researchers are trying to understand how to deal with a sig-nificant societal problem, while basic re-searchers are trying to understand and ex-plain the basic nature of some phenomenon.

Applied qualitative researchers are able to bring their personal insights and experi-ences into any recommendations that may emerge because they get especially close to the problems under study during fieldwork. Audiences for applied research are typically policymakers, directors and managers of in-tervention-oriented organizations, and pro-fessionals working on problems. Sometimes for applied research depend a great deal on the timeliness and urgency of the problem being researched. A good example of ap-plied research is *Emerging Drug Problems*, a work sponsored by the U.S. General Ac-counting Office (1998) that examined new street drugs, recent research on addiction, and alternatives for public policy.

In contrast to basic researchers, who ulti-mately seek to generalize across time and space, applied research findings typically are limited to a specific time, place, and con-dition. For example, a researcher studying the nature of family problems in the 1980s

Applied Research

found that local artisans do not have total freedom in their attempts to create market niches since they are restricted by their abili-ties and the labor and capital they can mobi-lize. This is a classic example of interdisci-plinary theory generation and testing bridging economics, ethnology, and ecology.

Applied researchers work on human and societal problems. In the example just cited, had Chibnik examined the problem of creat-ing new markets for rural artisans and of-fered possible solutions for increased mar-keting, the work would have constituted applied rather than basic research. The pur-pose of applied research is to contribute knowledge that will help people understand the nature of a problem in order to intervene, thereby allowing human beings to more ef-fectively control their environment. While in basic research the source of questions is the traditions within a scholarly discipline, in applied research the source of questions is in the problems and concerns experienced by people and articulated by policymakers.

Applied researchers are often guided by the findings, understandings, and explana-tions of basic research. They conduct studies that test *applications* of basic theory and dis-ciplinary knowledge to real-world problems and experiences. The results of applied re-search are published in journals that special-ize in applied research within the traditions of a problem area or a discipline.

Societal concerns have given rise to a vari-ety of new fields that are interdisciplinary in nature. These emerging fields reflect the long-standing criticism by policymakers that universities have departments but soci-ety has problems. Applied interdisciplinary fields are especially problem oriented rather than discipline oriented. For example, work

EXHIBIT 5.2 Sample Interdisciplinary Applied Research Questions

Applied economic anthropology	How can the prosperous economy of an isolated, small minority group be preserved when that group encounters new competition from the encroaching global economy?
Applied social psychology	How can a group within a large organization develop cohesion and identity within the mission and values of its parent structure and culture?
Applied political geography	How can people of previously isolated towns, each with its own history of local governance, come together to share power and engage in joint decision making at a regional level?
Applied educational and organizational development	How can students from different neighborhoods with varied ethnic and racial backgrounds be integrated in a new magnet school?

would not expect those problems to be the same as those experienced by families in the 1880s. While the research might include making such comparisons, applied researchers understand that problems emerge within particular time and space boundaries.

Evaluation Research: Summative and Formative

Once solutions to problems are identified, policies and programs are designed to intervene in society and bring about change. It is hoped that the intervention and changes will be effective in helping to solve problems. However, the effectiveness of any given human intervention is a matter subject to study. Thus, the next point on the theory-to-action research continuum is the conduct of evaluation and policy research to test the effectiveness of specific solutions and human interventions.

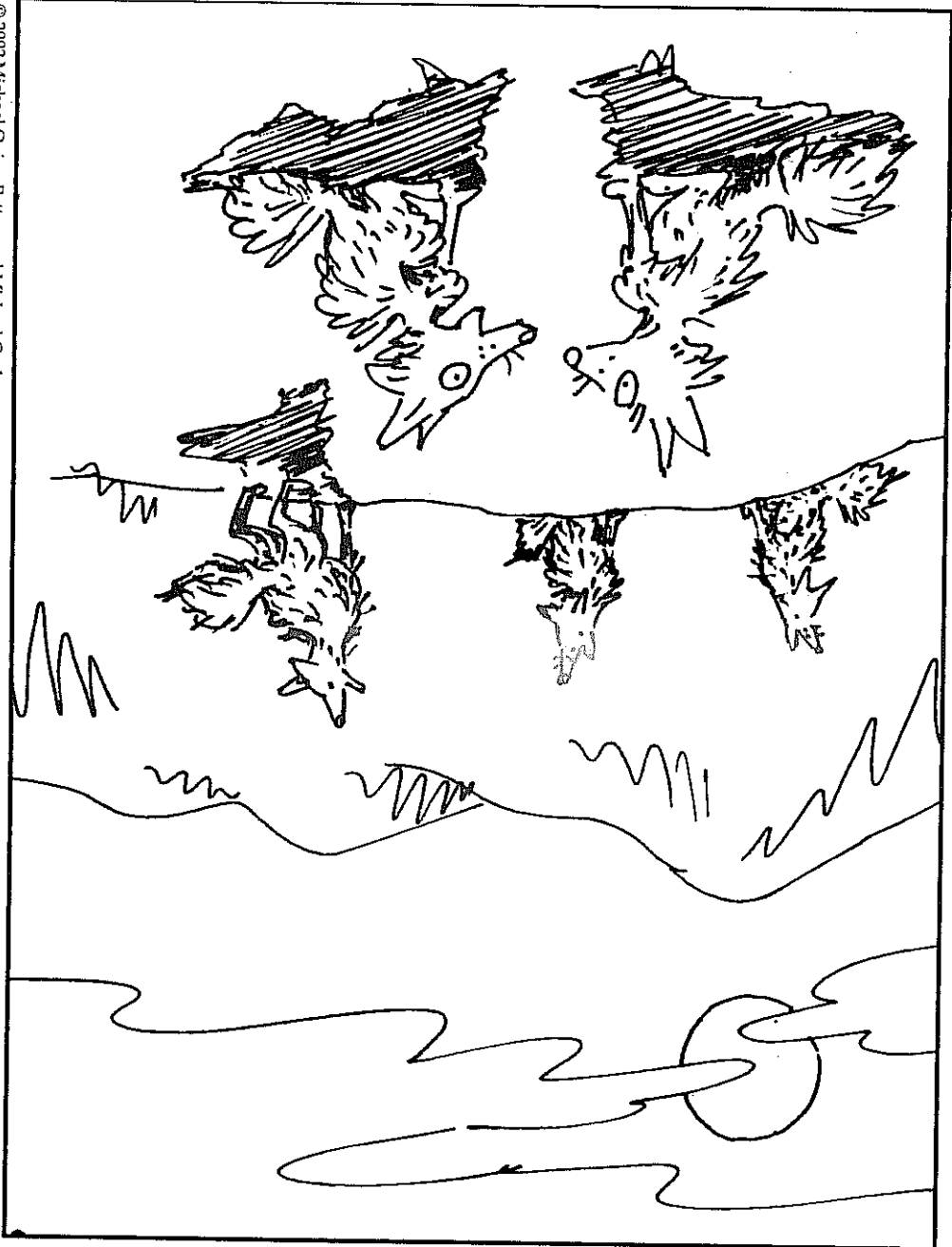
While applied research seeks to understand societal problems and identify potential

tial solutions, evaluations examine and judge the processes and outcomes aimed at attempted solutions. Evaluators study programs, policies, personnel, organizations, and products. Evaluation research can be conducted on virtually any explicit attempt to solve problems or bring about planned change. As illustrated in the Daniel story of history's "first evaluation" that opened this chapter, evaluators distinguish two quite different purposes for evaluation: (1) summative evaluations that judge overall effectiveness to inform major decisions about whether a program should continue and (2) formative evaluations that aim to improve programs.

Summative evaluations serve the purpose of rendering an overall judgment about the effectiveness of a program, policy, or product for the purpose of saying that the *evaluation* (thing being evaluated) is or is not effective and, therefore, should or should not be continued, and has or does not have the potential of being generalizable to other situations. A summative decision implies a

summing-up judgment or a summit (from the mountaintop) decision, for example, to expand a pilot program to new sites or move it from temporary (pilot test or demonstra- tion) funding to more permanent funding, or it may lead to program or policy termina- tion. Summative evaluations seldom rely entirely, or even primarily, on qualitative

Are they doing research or evaluation?



© 2002 Michael Quinn Patton and Michael Cochran

evaluation takes place. Formative evaluations rely heavily on process studies, implementation evaluations, case studies, and evaluability assessments (see Chapter 4). Formative evaluations often rely heavily, even primarily, on qualitative methods. Findings are context specific.

Although formative and summative remain the most basic and classic distinctions in evaluation, other evaluation purposes have emerged in recent years in "a world larger than formative and summative" (Patton 1996b). New purposes include ongoing "developmental evaluation" for program and organizational development and learning (Patton 1994; Freskill and Torres 1999); empowering local groups through evaluation participation (Fetterman 2000a; Patton 1997b); and using the processes of evaluation (*process use*) to build staff capacity for data-based decision making and continuous improvement (Patton 1997a: 87-113, 1998). For our analysis here, however, these and related approaches to evaluation share the general purpose of improvement and can be included within the broad category of formative research along our theory-to-action continuum. In addition, some evaluation studies are now designed to generate generalizable knowledge about effective practices across different projects or programs based on cluster evaluations, lessons learned, "better" practices, and meta-analyses (Patton 1997a: 70-75). This knowledge-generating approach to evaluation research, to the extent that it aims to discover general principles of effective practice rather than render judgment about the merit or worth of a specific intervention, falls roughly within the category "applied research" in this theory-to-action continuum. However, the emergence and increased importance of knowledge-generating evaluations illustrate why these five categories (basic, applied, summative, formative, and action re-

data and naturalistic inquiry because of decision makers' interest in measuring standardized outcomes, having controlled comparisons, and making judgments about effectiveness from relatively larger samples with statistical pre-post and follow-up results. Qualitative data in summative evaluations typically add depth, detail, and nuance to quantitative findings, rendering insights through illuminative case studies and examining individualized outcomes and issues of quality or excellence—applications discussed in Chapter 4. Harkreader and Henry (2000) have provided an excellent discussion of the challenges of rendering summative judgments about merit and worth; they use as their example comparative quantitative performance data from Georgia schools to assess a democratic reform initiative. Fetterman (2000b) shows how qualitative data can be the primary basis for a summative evaluation. His evaluation of STEP, a 12-month teacher education program in the Stanford University School of Education, included fieldwork immersion in the program, open-ended interviews with all students, focus groups, observations of classrooms, interviews with faculty, digital photography of classroom activities, and qualitative analysis of curriculum materials, as well as a variety of surveys and outcome measures. The summative evaluations of a democratic reform initiative in Georgia and of Stanford's STEP program both followed and built on extensive formative evaluation work, the purpose of which we now examine.

Formative evaluations, in contrast to summative ones, serve the purpose of improving a specific program, policy, group of staff (in a personnel evaluation), or product. Formative evaluations aim at *forming* (shaping) the thing being studied. No attempt is made in a formative evaluation to generalize findings beyond the setting in which the

informal, the people in the situation are often directly involved in gathering the information and then studying themselves, and the results are used internally to attack specific problems within a program, organization, or community. While action research may be used as part of an overall organizational or community development process, it most typically focuses on specific problems and issues within the organization or community rather than on the overall effectiveness of an entire program or organization. Thus, along this theory-to-action-research continuum, action research has the narrowest focus.

The findings of formative evaluation and action research are seldom disseminated beyond the immediate program or organization within which the study takes place. In many instances, there may not even be a full written research report. Publication and dissemination of findings are more likely to be through briefings, staff discussions, and oral communications. Summaries of findings and recommendations will be distributed for discussion, but the formality of the reporting and the nature of the research publications are quite different from those in basic, applied, or even summative evaluation research.

An example of action research comes from a small rural community in the Midwest in which the town board needed to decide what to do with a dilapidated building on a public park. They got a high school class to put together a simple telephone survey to solicit ideas about what to do with the building. They also conducted a few focus groups in local churches. The results showed that the townspeople preferred to fix up the building and restore it as a community center rather than tear it down. The action research process took about a month. Based on the findings, a local committee was formed to seek volunteers and funds for the restoration. In action research, by way of contrast, design and data collection tend to be more

unit. In action research, a product, or a staff policy, an organization, a program, a proving the effectiveness of a program, a formative evaluation focuses on ways of improving the effectiveness of a program, a design, at least in part, by an evaluator. For-analyzed, and the data are collected and/or designed in formative evaluation, there is a formal In formative evaluation, there is a formal being researched.

for the researcher as distinct from the people and the extent to which there is a special role and the different kinds of problems studied, the extent to which the research is systematic, the difference between formative evaluation and action research centers on things. The difference between formative evaluation and action research centers on points in time. There is no intention, typically, to generalize beyond those specific settings. Both formative evaluation and action research focus on specific programs at specific points in time. There is no intention, typically, to generalize beyond those specific settings. The difference between formative evaluation and action research centers on the extent to which the research is systematic, the different kinds of problems studied, and the extent to which there is a special role for the researcher as distinct from the people being researched.

which the research is undertaken. Both formative evaluation and action research focus on specific programs at specific points in time. There is no intention, typically, to generalize beyond those specific settings. The difference between formative evaluation and action research centers on the extent to which the research is systematic, the different kinds of problems studied, and the extent to which there is a special role for the researcher as distinct from the people being researched.

Action-Oriented, Problem-Solving Research

The final category along the theory-to-action continuum is action research. Action research aims at solving specific problems within a program, organization, or community. Action research explicitly and purposefully becomes part of the change process by engaging the people in the program or organization in studying their own problems in order to solve those problems (Whyte 1989). As a result, the distinction between research and action becomes quite blurred and the research methods tend to be less systematic, more informal, and quite specific to the problem, people, and organization for which the research is undertaken.

As a result, the distinction between research and action becomes quite blurred and the research methods tend to be less systematic, more informal, and quite specific to the problem, people, and organization for which the research is undertaken.

INACTION RESEARCH

Educational researcher, Ronald Gentile (1994) noticed the increasing popularity and importance of action research and wondered about the value of all the work involved: "While there are some advantages of action research, the disadvantages are that the researcher must collect a lot of data, carefully observing both the behaviors of interest and the conditions under which they occur. Following that, one has to score the data, a process that sometimes requires inventing ways to categorize them, analyze them, and draw inferences that are appropriate for the sample, design, and so forth. The problem with action research, in other words, is that it requires too much action. Fortunately, I discovered an alternative to action research that is probably best labeled 'inaction research'" (p. 30).

Three examples of Inaction Research:

Statistics in service of the id—"The invention or use of statistics to support some preconceived belief or entrepreneurial motive" (p. 30).

Scholarship as inaction research—Making up quotations where the researcher knows what the interviewee really meant to say but didn't say quite right, so the inaction researcher helps out by making up the needed quotation; also, "inventing new terminology that, by definition, has no historical usage" so no literature review is needed because there is no literature (p. 31).

Happiness quotients—"Exposing people to a program or product and then following up with a one-page questionnaire asking how well they liked the program or product" thereby giving the appearance of having done an evaluation without all the cost, inconvenience and difficulties of conducting real fieldwork (p. 31).

tion, thereby solving the town board's problem of what to do with the building—an example of action-oriented, problem-solving research.

The Purpose of Purpose Distinctions

It is important to understand variations in purpose along this theory-to-action continuum because different purposes typically lead to different ways of conceptualizing problems, different designs, different types of data gathering, and different ways of publishing and disseminating findings. These are only partly issues of scholarship. Politics, paradigms, and values are also part of the landscape.

Researchers engaged in inquiry at various points along the continuum can have very strong opinions and feelings about researchers at other points along the continuum, sometimes generating opposing opinions and strong emotions. Basic and applied researchers, for example, would often dispute even calling for formative and action research by the name *research*. The standards that basic researchers apply to what they would consider good research excludes even some applied research because it may not manifest the conceptual clarity and theoretical rigor in real-world situations that basic researchers value. Formative and action researchers, on the other hand, may attack basic research for being esoteric and irrelevant.

those choices for both the kind of research undertaken and the researcher's status as a professional within various social groups. Exhibit 5.3 summarizes some of the major differences among the different kinds of re-

Examples of Types

of Research Questions:

A Family Research Example

To further clarify these distinctions, it may be helpful to take a particular issue and look at how it would be approached for each type of research. For illustrative purposes, let's examine the different kinds of questions that can be asked about families for different research purposes. All of the research questions in Exhibit 5.4 focus on families, but the purpose and focus of each type of research are quite different. With clarity about purpose, it is possible to turn to consideration of specific design alternatives and strategies. Clarity about purpose helps in making decisions about critical trade-offs in research and designs, our next topic.

Critical Trade-Offs in Design

Purposes, strategies, and trade-offs—these themes go together. A discussion of design strategies and trade-offs is necessitated by the fact that there are no perfect research designs. There are always trade-offs. Limited resources, limited time, and limits on the human ability to grasp the complex nature of social reality necessitate trade-offs. The very first trade-offs come in framing the research or evaluation questions to be studied. The problem here is to determine the extent to which it is desirable to study one or a few questions in great depth or to

Debates about the meaningfulness, rigor, significance, and relevance of various approaches to research are regular features of university life. On the whole, within universities and among scholars, the status hierarchy in science attributes the highest status to basic research, secondary status to applied research, little status to summative evaluation research, and virtually no status to formation research, and virtually no status to formative and action research. The status hierarchy is reversed in real-world settings, where people with problems attribute the greatest significance to action and formative research that can help them solve their problems in a timely way and attach the least importance to basic research, which they consider remote and largely irrelevant to what they are doing on a day-to-day basis.

The distinctions along the continuum are not only distinctions about purpose and how one conducts research, but they also involve the issue of what one calls what one does. In other words, a person conducting basic research for the purpose of contributing to theory within a discipline may find it helpful to call that work *applied research* to get certain kinds of funding. Summative evaluation researchers may describe what they are doing as *formative evaluation* to make their work more acceptable to program staff resistant to being studied. On the other hand, applied researchers may call what they are doing *basic research* to increase its acceptability among scholars.

In short, there are no clear lines dividing the points along the continuum. Part of what determines where a particular kind of research falls along the continuum is how the researcher describes what is being done and its purpose. Different reviewers of the same piece of research might well use a different label to describe it. What is important for our purposes is that researchers understand the implications of these distinctions, the choices involved, and the implications of

EXHIBIT 5.3 A Typology of Research Purposes

Types of Research	Purpose	Focus of Research	Desired Results	Desired Level of Generalization	Key Assumptions	Publication Mode	Standard for Judging
Basic research	Knowledge as an end in itself; discover truth	Questions deemed important by one's discipline or personal intellectual interest	Contribution to theory	Across time and space (ideal)	The world is patterned; those patterns are knowable and explainable.	Major refereed scholarly journals in one's discipline, scholarly books	Rigor of research, universality and verifiability of theory
Applied research	Understand the nature and sources of human and societal problems	Questions deemed important by society	Contributions to theories that can be used to formulate problem-solving programs and interventions	Within as general a time and space as possible, but clearly limited application context	Human and societal problems can be understood and solved with knowledge.	Specialized academic journals, applied research journals within disciplines, interdisciplinary problem-focused journals	Rigor and theoretical insight into the problem
Summative evaluation	Determine effectiveness of human interventions and actions (programs, policies, personnel, products)	Goals of the intervention	Judgments and generalizations about effective types of interventions and the conditions under which those efforts are effective	All interventions with similar goals	What works one place under specified conditions should work elsewhere.	Evaluation reports for program funders and policymakers, specialized journals	Generalizability to future efforts and to other programs and policy issues
Formative evaluation	Improve an intervention: A program, policy, organization, or product	Strengths and weaknesses of the specific program, policy, product, or personnel being studied	Recommendations for improvements	Limited to specific setting studied	People can and will use information to improve what they're doing.	Oral briefings; conferences; internal report; limited circulation to similar programs, other evaluators	Usefulness to and actual use by intended users in the setting studied
Action research	Solve problems in a program, organization, or community	Organization and community problems	Immediate action; solving problems as quickly as possible	Here and now	People in a setting can solve problems by studying themselves.	Interpersonal interactions among research participants; informal unpublished	Feelings about the process among research participants, feasibility of the solution generated

Family Research Example: Research Questions Matched to Research Category

EXHIBIT 5.4

Basic research	What are variations in types of families and what functions do those variations serve?
Applied research	What is the divorce rate among different kinds of families in the United States and what explains different rates of divorce among different groups?
Summative evaluation	What is the overall effectiveness of a publicly funded educational program that teaches family members communication skills where the desired program outcomes are enhanced communications among family members, a greater sense of satisfaction with family life, effective parenting practices, and reduced risk of divorce?
Formative evaluation	How can a program teaching family communications skills be improved? What are the program's strengths and weaknesses? What do participants like and dislike?
Action research	A self-study in a particular organization (e.g., church, neighborhood center) to figure out what activities would be attractive to families with children of different ages to solve the problem of low participation in family activities.

study many questions but in less depth—the “boundary problem” in naturalistic inquiry (Guba 1978). Once a potential set of inquiry questions has been generated, it is necessary to begin the process of prioritizing those questions to decide which of them ought to be pursued. For example, for an evaluation, should all parts of the program be studied or only certain parts? Should all clients be interviewed or only some subset of clients? Should the evaluator aim at describing all program processes or only certain selected processes in depth? Should all outcomes be examined or only certain outcomes of particular interest to inform a pending decision? These are questions that are discussed and negotiated with intended users of the evaluation. In basic research, these kinds of questions are resolved by the nature of the theoretical contribution to be made. In dissertation research, the doctoral committee provides guidance on focusing questions.

And always there are fundamental constraints of time and resources. Converging on focused priorities typically proves more difficult than the challenge of generating potential questions at the beginning of a study or evaluation. Doctoral students can be especially adept at avoiding focus, conceiving instead of sweeping, comprehensive studies that make the whole world their fieldwork system. In evaluations, once involved users begin to take seriously the notion that they can learn from finding how whether what they think is being accomplished by a program is really being accomplished, they soon generate a long list of things they'd like to find out. The evaluation facilitator's role is to help them move from a rather extensive list of potential questions to a much shorter list of realistically possible questions and finally to a focused list of essential and necessary questions.

Review of relevant literature can also bring focus to a study. What is already known? Unknown? What are the cutting-edge theoretical issues? Yet, reviewing the literature can present a quandary in qualitative inquiry because it may bias the researcher's thinking and reduce openness to whatever emerges in the field. Thus, sometimes a literature review may not take place until after data collection. Alternatively, the literature review may go on simultaneously with fieldwork, permitting a creative interplay among the processes of data collection, literature review, and researcher introspection (Marshall and Rossman 1989:38-40). As with other qualitative design issues, trade-offs appear at every turn, for there are decided advantages and disadvantages to reviewing the literature before, during, or after fieldwork—or on a continual basis throughout the study.

A specific example of possible variations in focus will illustrate the kinds of trade-offs involved in designing a study. Suppose some educators are interested in studying how a school program affects the social development of school-age children. They want to know how the interactions of children with others in the school setting contribute to the development of social skills. They believe that those social skills will be different for different children, and they are not sure of the range of social interactions that may occur, so they are interested in a qualitative inquiry that will capture variations in program experience and relate those experiences to individualized outcomes. What, then, are trade-offs in determining the final focus?

We begin with the fact that any given child has social interactions with a great many people. The first problem in focusing, then, is to determine how much of the social reality experienced by children we should

attempt to study. In a narrowly focused study, we might select one particular set of interactions and limit our study to those teachers and children. Broadening the scope somewhat, we might decide to look only at those interactions that occur in the classroom, thereby increasing the scope of the study to include interactions not only between teacher and child but also among peers in the classroom and between any visitors and visitors to the classroom and children. Broadening the scope of the study still more, we might decide to look at all the social relationships that children experience in schools; in this case we would move beyond the classroom to look at interactions with other personnel in the school—fore example, the librarian, school counselors, social subject teachers, the custodian, and school administrative staff. Broadening the scope of the study still further, the educational might decide that it is important to look at the social relationships children experience at home as well as at school so as to understand better how children experience and are affected by both settings, so we would include in our design interactions with parents, siblings, and others in the home. Finally, one might look at the social relationships experienced throughout the full range of societal contacts that children have, including church, clubs, and even mass media contacts.

A case could be made for the importance and value of any of these approaches, from the narrowest focus, looking only at students' teacher interactions, to the broadest focus, looking at students' full, complex social world. Now let's add the real-world constraint of limited resources—say, \$50,000 and three months—to conduct the study. At some level, any of these research endeavors could be undertaken for \$50,000. But it be-