

Directorate for Education
and Human Resources



Division of Research,
Evaluation and Communication
National Science Foundation

User-Friendly Handbook for

Mixed Method Evaluations

The Foundation provides awards for research in the sciences and engineering. The awardee is wholly responsible for the conduct of such research and the preparation of the results for publication. The Foundation, therefore, does not assume responsibility for the research findings or their interpretation.

The Foundation welcomes proposals from all qualified scientists and engineers, and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and related programs described here.

In accordance with federal statutes, regulations, and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, denied the benefits of, or be subject to discrimination under any program or activity receiving financial assistance from the National Science Foundation.

Facilitation Awards for Scientists and Engineers with Disabilities (FASSED) provide funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on an NSF project. See the program announcement or contact the program coordinator at (703) 306-1636.

Privacy Act and Public Burden

The information requested on proposal forms is solicited under the authority of the National Science Foundation Act of 1950, as amended. It will be used in connection with the selection of qualified proposals and may be disclosed to qualified reviewers and staff assistants as part of the review process; to applicant institutions/grantees; to provide or obtain data regarding the application review process, award decisions, or the administration of awards; to government contractors, experts, volunteers, and researchers as necessary to complete assigned work; and to other government agencies in order to coordinate programs. See Systems of Records, NSF-50, Principal Investigators/Proposal File and Associated Records, and NSF-51, 60 Federal Register 4449 (January 23, 1995). Reviewer/Proposal File and Associated Records, 59 Federal Register 8031 (February 17, 1994). Submission of the information is voluntary. Failure to provide full and complete information, however, may reduce the possibility of your receiving an award.

Public reporting burden for this collection of information is estimated to average 120 hours per response, including the time for reviewing instructions. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Herman G. Fleming, Reports Clearance Officer, Contracts, Policy, and Oversight, National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230.

The National Science Foundation has TTD (Telephonic Device for the Deaf) capability, which enables individuals with hearing impairment to communicate with the Foundation about NSF programs, employment, or general information. This number is (703) 306-0090.

ACKNOWLEDGMENTS

Appreciation is expressed to our external advisory panel Dr. Frances Lawrenz, Dr. Jennifer Greene, Dr. Mary Ann Millsap, and Dr. Steve Dietz for their comprehensive reviews of this document and their helpful suggestions. We also appreciate the direction provided by Dr. Conrad Katzenmeyer and Mr. James Dietz of the Division of Research, Evaluation and Communication.

User-Friendly Handbook for Mixed Method Evaluations

Edited by

Joy Frechtling
Laure Sharp
Westat, Inc.

August 1997

NSF Program Officer
Conrad Katzenmeyer

Directorate for Education
and Human Resources



Division of Research,
Evaluation and Communication

This handbook was developed with support from the National Science Foundation RED 94-52965.

Table of Contents

Part I.	Introduction to Mixed Method Evaluations	Page
1	Introducing This Handbook <i>(Laure Sharp and Joy Frechtling)</i>	1-1
	The Need for a Handbook on Designing and Conducting Mixed Method Evaluations.....	1-1
	Key Concepts and Assumptions.....	1-2
2	Illustration: A Hypothetical Project <i>(Laure Sharp)</i>	2-1
	Project Title.....	2-1
	Project Description.....	2-1
	Project Goals as Stated in the Grant Application to NSF.....	2-2
	Overview of the Evaluation Plan.....	2-3
Part II.	Overview of Qualitative Methods and Analytic Techniques	
3	Common Qualitative Methods <i>(Colleen Mahoney)</i>	3-1
	Observations.....	3-1
	Interviews.....	3-5
	Focus Groups.....	3-9
	Other Qualitative Methods.....	3-13
	Appendix A: Sample Observation Instrument.....	A-1
	Appendix B: Sample Indepth Interview Guide.....	B-1
	Appendix C: Sample Focus Group Topic Guide.....	C-1
4	Analyzing Qualitative Data..... <i>(Suzanne Berkowitz)</i>	4-1
	What Is Qualitative Analysis?.....	4-1
	Processes in Qualitative Analysis.....	4-3
	Summary: Judging the Quality of Qualitative Analysis.....	4-17
	Practical Advice in Conducting Qualitative Analyses.....	4-19

Table of Contents (continued)

Part III.	Designing and Reporting Mixed Method Evaluations	Page
5	Overview of the Design Process for Mixed Method Evaluation <i>(Laure Sharp and Joy Frechtling)</i>	5-1
	Developing Evaluation Questions.....	5-2
	Selecting Methods for Gathering the Data: The Case for Mixed Method Designs.....	5-9
	Other Considerations in Designing Mixed Method Evaluations	5-10
6	Evaluation Design for the Hypothetical Project	6-1
	<i>(Laure Sharp)</i>	
	Step 1. Develop Evaluation Questions.....	6-1
	Step 2. Determine Appropriate Data Sources and Data Collection Approaches to Obtain Answers to the Final Set of Evaluation Questions.....	6-3
	Step 3. Reality Testing and Design Modifications: Staff Needs, Costs, Time Frame Within Which All Tasks (Data Collection, Data Analysis, and Reporting Writing) Must Be Completed.....	6-8
7	Reporting the Results of Mixed Method Evaluations..... <i>(Gary Silverstein and Laure Sharp)</i>	7-1
	Ascertaining the Interests and Needs of the Audience.....	7-2
	Organizing and Consolidating the Final Report.....	7-4
	Formulating Sound Conclusions and Recommendations.....	7-7
	Maintaining Confidentiality	7-8
	Tips for Writing Good Evaluation Reports.....	7-10
Part IV.	Supplementary Materials	
8	Annotated Bibliography	8-1
9	Glossary.....	9-1

Table of Contents

List of Exhibits

Exhibit		Page
1	Common techniques	1-4
2	Example of a mixed method design	1-9
3	Advantages and disadvantages of observations	3-2
4	Types of data for which observations are a good source	3-3
5	Advantages and disadvantages of indepth interviews.....	3-7
6	Considerations in conducting indepth interviews and focus groups.....	3-8
7	Which to use: Focus groups or indepth interviews?	3-11
8	Advantages and disadvantages of document studies.....	3-14
9	Advantages and disadvantages of using key informants.....	3-15
10	Data matrix for Campus A: What was done to share knowledge	4-7
11	Participants' views of information sharing at eight campuses.....	4-9
12	Matrix of cross-case analysis linking implementation and outcome factors ...	4-17
13	Goals, stakeholders, and evaluation questions for a formative evaluation	6-2
14	Goals, stakeholders, and evaluation questions for a summative evaluation	6-3
15	Evaluation questions, data sources, and data collection methods for a formative evaluation.....	6-5
16	Evaluation questions, data sources, and data collection methods for a summative evaluation.....	6-6
17	First data collection plan	6-7
18	Final data collection plan	6-9
19	Matrix of stakeholders.....	7-3
20	Example of an evaluation/methodology matrix	7-6

PART I.

INTRODUCTION TO

MIXED METHOD

EVALUATIONS

Chapter 1 INTRODUCING THIS HANDBOOK

The Need for a Handbook on Designing and Conducting Mixed Method Evaluations

Evaluation of the progress and effectiveness of projects funded by the National Science Foundation's (NSF) Directorate for Education and Human Resources (EHR) has become increasingly important. Project staff, participants, local stakeholders, and decisionmakers need to know how funded projects are contributing to knowledge and understanding of mathematics, science, and technology. To do so, some simple but critical questions must be addressed:

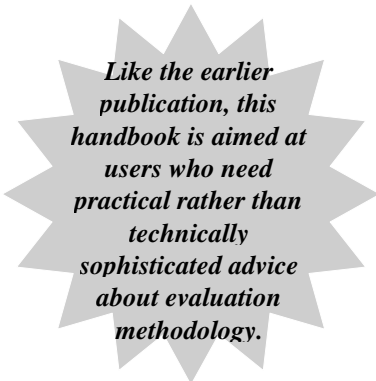
- What are we finding out about teaching and learning?
- How can we apply our new knowledge?
- Where are the dead ends?
- What are the next steps?

Although there are many excellent textbooks, manuals, and guides dealing with evaluation, few are geared to the needs of the EHR grantee who may be an experienced researcher but a novice evaluator. One of the ways that EHR seeks to fill this gap is by the publication of what have been called "user-friendly" handbooks for project evaluation.

The first publication, *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education*, issued in 1993, describes the types of evaluations principal investigators/project directors (PIs/PDs) may be called upon to perform over the lifetime of a project. It also describes in some detail the evaluation process, which includes the development of evaluation questions and the collection and analysis of appropriate data to provide answers to these questions. Although this first handbook discussed both qualitative and quantitative methods, it

covered techniques that produce numbers (quantitative data) in greater detail. This approach was chosen because decisionmakers usually demand quantitative (statistically documented) evidence of results. Indicators that are often selected to document outcomes include percentage of targeted populations participating in mathematics and science courses, test scores, and percentage of targeted populations selecting careers in the mathematics and science fields.

The current handbook, *User-Friendly Guide to Mixed Method Evaluations*, builds on the first but seeks to introduce a broader perspective. It was initiated because of the recognition that by focusing primarily on quantitative techniques, evaluators may miss important parts of a story. Experienced evaluators have found that most often the best results are achieved through the use of mixed method evaluations, which combine quantitative and qualitative techniques. Because the earlier handbook did not include an indepth discussion of the collection and analysis of qualitative data, this handbook was initiated to provide more information on qualitative techniques and discuss how they can be combined effectively with quantitative measures.



Like the earlier publication, this handbook is aimed at users who need practical rather than technically sophisticated advice about evaluation methodology.

Like the earlier publication, this handbook is aimed at users who need practical rather than technically sophisticated advice about evaluation methodology. The main objective is to make PIs and PDs "evaluation smart" and to provide the knowledge needed for planning and managing useful evaluations.

Key Concepts and Assumptions

Why Conduct an Evaluation?

There are two simple reasons for conducting an evaluation:

- To gain direction for improving projects as they are developing, and
- To determine projects' effectiveness after they have had time to produce results.

Formative evaluations (which include implementation and process evaluations) address the first set of issues. They examine the development of the project and may lead to changes in the way the project is structured and carried out. Questions typically asked include:

- To what extent do the activities and strategies match those described in the plan? If they do not match, are the changes in the activities justified and described?
- To what extent were the activities conducted according to the proposed timeline? By the appropriate personnel?
- To what extent are the actual costs of project implementation in line with initial budget expectations?
- To what extent are the participants moving toward the anticipated goals of the project?
- Which of the activities or strategies are aiding the participants to move toward the goals?
- What barriers were encountered? How and to what extent were they overcome?

Summative evaluations (also called outcome or impact evaluations) address the second set of issues. They look at what a project has actually accomplished in terms of its stated goals. Summative evaluation questions include:

- To what extent did the project meet its overall goals?
- Was the project equally effective for all participants?
- What components were the most effective?
- What significant unintended impacts did the project have?
- Is the project replicable and transportable?

For each of these questions, both quantitative data (data expressed in numbers) and qualitative data (data expressed in narratives or words) can be useful in a variety of ways.

The remainder of this chapter provides some background on the differing and complementary nature of quantitative and qualitative evaluation methodologies. The aim is to provide an overview of the advantages and disadvantages of each, as well as an idea of some of the more controversial issues concerning their use.

Before doing so, however, it is important to stress that there are many ways of performing project evaluations, and that there is no recipe or formula that is best for every case. Quantitative and qualitative methods each have advantages and drawbacks when it comes to an evaluation's design, implementation, findings, conclusions, and

utilization. The challenge is to find a judicious balance in any particular situation. According to Cronbach (1982),

There is no single best plan for an evaluation, not even for an inquiry into a particular program at a particular time, with a particular budget.

What Are the Major Differences Between Quantitative and Qualitative Techniques?

As shown in Exhibit 1, quantitative and qualitative measures are characterized by different techniques for data collection.


Exhibit 1. Common techniques	
Quantitative	Qualitative
Questionnaires	Observations
Tests	Interviews
Existing databases	Focus groups

Aside from the most obvious distinction between numbers and words, the conventional wisdom among evaluators is that qualitative and quantitative methods have different strengths, weaknesses, and requirements that will affect evaluators' decisions about which methodologies are best suited for their purposes. The issues to be considered can be classified as being primarily **theoretical** or **practical**.

Theoretical issues. Most often, these center on one of three topics:

- The value of the types of data;
- The relative scientific rigor of the data; or
- Basic, underlying philosophies of evaluation.

Value of the data. Quantitative and qualitative techniques provide a tradeoff between breadth and depth and between generalizability and targeting to specific (sometimes very limited) populations. For example, a sample survey of high school students who participated in a special science enrichment program (a quantitative technique) can yield representative and broadly generalizable information about the proportion of participants who plan to major in science when they get to college and how this proportion differs by gender. But at best, the survey can elicit only a few, often superficial reasons for this gender difference. On the other hand, separate focus groups (a qualitative



Quantitative and qualitative techniques provide a tradeoff between breadth and depth.

technique) conducted with small groups of male and female students will provide many more clues about gender differences in the choice of science majors and the extent to which the special science program changed or reinforced attitudes. But this technique may be limited in the extent to which findings apply beyond the specific individuals included in the focus groups.

Scientific rigor. Data collected through quantitative methods are often believed to yield more objective and accurate information because they were collected using standardized methods, can be replicated, and, unlike qualitative data, can be analyzed using sophisticated statistical techniques. In line with these arguments, traditional wisdom has held that qualitative methods are most suitable for formative evaluations, whereas summative evaluations require "hard" (quantitative) measures to judge the ultimate value of the project.

This distinction is too simplistic. Both approaches may or may not satisfy the canons of scientific rigor. Quantitative researchers are becoming increasingly aware that some of their data may not be accurate and valid, because some survey respondents may not understand the meaning of questions to which they respond, and because people's recall of even recent events is often faulty. On the other hand, qualitative researchers have developed better techniques for classifying and analyzing large bodies of descriptive data. It is also increasingly recognized that all data collection—quantitative and qualitative—operates within a cultural context and is affected to some extent by the perceptions and beliefs of investigators and data collectors.

Philosophical distinction. Some researchers and scholars differ about the respective merits of the two approaches largely because of different views about the nature of knowledge and how knowledge is best acquired. Many qualitative researchers argue that there is no objective social reality, and that all knowledge is "constructed" by observers who are the product of traditions, beliefs, and the social and political environment within which they operate. And while quantitative researchers no longer believe that their research methods yield absolute and objective truth, they continue to adhere to the scientific model and seek to develop increasingly sophisticated techniques and statistical tools to improve the measurement of social phenomena. The qualitative approach emphasizes the importance of understanding the context in which events and outcomes occur, whereas quantitative researchers seek to control the context by using random assignment and multivariate analyses. Similarly, qualitative researchers believe that the study of deviant cases provides important insights for the interpretation of findings; quantitative researchers tend to ignore the small number of deviant and extreme cases.

This distinction affects the nature of research designs. According to its most orthodox practitioners, qualitative research does not start with narrowly specified evaluation questions; instead, specific questions are formulated after open-ended field research has been completed (Lofland and Lofland, 1995). This approach may be difficult for program and project evaluators to adopt, since specific questions about the effectiveness of interventions being evaluated are usually expected to guide the evaluation. Some researchers have suggested that a distinction be made between Qualitative and qualitative work: Qualitative work (large Q) refers to methods that eschew prior evaluation questions and hypothesis testing, whereas qualitative work (small q) refers to open-ended data collection methods such as indepth interviews embedded in structured research (Kidder and Fine, 1987). The latter are more likely to meet EHR evaluators' needs.

Practical issues. On the practical level, there are four issues which can affect the choice of method:

- Credibility of findings;
- Staff skills;
- Costs; and
- Time constraints.

Credibility of findings. Evaluations are designed for various audiences, including funding agencies, policymakers in governmental and private agencies, project staff and clients, researchers in academic and applied settings, as well as various other "stakeholders" (individuals and organizations with a stake in the outcome of a project). Experienced evaluators know that they often deal with skeptical audiences or stakeholders who seek to discredit findings that are too critical or uncritical of a project's outcomes. For this reason, the evaluation methodology may be rejected as unsound or weak for a specific case.

The major stakeholders for EHR projects are policymakers within NSF and the federal government, state and local officials, and decisionmakers in the educational community where the project is located. In most cases, decisionmakers at the national level tend to favor quantitative information because these policymakers are accustomed to basing funding decisions on numbers and statistical indicators. On the other hand, many stakeholders in the educational community are often skeptical about statistics and "number crunching" and consider the richer data obtained through qualitative research to be more trustworthy and informative. A particular case in point is the use of traditional test results, a favorite outcome criterion

for policymakers, school boards, and parents, but one that teachers and school administrators tend to discount as a poor tool for assessing true student learning.

Staff skills. Qualitative methods, including indepth interviewing, observations, and the use of focus groups, require good staff skills and considerable supervision to yield trustworthy data. Some quantitative research methods can be mastered easily with the help of simple training manuals; this is true of small-scale, self-administered questionnaires, where most questions can be answered by yes/no checkmarks or selecting numbers on a simple scale. Large-scale, complex surveys, however, usually require more skilled personnel to design the instruments and to manage data collection and analysis.

Costs. It is difficult to generalize about the relative costs of the two methods; much depends on the amount of information needed, quality standards followed for the data collection, and the number of cases required for reliability and validity. A short survey based on a small number of cases (25-50) and consisting of a few “easy” questions would be inexpensive, but it also would provide only limited data. Even cheaper would be substituting a focus group session for a subset of the 25-50 respondents; while this method might provide more “interesting” data, those data would be primarily useful for generating new hypotheses to be tested by more appropriate qualitative or quantitative methods. To obtain robust findings, the cost of data collection is bound to be high regardless of method.

Time constraints. Similarly, data complexity and quality affect the time needed for data collection and analysis. Although technological innovations have shortened the time needed to process quantitative data, a good survey requires considerable time to create and pretest questions and to obtain high response rates. However, qualitative methods may be even more time consuming because data collection and data analysis overlap, and the process encourages the exploration of new evaluation questions (see Chapter 4). If insufficient time is allowed for the evaluation, it may be necessary to curtail the amount of data to be collected or to cut short the analytic process, thereby limiting the value of the findings. For evaluations that operate under severe time constraints—for example, where budgetary decisions depend on the findings—the choice of the best method can present a serious dilemma.

In summary, the debate over the merits of qualitative versus quantitative methods is ongoing in the academic community, but when it comes to the choice of methods for conducting project evaluations, a pragmatic strategy has been gaining increased support. Respected practitioners have argued for integrating the two

approaches building on their complimentary strengths.¹ Others have stressed the advantages of linking qualitative and quantitative methods when performing studies and evaluations, showing how the validity and usefulness of findings will benefit (Miles and Huberman, 1994).

Why Use a Mixed Method Approach?

The assumption guiding this handbook is that a strong case can be made for using an approach that combines quantitative and qualitative elements in most evaluations of EHR projects. We offer this assumption because most of the interventions sponsored by EHR are not introduced into a sterile laboratory, but rather into a complex social environment with feature that affect the success of the project. To ignore the complexity of the background is to impoverish the evaluation. Similarly, when investigating human behavior and attitudes, it is most fruitful to use a variety of data collection methods (Patton, 1990). By using different sources and methods at various points in the evaluation process, the evaluation team can build on the strength of each type of data collection and minimize the weaknesses of any single approach. A multimethod approach to evaluation can increase both the validity and reliability of evaluation data.

The range of possible benefits that carefully crafted mixed method designs can yield has been conceptualized by a number of evaluators.²

- The validity of results can be strengthened by using more than one method to study the same phenomenon. This approach—called **triangulation**—is most often mentioned as the main advantage of the mixed method approach.
- Combining the two methods pays off in improved instrumentation for all data collection approaches and in sharpening the evaluator's understanding of findings. A typical design might start out with a qualitative segment such as a focus group discussion, which will alert the evaluator to issues that should be explored in a survey of program participants, followed by the survey, which in turn is followed by indepth interviews to clarify some of the survey findings (Exhibit 2).

¹See especially the article by William R. Shadish in *Program Evaluation: A Pluralistic Enterprise*, New Directions for Program Evaluation, No. 60 (San Francisco: Jossey Bass. Winter 1993).

² For a full discussion of this topic, see Jennifer C. Greene, Valerie J. Caracelli, and Wendy F. Graham, Toward a Conceptual Framework for Mixed Method Evaluation Designs, *Educational Evaluation and Policy Analysis*, Vol. 11, No. 3, (Fall 1989), pp. 255-274.

Exhibit 2.		
Example of a mixed method design		
qualitative	quantitative	qualitative
(exploratory focus group)	(questionnaire)	(personal interview with subgroup)

But this sequential approach is only one of several that evaluators might find useful (Miles and Huberman, 1994). Thus, if an evaluator has identified subgroups of program participants or specific topics for which indepth information is needed, a limited qualitative data collection can be initiated while a more broad-based survey is in progress.

- A mixed method approach may also lead evaluators to modify or expand the evaluation design and/or the data collection methods. This action can occur when the use of mixed methods uncovers inconsistencies and discrepancies that alert the evaluator to the need for reexamining the evaluation framework and/or the data collection and analysis procedures used.

There is a growing consensus among evaluation experts that both qualitative and quantitative methods have a place in the performance of effective evaluations. Both formative and summative evaluations are enriched by a mixed method approach.

How To Use This Handbook

This handbook covers a lot of ground, and not all readers will want to read it from beginning to end. For those who prefer to sample sections, some organizational features are highlighted below.

- To provide practical illustrations throughout the handbook, we have invented a hypothetical project, which is summarized in the next chapter (Part 1, Chapter 2); the various stages of the evaluation design for this project will be found in Part 3, Chapter 6. These two chapters may be especially useful for evaluators who have not been involved in designing evaluations for major, multisite EHR projects.
- Part 2, Chapter 3 focuses on qualitative methodologies, and Chapter 4 deals with analysis approaches for qualitative data.

These two chapters are intended to supplement the information on quantitative methods in the previous handbook.

- Part 3, Chapters 5, 6, and 7 cover the basic steps in developing a mixed method evaluation design and describes ways of reporting findings to NSF and other stakeholders.
- Part 4 presents supplementary material, including an annotated bibliography and a glossary of common terms.

Before turning to these issues, however, we present the hypothetical NSF project that is used as an anchoring point for discussing the issues presented in the subsequent chapters.

References

- Cronbach, L. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Kidder, L., and Fine, M. (1987). *Qualitative and Quantitative Methods: When Stories Converge. Multiple Methods in Program Evaluation*. New Directions for Program Evaluation, No. 35. San Francisco: Jossey-Bass.
- Lofland, J., and Lofland, L.H. (1995). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Belmont, CA: Wadsworth Publishing Company.
- Miles, M.B., and Huberman, A.M. (1994). *Qualitative Data Analysis*, 2nd Ed. Newbury Park, CA: Sage, p. 40-43.
- National Science Foundation. (1993). *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education*. NSF 93-152. Arlington, VA: NSF.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*, 2nd Ed. Newbury Park, CA: Sage.

ILLUSTRATION: A HYPOTHETICAL PROJECT

Project Title

Undergraduate Faculty Enhancement: Introducing faculty in state universities and colleges to new concepts and methods in preservice mathematics instruction.

Project Description

In response to the growing national concern about the quality of American elementary and secondary education and especially about students' achievement in mathematics and science, considerable efforts have been directed at enhancing the skills of the teachers in the labor force through **inservice** training. Less attention has been focused on **preservice** training, especially for elementary school teachers, most of whom are educated in departments and schools of education. In many institutions, faculty members who provide this instruction need to become more conversant with the new standards and instructional techniques for the teaching of mathematics in elementary schools.

The proposed pilot project was designed to examine a strategy for meeting this need. The project attempts to improve preservice education by giving the faculty teaching courses in mathematics to future elementary school teachers new knowledge, skills, and approaches for incorporation into their instruction. In the project, the investigators ascertain the extent of faculty members' knowledge about standards-based instruction, engage them in expanding their understanding of standards-based reform and the instructional approaches that support high-quality teaching; and assess the extent to which the strategies emphasized and demonstrated in the pilot project are transferred to the participants' own classroom practices.

The project is being carried out on the main campus of a major state university under the leadership of the Director of the Center for Educational Innovation. Ten day-long workshops will be offered to two cohorts of faculty members from the main campus and branch campuses. These workshops will be supplemented by opportunities for networking among participating faculty members and the exchange of experiences and recommendations during a summer session following the academic year. The workshops are based on an integrated plan for reforming undergraduate education for future elementary teachers. The focus of the workshops is to provide carefully articulated information and practice on current approaches to mathematics instruction (content and pedagogy) in elementary grades, consistent with state frameworks and standards of excellence. The program uses and builds on the knowledge of content experts, master practitioners, and teacher educators.

The following strategies are being employed in the workshops: presentations, discussions, hands-on experiences with various traditional and innovative tools, coaching, and videotaped demonstrations of model teaching. The summer session is offered for sharing experiences, reflecting on successful and unsuccessful applications, and constructing new approaches. In addition, participants are encouraged to communicate with each other throughout the year via e-mail. Project activities are funded for 2 years and are expected to support two cohorts of participants; funding for an additional 6-month period to allow performance of the summative evaluation has been included.

Participation is limited to faculty members on the main campus and in the seven 4-year branch campuses of the state university where courses in elementary mathematics education are offered. Participants are selected on the basis of a written essay and a commitment to attend all sessions and to try suggested approaches in their classroom. A total of 25 faculty members are to be enrolled in the workshops each year. During the life of the project, roughly 1,000 undergraduate students will be enrolled in classes taught by the participating faculty members.

Project Goals as Stated in the Grant Application to NSF

As presented in the grant application, the project has four main goals:

- To further the knowledge of college faculty with respect to new concepts, standards, and methods for mathematics education in elementary schools;

- To enable and encourage faculty members to incorporate these approaches in their own classrooms activities and, hopefully, into the curricula of their institutions;
- To stimulate their students' interest in teaching mathematics and in using the new techniques when they become elementary school teachers; and
- To test a model for achieving these goals.

Overview of the Evaluation Plan

A staff member of the Center for Educational Innovation with prior evaluation experience was assigned responsibility for the evaluation. She will be assisted by undergraduate and graduate students. As required, consultation will be provided by members of the Center's statistical and research staff and by faculty members on the main campus who have played leadership roles in reforming mathematics education.

A formative (progress) evaluation will be carried out at the end of the first year. A summative (outcome) evaluation is to be completed 6 months after project termination. Because the project was conceived as a prototype for future expansion to other institutions, a thorough evaluation was considered an essential component, and the evaluation budget represented a higher-than-usual percentage of total costs (project costs were \$500,000, of which \$75,000 was allocated for evaluation).

The evaluation designs included in the application were specified only in general terms. The formative evaluation would look at the implementation of the program and be used for identifying its strengths and weaknesses. Suggested formative evaluation questions included the following:

- Were the workshops delivered and staffed as planned? If not, what were the reasons?
- Was the workshop content (disciplinary and pedagogical) accurate and up to date?
- Did the instructors communicate effectively with participants, stimulate questions, and encourage all participants to take part in discussions?
- Were appropriate materials available?

- Did the participants have the opportunity to engage in inquiry-based activities?
- Was there an appropriate balance of knowledge building and application?

The summative evaluation was intended to document the extent to which participants introduced changes in their classroom teaching and to determine which components of the workshops were especially effective in this respect. The proposal also promised to investigate the impact of the workshops on participating faculty members, especially on their acquisition of knowledge and skills. Furthermore, the impact on other faculty members, on the institution, and on students was to be part of the evaluation. Recommendations for replicating this project in other institutions, and suggestions for changes in the workshop content or administrative arrangements, were to be included in the summative evaluation. Proposed summative evaluation questions included the following:

- To what extent did the participants use what they were taught in their own instruction or activities? Which topics and techniques were most often (or least often) incorporated?
- To what extent did participants share their recently acquired knowledge and skills with other faculty? Which topics were frequently discussed? Which ones were not?
- To what extent was there an impact on the students of these teachers? Had they become more (or less) positive about making the teaching of elementary mathematics an important component of their future career?
- Did changes occur in the overall program of instruction offered to potential elementary mathematics teachers? What were the obstacles to the introduction of changes?

The proposal also enumerated possible data sources for conducting the evaluations, including self-administered questionnaires completed after each workshop, indepth interviews with knowledgeable informants, focus groups, observation of workshops, classroom observations, and surveys of students. It was stated that a more complete design for the formative and summative evaluations would be developed after contract award.

PART II.

OVERVIEW OF

QUALITATIVE **M**ETHODS

AND

ANALYTIC **T**ECHNIQUES

Chapter 3 COMMON QUALITATIVE METHODS

In this chapter we describe and compare the most common qualitative methods employed in project evaluations.³ These include observations, indepth interviews, and focus groups. We also cover briefly some other less frequently used qualitative techniques. Advantages and disadvantages are summarized. For those readers interested in learning more about qualitative data collection methods, a list of recommended readings is provided.

Observations

Observational techniques are methods by which an individual or individuals gather firsthand data on programs, processes, or behaviors being studied. They provide evaluators with an opportunity to collect data on a wide range of behaviors, to capture a great variety of interactions, and to openly explore the evaluation topic. By directly observing operations and activities, the evaluator can develop a holistic perspective, i.e., an understanding of the context within which the project operates. This may be especially important where it is not the event that is of interest, but rather how that event may fit into, or be impacted by, a sequence of events. Observational approaches also allow the evaluator to learn about things the participants or staff may be unaware of or that they are unwilling or unable to discuss in an interview or focus group.

When to use observations. Observations can be useful during both the formative and summative phases of evaluation. For example, during the formative phase, observations can be useful in determining whether or not the project is being delivered and operated as planned. In the hypothetical project, observations could be used to describe the faculty development sessions, examining the extent to which participants understand the concepts, ask the right questions, and are engaged in appropriate interactions. Such

³Information on common quantitative methods is provided in the earlier *User-Friendly Handbook for Project Evaluation* (NSF 93-152).

Exhibit 3. Advantages and disadvantages of observations
Advantages Provide direct information about behavior of individuals and groups Permit evaluator to enter into and understand situation/context Provide good opportunities for identifying unanticipated outcomes Exist in natural, unstructured, and flexible setting
Disadvantages Expensive and time consuming Need well-qualified, highly trained observers; may need to be content experts May affect behavior of participants Selective perception of observer may distort data Investigator has little control over situation Behavior or set of behaviors observed may be atypical

formative observations could also provide valuable insights into the teaching styles of the presenters and how they are covering the material.

Observations during the summative phase of evaluation can be used to determine whether or not the project is successful. The technique would be especially useful in directly examining teaching methods employed by the faculty in their own classes after program participation. Exhibits 3 and 4 display the advantages and disadvantages of observations as a data collection tool and some common types of data that are readily collected by observation.

Readers familiar with survey techniques may justifiably point out that surveys can address these same questions and do so in a less costly fashion. Critics of surveys find them suspect because of their reliance on self-report, which may not provide an accurate picture of what is happening because of the tendency, intentional or not, to try to give the “right answer.” Surveys also cannot tap into the contextual element. Proponents of surveys counter that properly constructed surveys with built in checks and balances can overcome these problems and provide highly credible data. This frequently debated issue is best decided on a case-by-case basis.

Recording Observational Data

Observations are carried out using a carefully developed set of steps and instruments. The observer is more than just an onlooker, but rather comes to the scene with a set of target concepts, definitions, and criteria for describing events. While in some studies observers may simply record and describe, in the majority of evaluations, their descriptions are, or eventually will be, judged against a continuum of expectations.

Observations usually are guided by a structured protocol. The protocol can take a variety of forms, ranging from the request for a narrative describing events seen to a checklist or a rating scale of specific behaviors/activities that address the evaluation question of interest. The use of a protocol helps assure that all observers are gathering the pertinent information and, with appropriate training, applying the same criteria in the evaluation. For example, if, as described earlier, an observational approach is selected to gather data on the faculty training sessions, the instrument developed would explicitly guide the observer to examine the kinds of activities in which participants were interacting, the role(s) of the trainers and the participants, the types of materials provided and used, the opportunity for hands-on interaction, etc. (See Appendix A to this chapter for an example of

Exhibit 4.

Types of information for which observations are a good source

The **setting** - The physical environment within which the project takes place.

The **human, social environment** - The ways in which all actors (staff, participants, others) interact and behave toward each other.

Project implementation activities - What goes on in the life of the project? What do various actors (staff, participants, others) actually do? How are resources allocated?

The **native language of the program** - Different organizations and agencies have their own language or jargon to describe the problems they deal with in their work; capturing the precise language of all participants is an important way to record how staff and participants understand their experiences.

Nonverbal communication - Nonverbal cues about what is happening in the project: on the way all participants dress, express opinions, physically space themselves during discussions, and arrange themselves in their physical setting.

Notable nonoccurrences - Determining what is not occurring although the expectation is that it should be occurring as planned by the project team, or noting the absence of some particular activity/factor that is noteworthy and would serve as added information.

observational protocol that could be applied to the hypothetical project.)

The protocol goes beyond a recording of events, i.e., use of identified materials, and provides an overall context for the data. The protocol should prompt the observer to

- **Describe the setting** of program delivery, i.e., where the observation took place and what the physical setting was like;
- **Identify the people** who participated in those activities, i.e., characteristics of those who were present;
- **Describe the content of the intervention**, i.e., actual activities and messages that were delivered;
- **Document the interactions** between implementation staff and project participants;
- **Describe and assess** the quality of the delivery of the intervention; and
- **Be alert to** unanticipated events that might require refocusing one or more evaluation questions.

Field notes are frequently used to provide more indepth background or to help the observer remember salient events if a form is not completed at the time of observation. Field notes contain the description of what has been observed. The descriptions must be factual, accurate, and thorough without being judgmental and cluttered by trivia. The date and time of the observation should be recorded, and everything that the observer believes to be worth noting should be included. No information should be trusted to future recall.

The use of technological tools, such as battery-operated tape recorder or dictaphone, laptop computer, camera, and video camera, can make the collection of field notes more efficient and the notes themselves more comprehensive. Informed consent must be obtained from participants before any observational data are gathered.

The Role of the Observer

There are various methods for gathering observational data, depending on the nature of a given project. The most fundamental distinction between various observational strategies concerns the extent to which the observer will be a *participant* in the setting being studied. The extent of participation is a continuum that varies from

The most fundamental distinction between various observational strategies concerns the extent to which the observer will be a participant in the setting being studied.

complete involvement in the setting as a full participant to complete separation from the setting as an outside observer or spectator. The participant observer is fully engaged in experiencing the project setting while at the same time trying to understand that setting through personal experience, observations, and interactions and discussions with other participants. The outside observer stands apart from the setting, attempts to be nonintrusive, and assumes the role of a “fly-on-the-wall.” The extent to which full participation is possible and desirable will depend on the nature of the project and its participants, the political and social context, the nature of the evaluation questions being asked, and the resources available. “The ideal is to negotiate and adopt that degree of participation that will yield the most meaningful data about the program given the characteristics of the participants, the nature of staff-participant interactions, and the sociopolitical context of the program” (Patton, 1990).


In some cases it may be beneficial to have two people observing at the same time. This can increase the quality of the data by providing a larger volume of data and by decreasing the influence of observer bias. However, in addition to the added cost, the presence of two observers may create an environment threatening to those being observed and cause them to change their behavior. Studies using observation typically employ intensive training experiences to make sure that the observer or observers know what to look for and can, to the extent possible, operate in an unbiased manner. In long or complicated studies, it is useful to check on an observer’s performance periodically to make sure that accuracy is being maintained. The issue of training is a critical one and may make the difference between a defensible study and what can be challenged as “one person’s perspective.”

A special issue with regard to observations relates to the amount of observation needed. While in participant observation this may be a moot point (except with regard to data recording), when an outside observer is used, the question of “how much” becomes very important. While most people agree that one observation (a single hour of a training session or one class period of instruction) is not enough, there is no hard and fast rule regarding how many samples need to be drawn. General tips to consider are to avoid atypical situations, carry out observations more than one time, and (where possible and relevant) spread the observations out over time.

Participant observation is often difficult to incorporate in evaluations; therefore, the use of outside observers is far more common. In the hypothetical project, observations might be scheduled for all training sessions and for a sample of classrooms, including some where faculty members who participated in training were teaching and some staffed by teachers who had not participated in the training.

Issues of privacy and access. Observational techniques are perhaps the most privacy-threatening data collection technique for staff and, to a lesser extent, participants. Staff fear that the data may be included in their performance evaluations and may have effects on their careers. Participants may also feel uncomfortable assuming that they are being judged. Evaluators need to assure everyone that evaluations of performance are not the purpose of the effort, and that no such reports will result from the observations. Additionally, because most educational settings are subject to a constant flow of observers from various organizations, there is often great reluctance to grant access to additional observers. Much effort may be needed to assure project staff and participants that they will not be adversely affected by the evaluators' work and to negotiate observer access to specific sites.

Interviews




Interviews allow the evaluation team to capture the perspectives of project participants, staff, and others associated with the project.

Interviews provide very different data from observations: they allow the evaluation team to capture the perspectives of project participants, staff, and others associated with the project. In the hypothetical example, interviews with project staff can provide information on the early stages of the implementation and problems encountered. The use of interviews as a data collection method begins with the assumption that the participants' perspectives are meaningful, knowable, and able to be made explicit, and that their perspectives affect the success of the project. An interview, rather than a paper and pencil survey, is selected when interpersonal contact is important and when opportunities for followup of interesting comments are desired.

Two types of interviews are used in evaluation research: structured interviews, in which a carefully worded questionnaire is administered; and indepth interviews, in which the interviewer does not follow a rigid form. In the former, the emphasis is on obtaining answers to carefully phrased questions. Interviewers are trained to deviate only minimally from the question wording to ensure uniformity of interview administration. In the latter, however, the interviewers seek to encourage free and open responses, and there may be a tradeoff between comprehensive coverage of topics and indepth exploration of a more limited set of questions. Indeath interviews also encourage capturing of respondents' perceptions in their own words, a very desirable strategy in qualitative data collection. This allows the evaluator to present the meaningfulness of the experience from the respondent's perspective. Indeath interviews are conducted with individuals or with a small group of individuals.⁴

⁴ A special case of the group interview is called a focus group. Although we discuss focus groups separately, several of the exhibits in this section will refer to both forms of data collection because of their similarities.



Indepth interviews are characterized by extensive probing and open-ended questions.

Indepth interviews. An indepth interview is a dialogue between a skilled interviewer and an interviewee. Its goal is to elicit rich, detailed material that can be used in analysis (Lofland and Lofland, 1995). Such interviews are best conducted face to face, although in some situations telephone interviewing can be successful.

Indepth interviews are characterized by extensive probing and open-ended questions. Typically, the project evaluator prepares an interview guide that includes a list of questions or issues that are to be explored and suggested probes for following up on key topics. The guide helps the interviewer pace the interview and make interviewing more systematic and comprehensive. Lofland and Lofland (1995) provide guidelines for preparing interview guides, doing the interview with the guide, and writing up the interview. Appendix B to this chapter contains an example of the types of interview questions that could be asked during the hypothetical study.

The dynamics of interviewing are similar to a guided conversation. The interviewer becomes an attentive listener who shapes the process into a familiar and comfortable form of social engagement—a conversation—and the quality of the information obtained is largely dependent on the interviewer’s skills and personality (Patton, 1990). In contrast to a good conversation, however, an indepth interview is not intended to be a two-way form of communication and sharing. The key to being a good interviewer is being a good listener and questioner. Tempting as it may be, it is not the role of the interviewer to put forth his or her opinions, perceptions, or feelings. Interviewers should be trained individuals who are sensitive, empathetic, and able to establish a nonthreatening environment in which participants feel comfortable. They should be selected during a process that weighs personal characteristics that will make them acceptable to the individuals being interviewed; clearly, age, sex, profession, race/ethnicity, and appearance may be key characteristics. Thorough training, including familiarization with the project and its goals, is important. Poor interviewing skills, poor phrasing of questions, or inadequate knowledge of the subject’s culture or frame of reference may result in a collection that obtains little useful data.

When to use indepth interviews. Indepth interviews can be used at any stage of the evaluation process. They are especially useful in answering questions such as those suggested by Patton (1990):

- What does the program look and feel like to the participants? To other stakeholders?
- What are the experiences of program participants?

<p>Exhibit 5. Advantages and disadvantages of indepth interviews</p>
<p style="text-align: center;">Advantages</p> <p>Usually yield richest data, details, new insights</p> <p>Permit face-to-face contact with respondents</p> <p>Provide opportunity to explore topics in depth</p> <p>Afford ability to experience the affective as well as cognitive aspects of responses</p> <p>Allow interviewer to explain or help clarify questions, increasing the likelihood of useful responses</p> <p>Allow interviewer to be flexible in administering interview to particular individuals or circumstances</p>
<p style="text-align: center;">Disadvantages</p> <p>Expensive and time-consuming</p> <p>Need well-qualified, highly trained interviewers</p> <p>Interviewee may distort information through recall error, selective perceptions, desire to please interviewer</p> <p>Flexibility can result in inconsistencies across interviews</p> <p>Volume of information too large; may be difficult to transcribe and reduce data</p>

- What do stakeholders know about the project?
- What thoughts do stakeholders knowledgeable about the program have concerning program operations, processes, and outcomes?
- What are participants' and stakeholders' expectations?
- What features of the project are most salient to the participants?
- What changes do participants perceive in themselves as a result of their involvement in the project?

Specific circumstances for which indepth interviews are particularly appropriate include

- complex subject matter;
- detailed information sought;
- busy, high-status respondents; and
- highly sensitive subject matter.

In the hypothetical project, indepth interviews of the project director, staff, department chairs, branch campus deans, and nonparticipant faculty would be useful. These interviews can address both formative and summative questions and be used in conjunction with other data collection methods. The advantages and disadvantages of indepth interviews are outlined in Exhibit 5.

When indepth interviews are being considered as a data collection technique, it is important to keep several potential pitfalls or problems in mind.

- **There may be substantial variation in the interview setting.** Interviews generally take place in a wide range of settings. This limits the interviewer's control over the environment. The interviewer may have to contend with disruptions and other problems that may inhibit the acquisition of information and limit the comparability of interviews.
- **There may be a large gap between the respondent's knowledge and that of the interviewer.** Interviews are often conducted with knowledgeable respondents yet administered by less knowledgeable interviewers or by interviewers not completely familiar with the pertinent social, political, or cultural context. Therefore, some of the responses may not be correctly understood or reported. The solution may be not only to employ highly trained and knowledgeable staff, but also to use interviewers with special skills for specific types of

respondents (for example, same status interviewers for high-level administrators or community leaders). It may also be most expedient for the project director or senior evaluation staff to conduct such interviews, if this can be done without introducing or appearing to introduce bias.

Exhibit 6.
Considerations in conducting indepth interviews and focus groups

Factors to consider in determining the setting for interviews (both individual and group) include the following:

- Select a setting that provides privacy for participants.
- Select a location where there are no distractions and it is easy to hear respondents speak.
- Select a comfortable location.
- Select a nonthreatening environment.
- Select a location that is easily accessible for respondents.
- Select a facility equipped for audio or video recording.
- Stop telephone or visitor interruptions to respondents interviewed in their office or homes.
- Provide seating arrangements that encourage involvement and interaction.

Exhibit 6 outlines other considerations in conducting interviews. These considerations are also important in conducting focus groups, the next technique that we will consider.

Recording interview data. Interview data can be recorded on tape (with the permission of the participants) and/or summarized in notes. As with observations, detailed recording is a necessary component of interviews since it forms the basis for analyzing the data. All methods, but especially the second and third, require carefully crafted interview guides with ample space available for recording the interviewee's responses. Three procedures for recording the data are presented below.

In the first approach, the interviewer (or in some cases the transcriber) listens to the tapes and writes a verbatim account of everything that was said. Transcription of the raw data includes word-for-word quotations of the participant's responses as well as the interviewer's descriptions of participant's characteristics, enthusiasm, body language, and overall mood during the interview. Notes from the interview can be used to identify speakers or to recall comments that are garbled or unclear on the tape. This approach is recommended when the necessary financial and human resources are available, when the transcriptions can be produced in a reasonable amount of time, when the focus of the interview is to make detailed comparisons, or when respondents' own words and phrasing are needed. The major advantages of this transcription method are its completeness and the opportunity it affords for the interviewer to remain attentive and focused during the interview. The major disadvantages are the amount of time and resources needed to produce complete transcriptions and the inhibitory impact tape recording has on some respondents. If this technique is selected, it is essential that the participants have been informed that their answers are being recorded, that they are assured confidentiality, and that their permission has been obtained.

A second possible procedure for recording interviews draws less on the word-by-word record and more on the notes taken by the interviewer or assigned notetaker. This method is called "note expansion." As soon as possible after the interview, the interviewer listens to the tape to clarify certain issues and to confirm that all the main points have been included in the notes. This approach is recommended when resources are scarce, when the results must be produced in a short period of time, and when the purpose of the interview is to get rapid feedback from members of the target population. The note expansion approach saves

time and retains all the essential points of the discussion. In addition to the drawbacks pointed out above, a disadvantage is that the interviewer may be more selective or biased in what he or she writes.

In the third approach, the interviewer uses no tape recording, but instead takes detailed notes during the interview and draws on memory to expand and clarify the notes immediately after the interview. This approach is useful if time is short, the results are needed quickly, and the evaluation questions are simple. Where more complex questions are involved, effective note-taking can be achieved, but only after much practice. Further, the interviewer must frequently talk and write at the same time, a skill that is hard for some to achieve.


Focus Groups

Focus groups combine elements of both interviewing and participant observation. The focus group session is, indeed, an interview (Patton, 1990) not a discussion group, problem-solving session, or decision-making group. At the same time, focus groups capitalize on group dynamics. The hallmark of focus groups is the explicit use of the group interaction to generate data and insights that would be unlikely to emerge without the interaction found in a group. The technique inherently allows observation of group dynamics, discussion, and firsthand insights into the respondents' behaviors, attitudes, language, etc.

Focus groups are a gathering of 8 to 12 people who share some characteristics relevant to the evaluation. Originally used as a market research tool to investigate the appeal of various products, the focus group technique has been adopted by other fields, such as education, as a tool for data gathering on a given topic. Focus groups conducted by experts take place in a focus group facility that includes recording apparatus (audio and/or visual) and an attached room with a one-way mirror for observation. There is an official recorder who may or may not be in the room. Participants are paid for attendance and provided with refreshments. As the focus group technique has been adopted by fields outside of marketing, some of these features, such as payment or refreshment, have been eliminated.

When to use focus groups. When conducting evaluations, focus groups are useful in answering the same type of questions as indepth interviews, except in a social context. Specific applications of the focus group method in evaluations include

- identifying and defining problems in project implementation;



*Focus groups and
indepth interviews
should not be used
interchangeably.*

- identifying project strengths, weaknesses, and recommendations;
- assisting with interpretation of quantitative findings;⁵
- obtaining perceptions of project outcomes and impacts; and
- generating new ideas.

In the hypothetical project, focus groups could be conducted with project participants to collect perceptions of project implementation and operation (e.g., Were the workshops staffed appropriately? Were the presentations suitable for all participants?), as well as progress toward objectives during the formative phase of evaluation (Did participants exchange information by e-mail and other means?). Focus groups could also be used to collect data on project outcomes and impact during the summative phase of evaluation (e.g., Were changes made in the curriculum? Did students taught by participants appear to become more interested in class work? What barriers did the participants face in applying what they had been taught?).

Although focus groups and indepth interviews share many characteristics, they should not be used interchangeably. Factors to consider when choosing between focus groups and indepth interviews are included in Exhibit 7.

Developing a Focus Group

An important aspect of conducting focus groups is the topic guide. (See Appendix C to this chapter for a sample guide applied to the hypothetical project.) The topic guide, a list of topics or question areas, serves as a summary statement of the issues and objectives to be covered by the focus group. The topic guide also serves as a road map and as a memory aid for the focus group leader, called a “moderator.” The topic guide also provides the initial outline for the report of finding.

Focus group participants are typically asked to reflect on the questions asked by the moderator. Participants are permitted to hear each other’s responses and to make additional comments beyond their own original responses as they hear what other people have to say. It is not necessary for the group to reach any kind of consensus, nor it is necessary for people to disagree. The moderator must keep the

⁵ Survey developers also frequently use focus groups to pretest topics or ideas that later will be used for quantitative data collection. In such cases, the data obtained are considered part of instrument development rather than findings. Qualitative evaluators feel that this is too limited an application and that the technique has broader utility.

Exhibit 7.		
Which to use: Focus groups or indepth interviews?		
<u>Factors to consider</u>	<u>Use focus groups when...</u>	<u>Use indepth interview when...</u>
Group interaction	interaction of respondents may stimulate a richer response or new and valuable thought.	group interaction is likely to be limited or nonproductive.
Group/peer pressure	group/peer pressure will be valuable in challenging the thinking of respondents and illuminating conflicting opinions.	group/peer pressure would inhibit responses and cloud the meaning of results.
Sensitivity of subject matter	subject matter is not so sensitive that respondents will temper responses or withhold information.	subject matter is so sensitive that respondents would be unwilling to talk openly in a group.
Depth of individual responses	the topic is such that most respondents can say all that is relevant or all that they know in less than 10 minutes.	the topic is such that a greater depth of response per individual is desirable, as with complex subject matter and very knowledgeable respondents.
Data collector fatigue	it is desirable to have one individual conduct the data collection; a few groups will not create fatigue or boredom for one person.	it is possible to use numerous individuals on the project; one interviewer would become fatigued or bored conducting all interviews.
Extent of issues to be covered	the volume of issues to cover is not extensive.	a greater volume of issues must be covered.
Continuity of information	a single subject area is being examined in depth and strings of behaviors are less relevant.	it is necessary to understand how attitudes and behaviors link together on an individual basis.
Experimentation with interview guide	enough is known to establish a meaningful topic guide.	it may be necessary to develop the interview guide by altering it after each of the initial interviews.
Observation by stakeholders	it is desirable for stakeholders to hear what participants have to say.	stakeholders do not need to hear firsthand the opinions of participants.
Logistics geographically	an acceptable number of target respondents can be assembled in one location.	respondents are dispersed or not easily assembled for other reasons.
Cost and training	quick turnaround is critical, and funds are limited.	quick turnaround is not critical, and budget will permit higher cost.
Availability of qualified staff	focus group facilitators need to be able to control and manage groups	interviewers need to be supportive and skilled listeners.

discussion flowing and make sure that one or two persons do not dominate the discussion. As a rule, the focus group session should not last longer than 1 1/2 to 2 hours. When very specific information is required, the session may be as short as 40 minutes. The objective is to get high-quality data in a social context where people can consider their own views in the context of the views of others, and where new ideas and perspectives can be introduced.

The participants are usually a relatively homogeneous group of people. Answering the question, “Which respondent variables represent relevant similarities among the target population?” requires some thoughtful consideration when planning the evaluation. Respondents’ social class, level of expertise, age, cultural background, and sex should always be considered. There is a sharp division among focus group moderators regarding the effectiveness of mixing sexes within a group, although most moderators agree that it is acceptable to mix the sexes when the discussion topic is not related to or affected by sex stereotypes.

Determining how many groups are needed requires balancing cost and information needs. A focus group can be fairly expensive, costing \$10,000 to \$20,000 depending on the type of physical facilities needed, the effort it takes to recruit participants, and the complexity of the reports required. A good rule of thumb is to conduct at least two groups for every variable considered to be relevant to the outcome (sex, age, educational level, etc.). However, even when several groups are sampled, conclusions typically are limited to the specific individuals participating in the focus group. Unless the study population is extremely small, it is not possible to generalize from focus group data.

Recording focus group data. The procedures for recording a focus group session are basically the same as those used for in-depth interviews. However, the focus group approach lends itself to more creative and efficient procedures. If the evaluation team does use a focus group room with a one-way mirror, a colleague can take notes and record observations. An advantage of this approach is that the extra individual is not in the view of participants and, therefore, not interfering with the group process. If a one-way mirror is not a possibility, the moderator may have a colleague present in the room to take notes and to record observations. A major advantage of these approaches is that the recorder focuses on observing and taking notes, while the moderator concentrates on asking questions, facilitating the group interaction, following up on ideas, and making smooth transitions from issue to issue. Furthermore, like observations, focus groups can be videotaped. These approaches allow for confirmation of what was seen and heard. Whatever the approach to gathering detailed data, informed consent is necessary and confidentiality should be assured.

Having highlighted the similarities between interviews and focus groups, it is important to also point out one critical difference. In focus groups, group dynamics are especially important. The notes, and resultant report, should include comments on group interaction and dynamics as they inform the questions under study.

Other Qualitative Methods

The last section of this chapter outlines less common but, nonetheless, potentially useful qualitative methods for project evaluation. These methods include document studies, key informants, alternative (authentic) assessment, and case studies.

Document Studies

Existing records often provide insights into a setting and/or group of people that cannot be observed or noted in another way. This information can be found in document form. Lincoln and Guba (1985) defined a document as “any written or recorded material” not prepared for the purposes of the evaluation or at the request of the inquirer. Documents can be divided into two major categories: public records, and personal documents (Guba and Lincoln, 1981).

Public records are materials created and kept for the purpose of “attesting to an event or providing an accounting” (Lincoln and Guba, 1985). Public records can be collected from outside (*external*) or within (*internal*) the setting in which the evaluation is taking place. Examples of *external* records are census and vital statistics reports, county office records, newspaper archives, and local business records that can assist an evaluator in gathering information about the larger community and relevant trends. Such materials can be helpful in better understanding the project participants and making comparisons between groups/communities.

For the evaluation of educational innovations, *internal* records include documents such as student transcripts and records, historical accounts, institutional mission statements, annual reports, budgets, grade and standardized test reports, minutes of meetings, internal memoranda, policy manuals, institutional histories, college/university catalogs, faculty and student handbooks, official correspondence, demographic material, mass media reports and presentations, and descriptions of program development and evaluation. They are particularly useful in describing institutional characteristics, such as backgrounds and academic performance of students, and in identifying institutional strengths and weaknesses. They can help the evaluator understand the

institution's resources, values, processes, priorities, and concerns. Furthermore, they provide a record or history not subject to recall bias.

Personal documents are first-person accounts of events and experiences. These "documents of life" include diaries, portfolios, photographs, artwork, schedules, scrapbooks, poetry, letters to the paper, etc. Personal documents can help the evaluator understand how the participant sees the world and what she or he wants to communicate to an audience. And unlike other sources of qualitative data, collecting data from documents is relatively invisible to, and requires minimal cooperation from, persons within the setting being studied (Fetterman, 1989).

The usefulness of existing sources varies depending on whether they are accessible and accurate. In the hypothetical project, documents can provide the evaluator with useful information about the culture of the institution and participants involved in the project, which in turn can assist in the development of evaluation questions. Information from documents also can be used to generate interview questions or to identify events to be observed. Furthermore, existing records can be useful for making comparisons (e.g., comparing project participants to project applicants, project proposal to implementation records, or documentation of institutional policies and program descriptions prior to and following implementation of project interventions and activities).

The advantages and disadvantages of document studies are outlined in Exhibit 8.

Key Informant

A key informant is a person (or group of persons) who has unique skills or professional background related to the issue/intervention being evaluated, is knowledgeable about the project participants, or has access to other information of interest to the evaluator. A key informant can also be someone who has a way of communicating that represents or captures the essence of what the participants say and do. Key informants can help the evaluation team better understand the issue being evaluated, as well as the project participants, their backgrounds, behaviors, and attitudes, and any language or ethnic considerations. They can offer expertise beyond the evaluation team. They are also very useful for assisting with the evaluation of curricula and other educational materials. Key informants can be surveyed or interviewed individually or through focus groups.

In the hypothetical project, key informants (i.e., expert faculty on main campus, deans, and department chairs) can assist with (1) developing

Exhibit 8. Advantages and disadvantages of document studies
<p style="text-align: center;">Advantages</p> <p>Available locally</p> <p>Inexpensive</p> <p>Grounded in setting and language in which they occur</p> <p>Useful for determining value, interest, positions, political climate, public attitudes, historical trends or sequences</p> <p>Provide opportunity for study of trends over time</p> <p>Unobtrusive</p>
<p style="text-align: center;">Disadvantages</p> <p>May be incomplete</p> <p>May be inaccurate; questionable authenticity</p> <p>Locating suitable documents may pose challenges</p> <p>Analysis may be time consuming</p> <p>Access may be difficult</p>

evaluation questions, and (2) answering formative and summative evaluation questions.

The use of advisory committees is another way of gathering information from key informants. Advisory groups are called together for a variety of purposes:

- To represent the ideas and attitudes of a community, group, or organization;
- To promote legitimacy for project;
- To advise and recommend; or
- To carry out a specific task.

Members of such a group may be specifically selected or invited to participate because of their unique skills or professional background; they may volunteer; they may be nominated or elected; or they may come together through a combination of these processes.

The advantages and disadvantages of using key informants are outlined in Exhibit 9.

<p>Exhibit 9. Advantages and disadvantages of using key informants</p>
<p style="text-align: center;">Advantages</p> <p>Information concerning causes, reasons, and/or best approaches from an “insider” point of view</p> <p>Advice/feedback increases credibility of study</p> <p>Pipeline to pivotal groups</p> <p>May have side benefit to solidify relationships between evaluators, clients, participants, and other stakeholders</p>
<p style="text-align: center;">Disadvantages</p> <p>Time required to select and get commitment may be substantial</p> <p>Relationship between evaluator and informants may influence type of data obtained</p> <p>Informants may interject own biases and impressions</p> <p>May result in disagreements among individuals leading to frustration/ conflicts</p>

Performance Assessment

The performance assessment movement is impacting education from preschools to professional schools. At the heart of this upheaval is the belief that for all of their virtues—particularly efficiency and economy—traditional objective, norm-referenced tests may fail to tell us what we most want to know about student achievement. In addition, these same tests exert a powerful and, in the eyes of many educators, detrimental influence on curriculum and instruction. Critics of traditional testing procedures are exploring alternatives to multiple-choice, norm-referenced tests. It is hoped that these alternative means of assessment, ranging from observations to exhibitions, will provide a more authentic picture of achievement.

Critics raise three main points against objective, norm-referenced tests.

- Tests themselves are flawed.
- Tests are a poor measure of anything except a student’s test-taking ability.
- Tests corrupt the very process they are supposed to improve (i.e., their structure puts too much emphasis on learning isolated facts).

The search for alternatives to traditional tests has generated a number of new approaches to assessment under such names as alternative assessment, performance assessment, holistic assessment, and authentic assessment. While each label suggests slightly different emphases, they all imply a movement toward assessment that supports exemplary teaching. Performance assessment appears to be the most popular term because it emphasizes the development of assessment tools that involve students in tasks that are worthwhile, significant, and meaningful. Such tasks involve higher order thinking skills and the coordination of a broad range of knowledge.

Performance assessment may involve “qualitative” activities such as oral interviews, group problem-solving tasks, portfolios, or personal documents/creations (poetry, artwork, stories). A performance assessment approach that could be used in the hypothetical project is work sample methodology (Schalock, Schalock, and Girad, in press). Briefly, work sample methodology challenges teachers to create unit plans and assessment techniques for students at several points during a training experience. The quality of this product is assessed (at least before and after training) in light of the goal of the professional development program. The actual performance of students on the assessment measures provides additional information on impact.

Case Studies

Classical case studies depend on ethnographic and participant observer methods. They are largely descriptive examinations, usually of a small number of sites (small towns, hospitals, schools) where the principal investigator is immersed in the life of the community or institution and combs available documents, holds formal and informal conversations with informants, observes ongoing activities, and develops an analysis of both individual and “cross-case” findings.

In the hypothetical study, for example, case studies of the experiences of participants from different campuses could be carried out. These might involve indepth interviews with the facility participants, observations of their classes over time, surveys of students, interviews with peers and department chairs, and analyses of student work samples at several points in the program. Selection of participants might be made based on factors such as their experience and training, type of students taught, or differences in institutional climate/supports.

Case studies can provide very engaging, rich explorations of a project or application as it develops in a real-world setting. Project evaluators must be aware, however, that doing even relatively modest, illustrative case studies is a complex task that cannot be accomplished through

occasional, brief site visits. Demands with regard to design, data collection, and reporting can be substantial.

For those wanting to become thoroughly familiar with this topic, a number of relevant texts are referenced here.

References

- Fetterman, D.M. (1989). *Ethnography: Step by Step*. Applied Social Research Methods Series, Vol. 17. Newbury Park, CA: Sage.
- Guba, E.G., and Lincoln, Y.S. (1981). *Effective Evaluation*. San Francisco: Jossey-Bass.
- Lincoln, Y.S., and Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lofland, J., and Lofland, L.H. (1995). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd Ed. Belmont, CA: Wadsworth.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*, 2nd Ed. Newbury Park, CA: Sage.
- Schalock, H.D., Schalock, M.D., and Girad, G.R. (In press). Teacher work sample methodology, as used at Western Oregon State College. In J. Millman, Ed., *Assuring Accountability? Using Gains in Student Learning to Evaluate Teachers and Schools*. Newbury Park, CA: Corwin.

Other Recommended Reading

- Debus, M. (1995). *Methodological Review: A Handbook for Excellence in Focus Group Research*. Washington, DC: Academy for Educational Development.
- Denzin, N.K., and Lincoln, Y.S. (Eds.). (1994). *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.
- Erlandson, D.A., Harris, E.L., Skipper, B.L., and Allen, D. (1993). *Doing Naturalist Inquiry: A Guide to Methods*. Newbury Park, CA: Sage.
- Greenbaum, T.L. (1993). *The Handbook of Focus Group Research*. New York: Lexington Books.

- Hart, D. (1994). *Authentic Assessment: A Handbook for Educators*. Menlo Park, CA: Addison-Wesley.
- Herman, J.L., and Winters, L. (1992). *Tracking Your School's Success: A Guide to Sensible Evaluation*. Newbury Park, CA: Corwin Press.
- Hymes, D.L., Chafin, A.E., and Gondor, R. (1991). *The Changing Face of Testing and Assessment: Problems and Solutions*. Arlington, VA: American Association of School Administrators.
- Krueger, R.A. (1988). *Focus Groups: A Practical Guide for Applied Research*. Newbury Park, CA: Sage.
- LeCompte, M.D., Millroy, W.L., and Preissle, J. (Eds.). (1992). *The Handbook of Qualitative Research in Education*. San Diego, CA: Academic Press.
- Merton, R.K., Fiske, M., and Kendall, P.L. (1990). *The Focused Interview: A Manual of Problems and Procedures*, 2nd Ed. New York: The Free Press.
- Miles, M.B., and Huberman, A.M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage.
- Morgan, D.L. (Ed.). (1993). *Successful Focus Groups: Advancing the State of the Art*. Newbury Park, CA: Sage.
- Morse, J.M. (Ed.). (1994). *Critical Issues in Qualitative Research Methods*. Thousand Oaks, CA: Sage.
- Perrone, V. (Ed.). (1991). *Expanding Student Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Reich, R.B. (1991). *The Work of Nations*. New York: Alfred A. Knopf.
- Schatzman, L., and Strauss, A.L. (1973). *Field Research*. Englewood Cliffs, NJ: Prentice-Hall.
- Seidman, I.E. (1991). *Interviewing as Qualitative Research: A Guide for Researchers in Education and Social Sciences*. New York: Teachers College Press.
- Stewart, D.W., and Shamdasani, P.N. (1990). *Focus Groups: Theory and Practice*. Newbury Park, CA: Sage.

United States General Accounting Office (GAO). (1990). *Case Study Evaluations*, Paper 10.1.9. Washington, DC: GAO.

Weiss, R.S. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York: Free Press.

Wiggins, G. (1989). A True Test: Toward More Authentic and Equitable Assessment. *Phi Delta Kappan*, May, 703-704.

Wiggins, G. (1989). Teaching to the (Authentic) Test. *Educational Leadership*, 46, 45.

Yin, R.K. (1989). *Case Study Research: Design and Method*. Newbury Park, CA: Sage.

Appendix A

Sample Observation Instrumentⁱ

ⁱ Developed from Weiss, Iris, 1997 Local Systemic Change Observation Protocol.

Faculty Development Observation Protocol

BACKGROUND INFORMATION

Observer _____ Date of Observation _____

Duration of Observation:

- 1 hour half day
 2 hours whole day
 Other, please specify _____

Total Number of Attendees _____

Name of Presenter(s) _____

SECTION ONE: CONTEXT BACKGROUND AND ACTIVITIES

This section provides a brief overview of the session being observed.

I. Session Context

In a few sentences, describe the session you observed. Include: (a) whether the observation covered a partial or complete session, (be) whether there were multiple break-out sessions, and (c) where this session fits in the project's sequence of faculty development for those in attendance.

II. Session Focus

Indicate the *major intended purpose(s)* of this session based on the information provided by the project staff.

III. Faculty Development Activities (*Check all activities observed and describe, as relevant*)

A. Indicate the major instructional resource(s) used in this faculty development session.

- Print materials
- Hands-on materials
- Outdoor resources
- Technology/audio-visual resources
- Other instructional resources (Please specify.)

B. Indicate the major way(s) in which participant activities were structured.

- As a whole group
- As small groups
- As pairs
- As individuals

C. Indicate the major activities of presenters and participants in this session. (Check circle to indicate applicability.)

- Formal presentations by presenter/facilitator: (*describe focus*)

- Formal presentations by participants: (*describe focus*)

- Hands-on/investigative/research/field activities: (*describe*)

- Problem-solving activities: (*describe*)

Proof and evidence: *(describe)*

Reading/reflection/written communication: *(describe)*

Explored technology use: *(describe focus)*

Explored assessment strategies: *(describe focus)*

Assessed participants' knowledge and/or skills: *(describe approach)*

Other activities: (Please specify)

D. Comments

Please provide any additional information you consider necessary to capture the activities or context of this faculty development session. Include comments on any feature of the session that is so salient that you need to get it “on the table” right away to help explain your ratings.

SECTION TWO: RATINGS

In Section One of this form, you documented what occurred in the session. In this section, you are asked to use that information, as well as any other pertinent observations, to rate each of a number of key indicators from 1 (not at all) to 5 (to a great extent) in five difference categories by circling the appropriate response. Note that any one session is not likely to provide evidence for every single indicator; use 6, “Don’t know” when there is not enough evidence for you to make a judgment. Use 7, “N/A” (Not Applicable) when you consider the indicator inappropriate given the purpose and context of the session. Similarly, there may be entire rating categories that are not applicable to a particular session.

Note that you may list any additional indicators you consider important in capturing the essence of this session and rate these as well.

Use your “Ratings of Key Indicators” (Part A) to inform your “Synthesis Ratings” (Part B) and indicate in “Supporting Evidence for Synthesis Ratings” (Part C) what factors were most influential in determining your synthesis ratings. Section Two concludes with ratings of the likely impact of faculty development and a capsule description of the session.

I. Design

A. Ratings of Key Indicators

	Not at all					To a great extent	Don't know	N/A
1. The strategies in this session were appropriate for accomplishing the purposes of the faculty development	1	2	3	4	5		6	7
2. The session effectively built on participants' knowledge of content, teaching, learning, and/or the reform/change process	1	2	3	4	5		6	7
3. The instructional strategies and activities used in this section reflected attention to participants':								
a. Experience, preparedness, and learning styles	1	2	3	4	5		6	7
b. Access to resources	1	2	3	4	5		6	7
4. The design of the session reflected careful planning and organization	1	2	3	4	5		6	7
5. The design of the session encouraged a collaborative approach to learning	1	2	3	4	5		6	7
6. The design of the session incorporated tasks, roles, and interactions consistent with a spirit of investigation	1	2	3	4	5		6	7
7. The design of the session provided opportunities for teachers to consider classroom applications of resources, strategies, and techniques	1	2	3	4	5		6	7
8. The design of the session appropriately balanced attention to multiple goals	1	2	3	4	5		6	7
9. Adequate time and structure were provided for reflection.....	1	2	3	4	5		6	7
10. Adequate time and structure were provided for participants to share experiences and insights	1	2	3	4	5		6	7
11. _____	1	2	3	4	5			

B. Synthesis Rating

1	2	3	4	5
Design of the session was not at all reflective of best practice for faculty development				Design of the session extremely reflective of best practice for faculty development

C. Supporting Evidence for Synthesis Rating

II. Implementation

A. Ratings of Key Indicators

	Not at all					To a great extent	Don't know	N/A
1. The session effectively incorporated instructional strategies that were appropriate for the purposes of the faculty development session and the needs of adult learners	1	2	3	4	5		6	7
2. The session effectively modeled questioning strategies that are likely to enhance the development of conceptual understanding (e.g., emphasis on higher-order questions, appropriate use of “wait time,” identifying perceptions and misconceptions)....	1	2	3	4	5		6	7
3. The pace of the session was appropriate for the purposes of the faculty development and the needs of adult learners	1	2	3	4	5		6	7
4. The session modeled effective assessment strategies ..	1	2	3	4	5		6	7
5. The presenter(s)' background, experience, and/or expertise enhanced the quality of the session	1	2	3	4	5		6	7
6. The presenter(s)' management style/strategies enhanced the quality of the session.....	1	2	3	4	5		6	7
7. _____	1	2	3	4	5			

B. Synthesis Rating

1	2	3	4	5
Implementation of the session not at all reflective of best practice for faculty development				Implementation of the session extremely reflective of best practice for faculty development

C. Supporting Evidence for Synthesis Rating

III. Disciplinary Content

Not applicable. (Disciplinary content not included in the session.)

A. Ratings of Key Indicators

	Not at all					To a great extent	Don't know	N/A
1. Disciplinary content was appropriate for the purposes of the faculty development session and the backgrounds of the participants.....	1	2	3	4	5		6	7
2. The content was sound and appropriately presented/ explored.....	1	2	3	4	5		6	7
3. Facilitator displayed an understanding of concepts (e.g., in his/her dialogue with participants).....	1	2	3	4	5		6	7
4. Content area was portrayed by a dynamic body of knowledge continually enriched by conjecture, investigation, analysis, and proof/justification	1	2	3	4	5		6	7
5. Depth and breadth of attention to disciplinary content was appropriate for the purposes of the session and the needs of adult learners	1	2	3	4	5		6	7
6. Appropriate connections were made to other areas of science/mathematics, to other disciplines, and/or to real world contexts	1	2	3	4	5		6	7
7. Degree of closure or resolution of conceptual understanding was appropriate for the purposes of the session and the needs of adult learners	1	2	3	4	5		6	7
8. _____	1	2	3	4	5			

B. Synthesis Rating

1	2	3	4	5
Disciplinary content of the session not at all reflective of best practice for faculty development				Disciplinary content of the session extremely reflective of best practice for faculty development

C. Supporting Evidence for Synthesis Rating

IV. Pedagogical Content

Not applicable. (Pedagogical content not included in the session.)

A. Ratings of Key Indicators

	Not at all					To a great extent	Don't know	N/A
1. Pedagogical content was appropriate for the purposes of the faculty development session and the backgrounds of the participants.....	1	2	3	4	5		6	7
2. Pedagogical content was sound and appropriately presented/explored	1	2	3	4	5		6	7
3. Presentor displayed an understanding of pedagogical concepts (e.g., in his/her dialogue with participants)...	1	2	3	4	5		6	7
4. The session included <i>explicit</i> attention to classroom implementation issues.....	1	2	3	4	5		6	7
5. Depth and breadth of attention to pedagogical content was appropriate for the purposes of the session and the needs of adult learners	1	2	3	4	5		6	7
6. Degree of closure or resolution of conceptual understanding was appropriate for the purposes of the session and the needs of adult learners	1	2	3	4	5		6	7
7. _____	1	2	3	4	5			

B. Synthesis Rating

1	2	3	4	5
Pedagogical content of the session not at all reflective of current standards for science/mathematics education				Pedagogical content of session extremely reflective of current standards for science/mathematics education

C. Supporting Evidence for Synthesis Rating

V. Culture/Equity

A. Ratings of Key Indicators

	Not at all					To a great extent	Don't know	N/A
1. Active participation of all was encouraged and valued	1	2	3	4	5		6	7
2. There was a climate of respect for participants' experiences, ideas, and contributions	1	2	3	4	5		6	7
3. Interactions reflected collaborative working relationships among participants	1	2	3	4	5		6	7
4. Interactions reflected collaborative working relationships between facilitator(s) and participants....	1	2	3	4	5		6	7
5. The presenter(s) language and behavior clearly demonstrated sensitivity to variations in participants': ¹								
a. Experience and/or preparedness.....	1	2	3	4	5		6	7
b. Access to resources	1	2	3	4	5		6	7
c. Gender, race/ethnicity, and/or culture	1	2	3	4	5		6	7
6. Opportunities were taken to recognize and challenge stereotypes and biases that became evident during the faculty development session	1	2	3	4	5		6	7
7. Participants were intellectually engaged with important ideas relevant to the focus of the session....	1	2	3	4	5		6	7
8. Faculty participants were encouraged to generate ideas, questions, conjectures, and propositions	1	2	3	4	5		6	7
9. Investigation and risk-taking were valued	1	2	3	4	5		6	7
10. Intellectual rigor, constructive criticism, and the challenging of ideas were valued.....	1	2	3	4	5		6	7
11. _____	1	2	3	4	5			

¹Use 1, "Not at all," when you have considerable evidence of insensitivity or inequitable behavior; 3, when there are no examples either way; and 5, "To a great extent," when there is considerable evidence of proactive efforts to achieve equity.

B. Synthesis Rating

1	2	3	4	5
Culture of the session interferes with engagement of participants as members of a faculty learning community				Culture of the session facilitates engagement of participants as members of a faculty learning community

C. Supporting Evidence for Synthesis Rating

VI. Overall Ratings of the Session

While the impact of a single faculty development session may well be limited in scope, it is important to judge whether it is helping move participants in the desired direction. For ratings in the section below, consider all available information (i.e., your previous ratings of design, implementation, content, and culture/equity; related interviews, and your knowledge of the overall faculty development program) as you assess likely impact of this session. Feel free to elaborate on ratings with comments in the space provided.

Likely Impact on Participants’ Capacity for Exemplary Instruction

Consider the likely impact of this session on the faculty participants’ capacity to teach exemplary science/mathematics instruction. Circle the response that best describes your overall assessment of the *likely effect* of this session in each of the following areas.

Not applicable. (The session did not focus on building capacity for classroom instruction.)

	Not at all					To a great extent	Don't know	N/A
	1	2	3	4	5			
1. Participants’ ability to identify and understand important ideas of science/mathematics	1	2	3	4	5		6	7
2. Participants’ understanding of science/mathematics as dynamic body of knowledge generated and enriched by investigation.....	1	2	3	4	5		6	7
3. Participants’ understanding of how students learn.....	1	2	3	4	5		6	7
4. Participants’ ability to plan/implement exemplary classroom instruction	1	2	3	4	5		6	7
5. Participants’ ability to implement exemplary classroom instructional materials	1	2	3	4	5		6	7
6. Participants’ self-confidence in instruction.....	1	2	3	4	5		6	7
7. Proactiveness of participants in addressing their faculty development needs.....	1	2	3	4	5		6	7
8. Professional networking among participants with regard to science/mathematics instruction.....	1	2	3	4	5		6	7

Comments (optional):

Name of Interviewer _____
Date _____
Name of Interviewee _____
Staff Position _____

Appendix B¹

Sample Indepth Interview Guide

Interview with Project Staff

“Good morning. I am _____ (introduce self).

This interview is being conducted to get your input about the implementation of the Undergraduate Faculty Enhancement workshops which you have been conducting/involved in. I am especially interested in any problems you have faced or are aware of and recommendations you have.”

“If it is okay with you, I will be tape recording our conversation. The purpose of this is so that I can get all the details but at the same time be able to carry on an attentive conversation with you. I assure you that all your comments will remain confidential. I will be compiling a report which will contain all staff comments without any reference to individuals. If you agree to this interview and the tape recording, please sign this consent form.”

“I’d like to start by having you briefly describe your responsibilities and involvement thus far with the Undergraduate Faculty Enhancement Project.” *(Note to interviewer: You may need to probe to gather the information you need).*

“I’m now going to ask you some questions that I would like you to answer to the best of your ability. If you do not know the answer, please say so.”

“Are you aware of any problems with the scheduling and location(s)?” *(Note to interviewer: If so, probe - “What have the problems been?”, “Do you know why these problems are occurring?”, “Do you have any suggestions on how to minimize these problems?”)*

“How were decisions made with respect to content and staffing of the first three workshops?” *(Note to interviewer: You may need to probe to gather the information about input from staff, participant reactions, availability of instructors, etc.)*

¹This guide was designed for interviews to be conducted after the project has been active for 3 months. For later interviews, the guide will need to be modified as appropriate.

“What is taking place in the workshops?” *(Note to interviewer: After giving individual time to respond, probe specific planned activities/strategies he/she may not have addressed - “What have the presentations been like?”, “Have there been demonstrations of model teaching? If so, please describe?”, “Has active participation been encouraged? Please describe for me how?”)*

“What do you think the strongest points of the workshops have been up to this point? Why do you say this?” *(Note to interviewer: You may need to probe why specific strong elements are mentioned - e.g., if interviewee replies “They work”, respond “How can you tell that they work?”)*

“What types of concerns have you had or heard regarding the availability of materials and equipment?” *(Note to interviewer: You may need to probe to gather the information you need)*

“What other problems are you aware of?” *(Note to interviewer: You may need to probe to gather the information you need)*

“What do you think about the project/workshops at this point?” *(Note to interviewer: You may need to probe to gather the information you need - e.g., “I’d like to know more about what your thinking is on that issue?”)*

“Is there any other information about the workshops or other aspects of the project that you think would be useful for me to know?” *(Note to interviewer: If so, you may need to probe to gather the information you need)*

Name of Moderator_____

Date_____

Attendees_____

Appendix C¹

Sample Focus Group Topic Guide

Workshop Participants

Evaluation Questions: Are the students taught by faculty participants exposed to new standards, materials, practices? Did this vary by faculty member? By students characteristics? Were there obstacles to changes?; What did the participants do to share knowledge with other faculty? Did other faculty adopt new concepts and practices?; Were changes made in curriculum? Examinations and other requirements? Expenditures for library and other resource materials? Did students taught by participants become more interested in class work? More active in class? Did they express interest in teaching math after graduation? Did they plan to use new concepts and techniques?

Introduction

Give an explanation

*Good afternoon. My name is _____ and this is my colleague _____.
Thank you for coming. A focus group is a relaxed discussion.....*

Present the purpose

We are here today to talk about your teaching experiences since you participated in the Undergraduate Faculty Enhancement workshops. The purpose is to get your perceptions of how the workshops have affected your teaching, your students, other faculty, and the curriculum. I am not here to share information, or to give you my opinions. Your perceptions are what matter. There are no right or wrong or desirable or undesirable answers. You can disagree with each other, and you can change your mind. I would like you to feel comfortable saying what you really think and how you really feel.

¹This guide was designed for year one participants one year after they had participated in training (month 22 of project).

Discuss procedure

_____ (colleague) will be taking notes and tape recording the discussion so that I do not miss anything you have to say. I explained these procedures to you when we set up this meeting. As you know everything is confidential. No one will know who said what. I want this to be a group discussion, so feel free to respond to me and to other members in the group without waiting to be called on. However, I would appreciate it if only one person did talk at a time. The discussion will last approximately one hour. There is a lot I want to discuss, so at times I may move us along a bit.

Participant introduction

Now, let's start by everyone sharing their name, what they teach, and how long they've been teaching.

Rapport building

I want each of you to think of an adjective that best described your teaching prior to the workshop experience and one that describes it following the experience. If you do not think your teaching has changed, you may select one adjective. We're going to go around the room so you can share your choices. Please briefly explain why you selected the adjective(s) you did.

Interview

What types of standards-based practice have you exposed students to since your participation in the workshops?

Probes: If there were standards not mentioned - Has anyone exposed students to _____? If not - Why not?

Would you have exposed students to these practices if you had not participated in the workshops?

Probes: Where would you have gotten this information? How would the information have been different?

How have you exposed students to these practices since completing the workshops?

Probes: Tell me more about that. How did that work?

Of the materials introduced to you through the workshops, which ones have you used?

Probes: Had you used these prior to the workshops? Tell me more about why you used these. Tell me more about how you used these. Of the materials not mentioned -- Has anyone used _____? Tell me why not.

Of these materials, which have you found the most useful?

Probes: Tell me more about why you have found this most useful. Of the materials not mentioned - Why haven't you found _____ useful? How could it be more useful?

Of the strategies introduced to you through the workshops, which ones have you applied to your teaching?

Probes: Tell me about how you have used this strategy. Of the strategies not mentioned - Has anyone tried _____? Tell me why not.

Of these strategies which ones have been most effective?

Probes: Tell me why you think they have been effective.

Which have you found to be least effective?

Probes: Tell me why you think they have not been effective. It's interesting, _____ found that strategy to be effective, what do you think may account for the difference?

What problems/obstacles have you faced in attempting to incorporate into your teaching the knowledge and skills you received through the workshops?

Probes: Tell me more about that.

How many of you have shared information from the workshops with other faculty?

Probes: Tell me about what you shared. Tell me about why you choose to share that aspect of the workshop? How did this happen (through presentations, faculty meetings, informal conversations, etc.) How have the other faculty responded? What concepts and practices have they adopted?

Has your experience in the workshops resulted in efforts, by you, your Chair, and/or Dean to make changes to the curriculum?

Probes: Tell me more about that. Tell me why you think this has/has not happened.

What about examinations and other requirements?

Similar probes to above

Since completing the workshops, describe for me any changes in your use of the library or resource center and purchase of educational materials.

Probes: How much more money would you say you've spent? Have you faced any problems with obtaining the resources you've requested?

Describe for me any changes you noticed in your students since your participation in the workshops.

Probes: Have their interest levels increased? How do you know that? Why do you think that is? How have your changes affected their active participation? What about their knowledge base? Skills? Anything else?

Describe for me the most beneficial aspects of the workshops for you as an instructor?

Probes: That's interesting, tell me more about that.

If you were designing these workshops in the future, how would you improve them?

Probes: Any ideas of how to best do that?

What areas do you feel you need more training in?

Probes: Why do you say that? What would be the best avenue(s) for receiving that training?

Closure

Though there were many different opinions about _____, it appears unanimous that _____. Does anyone see it differently? It seems most of you agree _____, but some think that _____. Does anyone want to add or clarify an opinion on this?

Is there any other information regarding your experience with or following the workshops that you think would be useful for me to know?

Thank you very much for coming this afternoon. Your time is very much appreciated and your comments have been very helpful.

Chapter 4 ANALYZING QUALITATIVE DATA

What Is Qualitative Analysis?

Qualitative modes of data analysis provide ways of discerning, examining, comparing and contrasting, and interpreting meaningful patterns or themes. Meaningfulness is determined by the particular goals and objectives of the project at hand: the same data can be analyzed and synthesized from multiple angles depending on the particular research or evaluation questions being addressed. The varieties of approaches—including ethnography, narrative analysis, discourse analysis, and textual analysis—correspond to different types of data, disciplinary traditions, objectives, and philosophical orientations. However, all share several common characteristics that distinguish them from quantitative analytic approaches.

In quantitative analysis, numbers and what they stand for are the material of analysis. By contrast, qualitative analysis deals in words and is guided by fewer universal rules and standardized procedures than statistical analysis.

We have few agreed-on canons for qualitative data analysis, in the sense of shared ground rules for drawing conclusions and verifying their sturdiness (Miles and Huberman, 1984).

This relative lack of standardization is at once a source of versatility and the focus of considerable misunderstanding. That qualitative analysts will not specify uniform procedures to follow in all cases draws critical fire from researchers who question whether analysis can be truly rigorous in the absence of such universal criteria; in fact, these analysts may have helped to invite this criticism by failing to adequately articulate their standards for assessing qualitative analyses, or even denying that such standards are possible. Their stance has fed a fundamentally mistaken but relatively common idea of qualitative analysis as unsystematic, undisciplined, and “purely subjective.”

Although distinctly different from quantitative statistical analyses both in procedures and goals, good qualitative analysis is both systematic and intensely disciplined.

Although distinctly different from quantitative statistical analysis both in procedures and goals, good qualitative analysis is both systematic and intensely disciplined. If not “objective” in the strict positivist sense, qualitative analysis is arguably replicable insofar as others can be “walked through” the analyst's thought processes and assumptions. Timing also works quite differently in qualitative evaluation. Quantitative evaluation is more easily divided into discrete stages of instrument development, data collection, data processing, and data analysis. By contrast, in qualitative evaluation, data collection and data analysis are not temporally discrete stages: as soon as the first pieces of data are collected, the evaluator begins the process of making sense of the information. Moreover, the different processes involved in qualitative analysis also overlap in time. Part of what distinguishes qualitative analysis is a loop-like pattern of multiple rounds of revisiting the data as additional questions emerge, new connections are unearthed, and more complex formulations develop along with a deepening understanding of the material. Qualitative analysis is fundamentally an *iterative* set of processes.

At the simplest level, qualitative analysis involves examining the assembled relevant data to determine how they answer the evaluation question(s) at hand. However, the data are apt to be in formats that are unusual for quantitative evaluators, thereby complicating this task. In quantitative analysis of survey results, for example, frequency distributions of responses to specific items on a questionnaire often structure the discussion and analysis of findings. By contrast, qualitative data most often occur in more embedded and less easily reducible or distillable forms than quantitative data. For example, a relevant “piece” of qualitative data might be interspersed portions of an interview transcript, multiple excerpts from a set of field notes, or a comment or cluster of comments from a focus group.

Throughout the course of qualitative analysis, the analyst should be asking and reasking the following questions:

- What patterns and common themes emerge in responses dealing with specific items? How do these patterns (or lack thereof) help to illuminate the broader study question(s)?
- Are there any deviations from these patterns? If yes, are there any factors that might explain these atypical responses?
- What interesting stories emerge from the responses? How can these stories help to illuminate the broader study question(s)?
- Do any of these patterns or findings suggest that additional data may need to be collected? Do any of the study questions need to be revised?

- Do the patterns that emerge corroborate the findings of any corresponding qualitative analyses that have been conducted? If not, what might explain these discrepancies?

Two basic forms of qualitative analysis, essentially the same in their underlying logic, will be discussed: **intra-case analysis** and **cross-case analysis**. A case may be differently defined for different analytic purposes. Depending on the situation, a case could be a single individual, a focus group session, or a program site (Berkowitz, 1996). In terms of the hypothetical project described in Chapter 2, a case will be a single campus. Intra-case analysis will examine a single project site, and cross-case analysis will systematically compare and contrast the eight campuses.

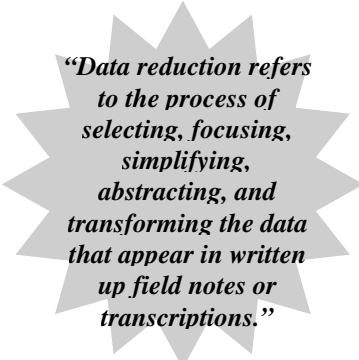
Processes in Qualitative Analysis

Qualitative analysts are justifiably wary of creating an unduly reductionistic or mechanistic picture of an undeniably complex, iterative set of processes. Nonetheless, evaluators have identified a few basic commonalities in the process of making sense of qualitative data. In this chapter we have adopted the framework developed by Miles and Huberman (1994) to describe the major phases of data analysis: data reduction, data display, and conclusion drawing and verification.

Data Reduction

First, the mass of data has to be organized and somehow meaningfully reduced or reconfigured. Miles and Huberman (1994) describe this first of their three elements of qualitative data analysis as data reduction. “Data reduction refers to the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written up field notes or transcriptions.” Not only do the data need to be condensed for the sake of manageability, they also have to be transformed so they can be made intelligible in terms of the issues being addressed.

Data reduction often forces choices about which aspects of the assembled data should be emphasized, minimized, or set aside completely for the purposes of the project at hand. Beginners often fail to understand that even at this stage, the data do not speak for themselves. A common mistake many people make in quantitative as well as qualitative analysis, in a vain effort to remain “perfectly



“Data reduction refers to the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written up field notes or transcriptions.”

objective,” is to present a large volume of unassimilated and uncategorized data for the reader's consumption.

In qualitative analysis, the analyst decides which data are to be singled out for description according to principles of selectivity. This usually involves some combination of deductive and inductive analysis. While initial categorizations are shaped by preestablished study questions, the qualitative analyst should remain open to inducing new meanings from the data available.

In evaluation, such as the hypothetical evaluation project in this handbook, data reduction should be guided primarily by the need to address the salient evaluation question(s). This selective winnowing is difficult, both because qualitative data can be very rich, and because the person who analyzes the data also often played a direct, personal role in collecting them. The words that make up qualitative analysis represent real people, places, and events far more concretely than the numbers in quantitative data sets, a reality that can make cutting any of it quite painful. But the acid test has to be the relevance of the particular data for answering particular questions. For example, a formative evaluation question for the hypothetical study might be whether the presentations were suitable for all participants. Focus group participants may have had a number of interesting things to say about the presentations, but remarks that only tangentially relate to the issue of suitability may have to be bracketed or ignored. Similarly, a participant's comments on his department chair that are unrelated to issues of program implementation or impact, however fascinating, should not be incorporated into the final report. The approach to data reduction is the same for intra-case and cross-case analysis.

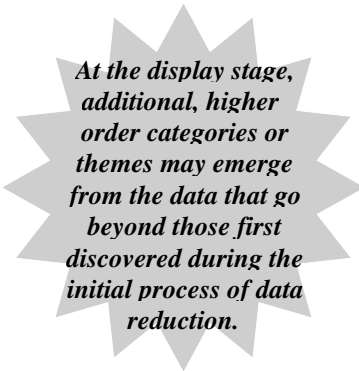
With the hypothetical project of Chapter 2 in mind, it is illustrative to consider ways of reducing data collected to address the question “what did participating faculty do to share knowledge with nonparticipating faculty?” The first step in an intra-case analysis of the issue is to examine all the relevant data sources to extract a description of what they say about the sharing of knowledge between participating and nonparticipating faculty on the one campus. Included might be information from focus groups, observations, and indepth interviews of key informants, such as the department chair. The most salient portions of the data are likely to be concentrated in certain sections of the focus group transcripts (or write-ups) and indepth interviews with the department chair. However, it is best to also quickly peruse all notes for relevant data that may be scattered throughout.

In initiating the process of data reduction, the focus is on distilling what the different respondent groups suggested about the activities used to share knowledge between faculty who participated in the

project and those who did not. How does what the participating faculty say compare to what the nonparticipating faculty and the department chair report about knowledge sharing and adoption of new practices? In setting out these differences and similarities, it is important not to so “flatten” or reduce the data that they sound like close-ended survey responses. The tendency to treat qualitative data in this manner is not uncommon among analysts trained in quantitative approaches. Not surprisingly, the result is to make qualitative analysis look like watered down survey research with a tiny sample size. Approaching qualitative analysis in this fashion unfairly and unnecessarily dilutes the richness of the data and, thus, inadvertently undermines one of the greatest strengths of the qualitative approach.

Answering the question about knowledge sharing in a truly qualitative way should go beyond enumerating a list of knowledge-sharing activities to also probe the respondents' assessments of the relative effectiveness of these activities, as well as their reasons for believing some more effective than others. Apart from exploring the specific content of the respondents' views, it is also a good idea to take note of the relative frequency with which different issues are raised, as well as the intensity with which they are expressed.

Data Display



At the display stage, additional, higher order categories or themes may emerge from the data that go beyond those first discovered during the initial process of data reduction.

Data display is the second element or level in Miles and Huberman's (1994) model of qualitative data analysis. Data display goes a step beyond data reduction to provide “an organized, compressed assembly of information that permits conclusion drawing...” A display can be an extended piece of text or a diagram, chart, or matrix that provides a new way of arranging and thinking about the more textually embedded data. Data displays, whether in word or diagrammatic form, allow the analyst to extrapolate from the data enough to begin to discern systematic patterns and interrelationships. At the display stage, additional, higher order categories or themes may emerge from the data that go beyond those first discovered during the initial process of data reduction.

From the perspective of program evaluation, data display can be extremely helpful in identifying why a system (e.g., a given program or project) is or is not working well and what might be done to change it. The overarching issue of why some projects work better or are more successful than others almost always drives the analytic process in any evaluation. In our hypothetical evaluation example, faculty from all eight campuses come together at the central campus to attend workshops. In that respect, all participants are exposed to the identical program. However, implementation of teaching

techniques presented at the workshop will most likely vary from campus to campus based on factors such as the participants' personal characteristics, the differing demographics of the student bodies, and differences in the university and departmental characteristics (e.g., size of the student body, organization of preservice courses, department chair's support of the program goals, departmental receptivity to change and innovation). The qualitative analyst will need to discern patterns of interrelationships to suggest why the project promoted more change on some campuses than on others.

One technique for displaying narrative data is to develop a series of flow charts that map out any critical paths, decision points, and supporting evidence that emerge from establishing the data for a single site. After the first flow chart has been developed, the process can be repeated for all remaining sites. Analysts may (1) use the data from subsequent sites to modify the original flow chart; (2) prepare an independent flow chart for *each* site; and/or (3) prepare a single flow chart for some events (if most sites adopted a generic approach) and multiple flow charts for others. Examination of the data display across the eight campuses might produce a finding that implementation proceeded more quickly and effectively on those campuses where the department chair was highly supportive of trying new approaches to teaching but was stymied and delayed when department chairs had misgivings about making changes to a tried-and-true system.

Data display for intra-case analysis. Exhibit 10 presents a data display matrix for analyzing patterns of response concerning perceptions and assessments of knowledge-sharing activities for one campus. We have assumed that three respondent units—participating faculty, nonparticipating faculty, and department chairs—have been asked similar questions. Looking at column (a), it is interesting that the three respondent groups were not in total agreement even on which activities they named. Only the participants considered e-mail a means of sharing what they had learned in the program with their colleagues. The nonparticipant colleagues apparently viewed the situation differently, because they did not include e-mail in their list. The department chair—perhaps because she was unaware they were taking place—did not mention e-mail or informal interchanges as knowledge-sharing activities.

Column (b) shows which activities each group considered most effective as a way of sharing knowledge, in order of perceived importance; column (c) summarizes the respondents' reasons for regarding those particular activities as most effective. Looking down column (b), we can see that there is some overlap across groups—for example, both the participants and the department chair believed structured seminars were the most effective knowledge-sharing

activity. Nonparticipants saw the structured seminars as better than lunchtime meetings, but not as effective as informal interchanges.

Exhibit 10.			
Data matrix for Campus A: What was done to share knowledge			
Respondent group	(a) Activities named	(b) Which most effective	(c) Why
Participants	<ul style="list-style-type: none"> • Structured seminars • E-mail • Informal interchanges • Lunch time meetings 	<ul style="list-style-type: none"> • Structured seminars • E-mail 	<ul style="list-style-type: none"> • Concise way of communicating a lot of information
Nonparticipants	<ul style="list-style-type: none"> • Structured seminars • Informal interchanges • Lunch time meetings 	<ul style="list-style-type: none"> • Informal interchanges • Structured seminars 	<ul style="list-style-type: none"> • Easier to assimilate information in less formal settings • Smaller bits of information at a time
Department chair	<ul style="list-style-type: none"> • Structured seminars • Lunch time meetings 	<ul style="list-style-type: none"> • Structured seminars 	<ul style="list-style-type: none"> • Highest attendance by nonparticipants • Most comments (positive) to chair

Simply knowing what each set of respondents considered most effective, without knowing why, would leave out an important piece of the analytic puzzle. It would rob the qualitative analyst of the chance to probe potentially meaningful variations in underlying conceptions of what defines effectiveness in an educational exchange. For example, even though both participating faculty and the department chair agreed on the structured seminars as the most effective knowledge-sharing activity, they gave somewhat different reasons for making this claim. The participants saw the seminars as the most effective way of communicating a lot of information concisely. The department chair used indirect indicators—attendance rates of nonparticipants at the seminars, as well as favorable comments on the seminars volunteered to her—to formulate her judgment of effectiveness. It is important to recognize the different bases on which the respondents reached the same conclusions.

Several points concerning qualitative analysis emerge from this relatively straightforward and preliminary exercise. First, a pattern of cross-group differences can be discerned even before we analyze the responses concerning the activities regarded as most effective, and why. The open-ended format of the question allowed each group to give its own definition of “knowledge-sharing activities.” The point of the analysis is not primarily to determine which activities

were used and how often; if that were the major purpose of asking this question, there would be far more efficient ways (e.g., a checklist or rating scale) to find the answer. From an analytic perspective, it is more important to begin to uncover relevant group differences in perceptions.

Differences in reasons for considering one activity more effective than another might also point to different conceptions of the primary goals of the knowledge-sharing activities. Some of these variations might be attributed to the fact that the respondent groups occupy different structural positions in life and different roles in this specific situation. While both participating and nonparticipating faculty teach in the same department, in this situation the participating faculty are playing a teaching role vis-a-vis their colleagues. The data in column (c) indicate the participants see their main goal as imparting a great deal of information as concisely as possible. By contrast, the nonparticipants—in the role of students—believe they assimilate the material better when presented with smaller quantities of information in informal settings. Their different approaches to the question might reflect different perceptions based on this temporary rearrangement in their roles. The department chair occupies a different structural position in the university than either the participating or nonparticipating faculty. She may be too removed from day-to-day exchanges among the faculty to see much of what is happening on this more informal level. By the same token, her removal from the grassroots might give her a broader perspective on the subject.

Data display in cross-case analysis. The principles applied in analyzing across cases essentially parallel those employed in the intra-case analysis. Exhibit 11 shows an example of a hypothetical data display matrix that might be used for analysis of program participants' responses to the knowledge-sharing question across all eight campuses. Looking down column (a), one sees differences in the number and variety of knowledge-sharing activities named by participating faculty at the eight schools. Brown bag lunches, department newsletters, workshops, and dissemination of written (hard-copy) materials have been added to the list, which for branch campus A included only structured seminars, e-mail, informal interchanges, and lunchtime meetings. This expanded list probably encompasses most, if not all, such activities at the eight campuses. In addition, where applicable, we have indicated whether the nonparticipating faculty involvement in the activity was compulsory or voluntary.

In Exhibit 11, we are comparing the same group on different campuses, rather than different groups on the same campus, as in Exhibit 10. Column (b) reveals some overlap across participants in which activities were considered most effective: structured seminars were named by participants at campuses A and C, brown bag lunches

Exhibit 11. Participants' views of information sharing at eight campuses			
Branch campus	(a) Activities named	(b) Which most effective	(c) Why
A	<ul style="list-style-type: none"> • Structured seminar (voluntary) • E-mail • Informal interchanges • Lunchtime meetings 	<ul style="list-style-type: none"> • Structured seminar • E-mail 	<ul style="list-style-type: none"> • Concise way of communicating a lot of information
B	<ul style="list-style-type: none"> • Brown bags • E-mail • Department newsletter 	<ul style="list-style-type: none"> • Brown bags 	<ul style="list-style-type: none"> • Most interactive
C	<ul style="list-style-type: none"> • Workshops (voluntary) • Structured seminar (compulsory) 	<ul style="list-style-type: none"> • Structured seminar 	<ul style="list-style-type: none"> • Compulsory • Structured format works well
D	<ul style="list-style-type: none"> • Informal interchanges • Dissemination of written materials 	<ul style="list-style-type: none"> • Combination of the two 	<ul style="list-style-type: none"> • Dissemination important but not enough without "personal touch"
E	<ul style="list-style-type: none"> • Structured seminars (compulsory) • Workshops (voluntary) 	<ul style="list-style-type: none"> • Workshops 	<ul style="list-style-type: none"> • Voluntary hands-on approach works best
F	<ul style="list-style-type: none"> • E-mail • Dissemination of materials • Workshops (compulsory) 	<ul style="list-style-type: none"> • Dissemination of materials 	<ul style="list-style-type: none"> • Not everyone regularly uses e-mail • Compulsory workshops rested as coercive
G	<ul style="list-style-type: none"> • Structured seminar • Informal interchanges • Lunch meetings 	<ul style="list-style-type: none"> • Lunch meetings 	<ul style="list-style-type: none"> • Best time
H	<ul style="list-style-type: none"> • Brown bags • E-mail • Dissemination of materials 	<ul style="list-style-type: none"> • Brown bags 	<ul style="list-style-type: none"> • Relaxed environment

by those at campuses B and H. However, as in Exhibit 10, the primary reasons for naming these activities were not always the same. Brown bag lunches were deemed most effective because of their interactive nature (campus B) and the relaxed environment in which they took place (campus H), both suggesting a preference for less formal learning situations. However, while campus A participants judged voluntary structured seminars the most effective way to communicate a great deal of information, campus C participants also liked that the structured seminars on their campus were compulsory. Participants at both campuses appear to favor structure, but may part company on whether requiring attendance is a good idea. The voluntary/compulsory distinction was added to illustrate different aspects of effective knowledge sharing that might prove analytically relevant.

It would also be worthwhile to examine the reasons participants gave for deeming one activity more effective than another, regardless of the activity. Data in column (c) show a tendency for participants on campuses B, D, E, F, and H to prefer voluntary, informal, hands-on, personal approaches. By contrast, those from campuses A and C seemed to favor more structure (although they may disagree on voluntary versus compulsory approaches). The answer supplied for campus G (“best time”) is ambiguous and requires returning to the transcripts to see if more material can be found to clarify this response.

To have included all the knowledge-sharing information from four different respondent groups on all eight campuses in a single matrix would have been quite complicated. Therefore, for clarity's sake, we present only the participating faculty responses. However, to complete the cross-case analysis of this evaluation question, the same procedure should be followed—if not in matrix format, then conceptually—for nonparticipating faculty and department chairpersons. For each group, the analysis would be modeled on the above example. It would be aimed at identifying important similarities and differences in what the respondents said or observed and exploring the possible bases for these patterns at different campuses. Much of qualitative analysis, whether intra-case or cross-case, is structured by what Glaser and Strauss (1967) called the “method of constant comparison,” an intellectually disciplined process of comparing and contrasting across instances to establish significant patterns, then further questioning and refinement of these patterns as part of an ongoing analytic process.

Conclusion Drawing and Verification

Conclusion drawing involves stepping back to consider what the analyzed data mean and to assess their implications for the questions at hand.

If the unprocessed “data heap” is the result of not taking responsibility for shaping the “story line” of the analysis, the opposite tendency is to take conclusion drawing well beyond what the data reasonably warrant.

This activity is the third element of qualitative analysis. Conclusion drawing involves stepping back to consider what the analyzed data mean and to assess their implications for the questions at hand.⁶ Verification, integrally linked to conclusion drawing, entails revisiting the data as many times as necessary to cross-check or verify these emergent conclusions. “The meanings emerging from the data have to be tested for their plausibility, their sturdiness, their ‘confirmability’—that is, their validity” (Miles and Huberman, 1994, p. 11). Validity means something different in this context than in quantitative evaluation, where it is a technical term that refers quite specifically to whether a given construct measures what it purports to measure. Here validity encompasses a much broader concern for whether the conclusions being drawn from the data are credible, defensible, warranted, and able to withstand alternative explanations.

For many qualitative evaluators, it is above all this third phase that gives qualitative analysis its special appeal. At the same time, it is probably also the facet that quantitative evaluators and others steeped in traditional quantitative techniques find most disquieting. Once qualitative analysts begin to move beyond cautious analysis of the factual data, the critics ask, what is to guarantee that they are not engaging in purely speculative flights of fancy? Indeed, their concerns are not entirely unfounded. If the unprocessed “data heap” is the result of not taking responsibility for shaping the “story line” of the analysis, the opposite tendency is to take conclusion drawing well beyond what the data reasonably warrant or to prematurely leap to conclusions and draw implications without giving the data proper scrutiny.

The question about knowledge sharing provides a good example. The underlying expectation, or hope, is for a diffusion effort, wherein participating faculty stimulate innovation in teaching mathematics among their colleagues. A cross-case finding might be that participating faculty at three of the eight campuses made active, ongoing efforts to share their new knowledge with their colleagues in a variety of formal and informal settings. At two other campuses, initial efforts at sharing started strong but soon fizzled out and were not continued. In the remaining three cases, one or two faculty participants shared bits and pieces of what they had learned with a few selected colleagues on an ad hoc basis, but otherwise took no steps to diffuse their new knowledge and skills more broadly.

⁶When qualitative data are used as a precursor to the design/development of quantitative instruments, this step may be postponed. Reducing the data and looking for relationships will provide adequate information for developing other instruments.

Taking these findings at face value might lead one to conclude that the project had largely failed in encouraging diffusion of new pedagogical knowledge and skills to nonparticipating faculty. After all, such sharing occurred in the desired fashion at only three of the eight campuses. However, before jumping ahead to conclude that the project was disappointing in this respect, or to generalize beyond this case to other similar efforts at spreading pedagogic innovations among faculty, it is vital to examine more closely the likely reasons why sharing among participating and nonparticipating faculty occurred, and where and how it did. The analysts would first look for factors distinguishing the three campuses where ongoing organized efforts at sharing did occur from those where such efforts were either not sustained or occurred in largely piecemeal fashion. However, it will also be important to differentiate among the less successful sites to tease out factors related both to the extent of sharing and the degree to which activities were sustained.

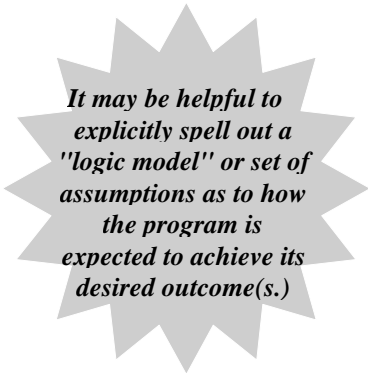
One possible hypothesis would be that successfully sustaining organized efforts at sharing on an ongoing basis requires structural supports at the departmental level and/or conducive environmental conditions at the home campus. In the absence of these supports, a great burst of energy and enthusiasm at the beginning of the academic year will quickly give way under the pressure of the myriad demands, as happened for the second group of two campuses. Similarly, under most circumstances, the individual good will of one or two participating faculty on a campus will in itself be insufficient to generate the type and level of exchange that would make a difference to the nonparticipating faculty (the third set of campuses).

At the three "successful" sites, for example, faculty schedules may allow regularly scheduled common periods for colleagues to share ideas and information. In addition, participation in such events might be encouraged by the department chair, and possibly even considered as a factor in making promotion and tenure decisions. The department might also contribute a few dollars for refreshments in order to promote a more informal, relaxed atmosphere at these activities. In other words, at the campuses where sharing occurred as desired, conditions were conducive in one or more ways: a new time slot did not have to be carved out of already crowded faculty schedules, the department chair did more than simply pay "lip service" to the importance of sharing (faculty are usually quite astute at picking up on what really matters in departmental culture), and efforts were made to create a relaxed ambiance for transfer of knowledge.

At some of the other campuses, structural conditions might not be conducive, in that classes are taught continuously from 8 a.m. through 8 p.m., with faculty coming and going at different times and on alternating days. At another campus, scheduling might not present so great a hurdle. However, the department chair may be so busy that despite philosophic agreement with the importance of diffusing the newly learned skills, she can do little to actively encourage sharing among participating and nonparticipating faculty. In this case, it is not structural conditions or lukewarm support so much as competing priorities and the department chair's failure to act concretely on her commitment that stood in the way. By contrast, at another campus, the department chairperson may publicly acknowledge the goals of the project but really believe it a waste of time and resources. His failure to support sharing activities among his faculty stems from more deeply rooted misgivings about the value and viability of the project. This distinction might not seem to matter, given that the outcome was the same on both campuses (sharing did not occur as desired). However, from the perspective of an evaluation researcher, whether the department chair believes in the project could make a major difference to what would have to be done to change the outcome.

We have begun to develop a reasonably coherent explanation for the cross-site variations in the degree and nature of sharing taking place between participating and nonparticipating faculty. Arriving at this point required stepping back and systematically examining and re-examining the data, using a variety of what Miles and Huberman (1994, pp. 245-262) call "tactics for generating meaning." They describe 13 such tactics, including noting patterns and themes, clustering cases, making contrasts and comparisons, partitioning variables, and subsuming particulars in the general. Qualitative analysts typically employ some or all of these, simultaneously and iteratively, in drawing conclusions.

One factor that can impede conclusion drawing in evaluation studies is that the theoretical or logical assumptions underlying the research are often left unstated. In this example, as discussed above, these are assumptions or expectations about knowledge sharing and diffusion of innovative practices from participating to nonparticipating faculty, and, by extension, to their students. For the analyst to be in a position to take advantage of conclusion-drawing opportunities, he or she must be able to recognize and address these assumptions, which are often only implicit in the evaluation questions. Toward that end, it may be helpful to explicitly spell out a "logic model" or set of assumptions as to how the program is expected to achieve its desired outcome(s.) Recognizing these assumptions becomes even more important when there is a need or

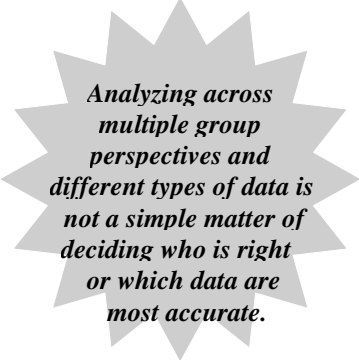


It may be helpful to explicitly spell out a "logic model" or set of assumptions as to how the program is expected to achieve its desired outcome(s.)

desire to place the findings from a single evaluation into wider comparative context vis-a-vis other program evaluations.

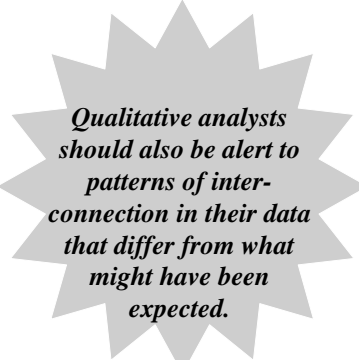
Once having created an apparently credible explanation for variations in the extent and kind of sharing that occurs between participating and nonparticipating faculty across the eight campuses, how can the analyst verify the validity—or truth value—of this interpretation of the data? Miles and Huberman (1994, pp. 262-277) outline 13 tactics for testing or confirming findings, all of which address the need to build systematic "safeguards against self-delusion" (p. 265) into the process of analysis. We will discuss only a few of these, which have particular relevance for the example at hand and emphasize critical contrasts between quantitative and qualitative analytic approaches. However, two points are very important to stress at the outset: several of the most important safeguards on validity—such as using multiple sources and modes of evidence—must be built into the design from the beginning; and the analytic objective is to create a plausible, empirically grounded account that is maximally responsive to the evaluation questions at hand. As the authors note: "You are not looking for one account, forsaking all others, but for the best of several alternative accounts" (p. 274).

One issue of analytic validity that often arises concerns the need to weigh evidence drawn from multiple sources and based on different data collection modes, such as self-reported interview responses and observational data. Triangulation of data sources and modes is critical, but the results may not necessarily corroborate one another, and may even conflict. For example, another of the summative evaluation questions proposed in Chapter 2 concerns the extent to which nonparticipating faculty adopt new concepts and practices in their teaching. Answering this question relies on a combination of observations, self-reported data from participant focus groups, and indepth interviews with department chairs and nonparticipating faculty. In this case, there is a possibility that the observational data might be at odds with the self-reported data from one or more of the respondent groups. For example, when interviewed, the vast majority of nonparticipating faculty might say, and really believe, that they are applying project-related innovative principles in their teaching. However, the observers may see very little behavioral evidence that these principles are actually influencing teaching practices in these faculty members' classrooms. It would be easy to brush off this finding by concluding that the nonparticipants are saving face by parroting what they believe they are expected to say about their teaching. But there are other, more analytically interesting, possibilities. Perhaps the nonparticipants have an incomplete understanding of these principles, or they were not adequately trained in how to translate them effectively into classroom practice.



Analyzing across multiple group perspectives and different types of data is not a simple matter of deciding who is right or which data are most accurate.

The important point is that analyzing across multiple group perspectives and different types of data is not a simple matter of deciding who is right or which data are most accurate. Weighing the evidence is a more subtle and delicate matter of hearing each group's viewpoint, while still recognizing that any single perspective is partial and relative to the respondent's experiences and social position. Moreover, as noted above, respondents' perceptions are no more or less real than observations. In fact, discrepancies between self-reported and observational data may reveal profitable topics or areas for further analysis. It is the analyst's job to weave the various voices and sources together in a narrative that responds to the relevant evaluation question(s). The more artfully this is done, the simpler, more natural it appears to the reader. To go to the trouble to collect various types of data and listen to different voices, only to pound the information into a flattened picture, is to do a real disservice to qualitative analysis. However, if there is a reason to believe that some of the data are stronger than others (some of the respondents are highly knowledgeable on the subject, while others are not), it is appropriate to give these responses greater weight in the analysis.



Qualitative analysts should also be alert to patterns of inter-connection in their data that differ from what might have been expected.

Qualitative analysts should also be alert to patterns of inter-connection in their data that differ from what might have been expected. Miles and Huberman define these as “following up surprises” (1994, p. 270). For instance, at one campus, systematically comparing participating and nonparticipating faculty responses to the question about knowledge-sharing activities (see Exhibit 10) might reveal few apparent cross-group differences. However, closer examination of the two sets of transcripts might show meaningful differences in perceptions dividing along other, less expected lines. For purposes of this evaluation, it was tacitly assumed that the relevant distinctions between faculty would most likely be between those who had and had not participated in the project. However, both groups also share a history as faculty in the same department. Therefore, other factors—such as prior personal ties—might have overridden the participant/nonparticipant faculty distinction. One strength of qualitative analysis is its potential to discover and manipulate these kinds of unexpected patterns, which can often be very informative. To do this requires an ability to listen for, and be receptive to, surprises.

Unlike quantitative researchers, who need to explain away deviant or exceptional cases, qualitative analysts are also usually delighted when they encounter twists in their data that present fresh analytic insights or challenges. Miles and Huberman (1994, pp. 269, 270) talk about “checking the meaning of outliers” and “using extreme cases.” In qualitative analysis deviant instances or cases that do not appear to fit the pattern or trend are not treated as outliers,

as they would be in statistical, probability-based analysis. Rather, deviant or exceptional cases should be taken as a challenge to further elaboration and verification of an evolving conclusion. For example, if the department chair strongly supports the project's aims and goals for all successful projects but one, perhaps another set of factors is fulfilling the same function(s) at the “deviant” site. Identifying those factors will, in turn, help to clarify more precisely what it is about strong leadership and belief in a project that makes a difference. Or, to elaborate on another extended example, suppose at one campus where structural conditions are not conducive to sharing between participating and nonparticipating faculty, such sharing is occurring nonetheless, spearheaded by one very committed participating faculty member. This example might suggest that a highly committed individual who is a natural leader among his faculty peers is able to overcome the structural constraints to sharing. In a sense, this “deviant” case analysis would strengthen the general conclusion by showing that it takes exceptional circumstances to override the constraints of the situation.

Elsewhere in this handbook, we noted that summative and formative evaluations are often linked by the premise that variations in project implementation will, in turn, effect differences in project outcomes. In the hypothetical example presented in this handbook, all participants were exposed to the same activities on the central campus, eliminating the possibility of analyzing the effects of differences in implementation features. However, using a different model and comparing implementation and outcomes at three different universities, with three campuses participating per university, would give some idea of what such an analysis might look like.

A display matrix for a cross-site evaluation of this type is given in Exhibit 12. The upper portion of the matrix shows how the three campuses varied in key implementation features. The bottom portion summarizes outcomes at each campus. While we would not necessarily expect a one-to-one relationship, the matrix loosely pairs implementation features with outcomes with which they might be associated. For example, workshop staffing and delivery are paired with knowledge-sharing activities, accuracy of workshop content with curricular change. However, there is nothing to preclude looking for a relationship between use of appropriate techniques in the workshops (formative) and curricular changes on the campuses (summative). Use of the matrix would essentially guide the analysis along the same lines as in the examples provided earlier.

Exhibit 12.					
Matrix of cross-case analysis linking implementation and outcome factors					
Implementation Features					
Branch campus	Workshops delivered and staffed as planned?	Content accurate/ up to date?	Appropriate techniques used?	Materials available?	Suitable presentation?
Campus A	Yes	Yes	For most participants	Yes, but delayed	Mostly
Campus B	No	Yes	Yes	No	Very mixed reviews
Campus C	Mostly	Yes	For a few participants	Yes	Some
Outcome Features - Participating Campuses					
Branch campus	Knowledge -sharing with nonparticipants?	Curricular changes?	Changes to exams and requirements?	Expenditures?	Students more interested/ active in class?
Campus A	High level	Many	Some	No	Some campuses
Campus B	Low level	Many	Many	Yes	Mostly participants' students
Campus C	Moderate level	Only a few	Few	Yes	Only minor improvement

In this cross-site analysis, the overarching question would address the similarities and differences across these three sites—in terms of project implementation, outcomes, and the connection between them—and investigate the bases of these differences. Was one of the projects discernibly more successful than others, either overall or in particular areas—and if so, what factors or configurations of factors seem to have contributed to these successes? The analysis would then continue through multiple iterations until a satisfactory resolution is achieved.

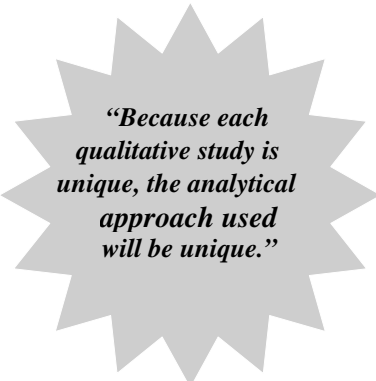
Summary: Judging the Quality of Qualitative Analysis

Issues surrounding the value and uses of conclusion drawing and verification in qualitative analysis take us back to larger questions raised at the outset about how to judge the validity and quality of qualitative research. A lively debate rages on these and related issues. It goes beyond the scope of this chapter to enter this discussion in any depth, but it is worthwhile to summarize emerging areas of agreement.

First, although stated in different ways, there is broad consensus concerning the qualitative analyst's need to be self-aware, honest, and reflective about the analytic process. Analysis is not just the end product, it is also the repertoire of processes used to arrive at that particular place. In qualitative analysis, it is not necessary or even desirable that anyone else who did a similar study should find exactly the same thing or interpret his or her findings in precisely the same way. However, once the notion of analysis as a set of uniform, impersonal, universally applicable procedures is set aside, qualitative analysts are obliged to describe and discuss how they did their work in ways that are, at the very least, accessible to other researchers. Open and honest presentation of analytic processes provides an important check on an individual analyst's tendencies to get carried away, allowing others to judge for themselves whether the analysis and interpretation are credible in light of the data.

Second, qualitative analysis, as all of qualitative research, is in some ways craftsmanship (Kvale, 1995). There *is* such a thing as poorly crafted or bad qualitative analysis, and despite their reluctance to issue universal criteria, seasoned qualitative researchers of different bents can still usually agree when they see an example of it. Analysts should be judged partly in terms of how skillfully, artfully, and persuasively they craft an argument or tell a story. Does the analysis flow well and make sense in relation to the study's objectives and the data that were presented? Is the story line clear and convincing? Is the analysis interesting, informative, provocative? Does the analyst explain how and why she or he drew certain conclusions, or on what bases she or he excluded other possible interpretations? These are the kinds of questions that can and should be asked in judging the quality of qualitative analyses. In evaluation studies, analysts are often called upon to move from conclusions to recommendations for improving programs and policies. The recommendations should fit with the findings and with the analysts' understanding of the context or milieu of the study. It is often useful to bring in stakeholders at the point of "translating" analytic conclusions to implications for action.

As should by now be obvious, it is truly a mistake to imagine that qualitative analysis is easy or can be done by untrained novices. As Patton (1990) comments:



"Because each qualitative study is unique, the analytical approach used will be unique."

Applying guidelines requires judgment and creativity. Because each qualitative study is unique, the analytical approach used will be unique. Because qualitative inquiry depends, at every stage, on the skills, training, insights, and capabilities of the researcher, qualitative analysis ultimately depends on the analytical intellect and style of the analyst. The human factor is the greatest

strength and the fundamental weakness of qualitative inquiry and analysis.

Practical Advice in Conducting Qualitative Analyses

Start the analysis right away and keep a running account of it in your notes: It cannot be overstressed that analysis should begin almost in tandem with data collection, and that it is an iterative set of processes that continues over the course of the field work and beyond. It is generally helpful for field notes or focus group or interview summaries to include a section containing comments, tentative interpretations, or emerging hypotheses. These may eventually be overturned or rejected, and will almost certainly be refined as more data are collected. But they provide an important account of the unfolding analysis and the internal dialogue that accompanied the process.

Involve more than one person: Two heads are better than one, and three may be better still. Qualitative analysis need not, and in many cases probably should not, be a solitary process. It is wise to bring more than one person into the analytic process to serve as a cross-check, sounding board, and source of new ideas and cross-fertilization. It is best if all analysts know something about qualitative analysis as well as the substantive issues involved. If it is impossible or impractical for a second or third person to play a central role, his or her skills may still be tapped in a more limited way. For instance, someone might review only certain portions of a set of transcripts.

Leave enough time and money for analysis and writing: Analyzing and writing up qualitative data almost always takes more time, thought, and effort than anticipated. A budget that assumes a week of analysis time and a week of writing for a project that takes a year's worth of field work is highly unrealistic. Along with revealing a lack of understanding of the nature of qualitative analysis, failing to build in enough time and money to complete this process adequately is probably the major reason why evaluation reports that include qualitative data can disappoint.

Be selective when using computer software packages in qualitative analysis: A great proliferation of software packages that can be used to aid analysis of qualitative data has been developed in recent years. Most of these packages were reviewed by Weitzman and Miles (1995), who grouped them into six types: word processors, word retrievers, textbase managers, code-and-retrieve programs, code-based theory builders, and conceptual network

builders. All have strengths and weaknesses. Weitzman and Miles suggested that when selecting a given package, researchers should think about the amount, types, and sources of data to be analyzed and the types of analyses that will be performed.

Two caveats are in order. First, computer software packages for qualitative data analysis essentially aid in the manipulation of relevant segments of text. While helpful in marking, coding, and moving data segments more quickly and efficiently than can be done manually, the software cannot determine meaningful categories for coding and analysis or define salient themes or factors. In qualitative analysis, as seen above, concepts must take precedence over mechanics: the analytic underpinnings of the procedures must still be supplied by the analyst. Software packages cannot and should not be used as a way of evading the hard intellectual labor of qualitative analysis. Second, since it takes time and resources to become adept in utilizing a given software package and learning its peculiarities, researchers may want to consider whether the scope of their project, or their ongoing needs, truly warrant the investment.

References

- Berkowitz, S. (1996). Using Qualitative and Mixed Method Approaches. Chapter 4 in *Needs Assessment: A Creative and Practical Guide for Social Scientists*, R. Reviere, S. Berkowitz, C.C. Carter, and C. Graves-Ferguson, Eds. Washington, DC: Taylor & Francis.
- Glaser, B., and Strauss, A. (1967). *The Discovery of Grounded Theory*. Chicago: Aldine.
- Kvale, S. (1995). The Social Construction of Validity. *Qualitative Inquiry*, (1):19-40.
- Miles, M.B., and Huberman, A.M. (1984). *Qualitative Data Analysis*, 16. Newbury Park, CA: Sage.
- Miles, M.B., and Huberman, A.M. (1994). *Qualitative Data Analysis*, 2nd Ed., p. 10-12. Newbury Park, CA: Sage.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*, 2nd Ed. Newbury Park: CA, Sage.
- Weitzman, E.A., and Miles, M.B. (1995). *A Software Sourcebook: Computer Programs for Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Other Recommended Reading

Coffey, A., and Atkinson, P. (1996). *Making Sense of Qualitative Data: Complementary Research Strategies*. Thousand Oaks, CA: Sage.

Howe, K., and Eisenhart, M. (1990). Standards for Qualitative (and Quantitative) Research: A Prolegomenon. *Educational Researcher*, 19(4):2-9.

Wolcott, H.F. (1994). *Transforming Qualitative Data: Description, Analysis and Interpretation*, Thousand Oaks: CA, Sage.

PART III.

DESIGNING AND REPORTING

MIXED METHOD

EVALUATIONS

OVERVIEW OF THE DESIGN PROCESS FOR MIXED METHOD EVALUATIONS

One size does not fit all. Consequently, when it comes to designing an evaluation, experience has proven that the evaluator must keep in mind that the specific questions being addressed and the audience for the answers must influence the selection of an evaluation design and tools for data collection.

Chapter 2 of the earlier *User-Friendly Handbook for Project Evaluation* (National Science Foundation, 1993) deals at length with designing and implementing an evaluation, identifying the following steps for carrying out an evaluation:

- Developing evaluation questions;
- Matching questions with appropriate information-gathering techniques;
- Collecting data;
- Analyzing the data; and
- Providing information to interested audiences.

Readers of this volume who are unfamiliar with the overall process are urged to read that chapter. In this chapter, we are briefly reviewing the process of designing an evaluation, including the development of evaluation questions, the selection of data collection methodologies, and related technical issues, with special attention to the advantages of mixed method designs. We are stressing mixed method designs because such designs frequently provide a more comprehensive and believable set of understandings about a project's accomplishments than studies based on either quantitative or qualitative data alone.

Developing Evaluation Questions

The development of evaluation questions consists of several steps:

- Clarifying the goals and objectives of the project;
- Identifying key stakeholders and audiences;
- Listing and prioritizing evaluation questions of interest to various stakeholders; and
- Determining which questions can be addressed given the resources and constraints for the evaluation (money, deadlines, access to informants and sites).

The process is not an easy one. To quote an experienced evaluator (Patton, 1990):

Once a group of intended evaluation users begins to take seriously the notion that they can learn from the collection and analysis of evaluative information, they soon find that there are lots of things they would like to find out. The evaluator'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.

We have developed a set of tools intended to help navigate these initial steps of evaluation design. These tools are simple forms or matrices that help to organize the information needed to identify and select among evaluation questions. Since the objectives of the formative and summative evaluations are usually different, separate forms need to be completed for each.

Worksheet 1 provides a form for briefly describing the project, the conceptual framework that led to the initiation of the project, and its proposed activities, and for summarizing its salient features. Information on this form will be used in the design effort. A side benefit of filling out this form and sharing it among project staff is that it can be used to make sure that there is a common understanding of the project's basic characteristics. Sometimes newcomers to a project, and even those who have been with it from the start, begin to develop some divergent ideas about emphases and goals.

WORKSHEET 1: DESCRIBE THE INTERVENTION

1. State the problem/question to be addressed by the project:

2. What is the intervention(s) under investigation?

3. State the conceptual framework which led to the decision to undertake this intervention and its proposed activities.


4. Who is the target group(s)?

5. Who are the stakeholders?

6. How is the project going to be managed?

7. What is the total budget for this project? How are major components budgeted?

8. List any other key points/issues.



*Stakeholder involvement
has become an
increasingly important
part of evaluation
design.*

Worksheet 2 provides a format for further describing the goals and objectives of the project in measurable terms. This step, essential in developing an evaluation design, can prove surprisingly difficult. A frequent problem is that goals or objectives may initially be stated in such global terms that it is not readily apparent how they might be measured. For example, the statement “improve the education of future mathematics and science educators” needs more refinement before it can be used as the basis for structuring an evaluation.

Worksheets 3 and 4 assist the evaluator in identifying the key stakeholders in the project and clarifying what it is each might want to address in an evaluation. Stakeholder involvement has become an important part of evaluation design, as it has been recognized that an evaluation must address the needs of individuals beyond the funding agency and the project director.

Worksheet 5 provides a tool for organizing and selecting among possible evaluation questions. It points to several criteria that should be considered. Who wants to know? Will the information be new or confirmatory? How important is the information to various stakeholders? Are there sufficient resources to collect and analyze the information needed to answer the questions? Can the question be addressed in the time available for the evaluation?

Once the set of evaluation questions is determined, the next step is selecting how each will be addressed and developing an overall evaluation design. It is at this point that decisions regarding the types and mixture of data collection methodologies, sampling, scheduling of data collection, and data analysis need to be made. These decisions are quite interdependent, and the data collection techniques selected will have important implications for both scheduling and analysis plans.

WORKSHEET 2: DESCRIBE PROJECT GOAL AND OBJECTIVE

1. Briefly describe the purpose of the project.

2. State the above in terms of a general goal.

3. State the first objective to be evaluated as clearly as you can.

4. Can this objective be broken down further? Break it down to the smallest unit. It must be clear what specifically you hope to see documented or changed.

5. Is this objective measurable (can indicators and standards be developed for it)?
If not, restate it.

6. Formulate one or more questions that will yield information about the extent to which the objective was addressed.

7. Once you have completed the above steps, go back to #3 and write the next objective.
Continue with steps 4, and 5, and 6.

WORKSHEET 3:

IDENTIFY KEY STAKEHOLDERS AND AUDIENCES

Audience	Spokesperson	Values, Interests, Expectations, etc. That Evaluation Should Address

WORKSHEET 4:
STAKEHOLDERS' INTEREST IN
POTENTIAL EVALUATION QUESTIONS

Question	Stakeholder Group(s)

WORKSHEET 5: PRIORITIZE AND ELIMINATE QUESTIONS

Take each question from worksheet 4 and apply criteria below.

Question	Which stakeholder(s)?	Importance to Stakeholders	New Data Collection?	Resources Required	Timeframe	Priority (High, Medium, Low, or Eliminate)			
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E
						H	M	L	E


Selecting Methods for Gathering the Data: The Case for Mixed Method Designs

As discussed in Chapter 1, mixed method designs can yield richer, more valid, and more reliable findings than evaluations based on either the qualitative or quantitative method alone. A further advantage is that a mixed method approach is likely to increase the acceptance of findings and conclusions by the diverse groups that have a stake in the evaluation.

When designing a mixed method evaluation, the investigator must consider two factors:

- Which is the most suitable data collection method for the type of data to be collected?
- How can the data collected be most effectively combined or integrated?

To recapitulate the earlier summary of the main differences between the two methods, qualitative methods provide a better understanding of the **context** in which the intervention is embedded; when a major goal of the evaluation is the **generalizability of findings**, quantitative data are usually needed. When the answer to an evaluation question calls for understanding the **perceptions and reactions of the target population**, a qualitative method (indepth interview, focus group) is most appropriate. If a major evaluation question calls for the assessment of the **behavior** of participants or other individuals involved in the intervention, trained observers will provide the most useful data.



While in any good evaluation data analysis is to some extent an iterative process, it is important to think things through as much as possible at the outset.

In Chapter 1, we also showed some of the many ways in which the quantitative and qualitative techniques can be combined to yield more meaningful findings. Specifically, the two methods have been successfully combined by evaluators to test the validity of results (triangulation), to improve data collection instruments, and to explain findings. A good design for mixed method evaluations should include specific plans for collecting and analyzing the data through the combined use of both methods; while it may often be difficult to come up with a detailed analysis plan at the outset, it is very useful to have such a plan when designing data collection instruments and when organizing narrative data obtained through qualitative methods. There needs to be considerable up-front thinking regarding probable data

analysis plans and strategies for synthesizing the information from various sources. Initial decisions can be made regarding the extent to which qualitative techniques will be used to provide full-blown stand-alone descriptions versus commentaries or illustrations to give greater meaning to quantitative data. Preliminary strategies for combining information from different data sources need to be formulated. Schedules for initiating the data analysis need to be established. The early findings thus generated should be used to reflect on the evaluation design and initiate any changes that might be warranted. While in any good evaluation data analysis is to some extent an iterative process, it is important to think things through as much as possible at the outset to avoid being left awash in data or with data focusing more on peripheral questions, rather than those that are germane to the study's goals and objectives (see Chapter 4; also see Miles and Huberman, 1994, and Greene, Caracelli, and Graham, 1989).

Other Considerations in Designing Mixed Method Evaluations

Sampling. Except in the rare cases when a project is very small and affects only a few participants and staff members, it will be necessary to deal with a subset of sites and/or informants for budgetary and managerial reasons. Sampling thus becomes an issue in the use of mixed methods, just as in the use of quantitative methods. However, the sampling approaches differ sharply depending on the method used.

The preferred sampling methods for quantitative studies are those that will enable researchers to make generalizations from the sample to the universe, i.e., all project participants, all sites, all parents. Random sampling is the appropriate method for this purpose. Statistically valid generalizations are seldom a goal of qualitative research; rather, the qualitative investigation is primarily interested in locating information-rich cases for study in depth. Purposeful sampling is therefore practiced, and it may take many forms. Instead of studying a random sample of a project's participants, evaluators may choose to concentrate their investigation on the lowest achievers admitted to the program. When selecting classrooms for observation of the implementation of an innovative practice, the evaluator may use deviant-case sampling, choosing one classroom where the innovation was reported "most successfully" implemented and another where major problems have been reported. Depending on the evaluation questions to be answered, many other sampling methods, including maximum variation sampling, critical case sampling, or

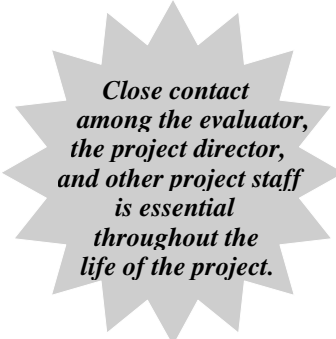
even typical case sampling, may be appropriate (Patton, 1990). When sampling subjects for indepth interviews, the investigator has considerable flexibility with respect to sample size.

In many evaluation studies, the design calls for studying a population at several points in time, e.g., students in the 9th grade and then again in the 12th grade. There are two ways of carrying out such studies that seek to measure trends. In a longitudinal approach, data are collected from the same individuals at designated time intervals; in a cross-sectional approach, new samples are drawn for each successive data collection. While in most cases, longitudinal designs that require collecting information from the same students or teachers at several points in time are best, they are often difficult and expensive to carry out because students move and teachers are reassigned. Furthermore, loss of respondents due to failure to locate or to obtain cooperation from some segment of the original sample is often a major problem. Depending on the nature of the evaluation and the size of the population studied, it may be possible to obtain good results with successive cross-sectional designs.

Timing, sequencing, frequency of data collection, and cost. The evaluation questions and the analysis plan will largely determine when data should be collected and how often focus groups, interviews, or observations should be scheduled. In mixed method designs, when the findings of qualitative data collection will affect the structuring of quantitative instruments (or vice versa), proper sequencing is crucial. As a general rule, project evaluations are strongest when data are collected at least at two points in time: before the time an innovation is first introduced, and after it has been in operation for a sizable period of time.

Throughout the design process, it is essential to keep an eye on the budgetary implications of each decision. As was pointed out in Chapter 1, costs depend not on the choice between qualitative and quantitative methods, but on the number of cases required for analysis and the quality of the data collection. Evaluators must resist the temptation to plan for a more extensive data collection than the budget can support, which may result in lower data quality or the accumulation of raw data that cannot be processed and analyzed.

Tradeoffs in the design of evaluations based on mixed methods. All evaluators find that both during the design phase, when plans are carefully crafted according to experts' recommendations, and later when fieldwork gets under way, modifications and tradeoffs become a necessity. Budget limitations, problems in accessing fieldwork sites and administrative records, and difficulties in recruiting staff with appropriate skills are among the recurring problems that should be anticipated as far as possible during the design phase, but that also may require modifying the design at a later time.



*Close contact
among the evaluator,
the project director,
and other project staff
is essential
throughout the
life of the project.*

What tradeoffs are least likely to impair the integrity and usefulness of mixed method evaluations if the evaluation plan as designed cannot be fully implemented? A good general rule for dealing with budget problems is to sacrifice the number of cases or the number of questions to be explored (this may mean ignoring the needs of some low priority stakeholders), but to preserve the depth necessary to fully and rigorously address the issues targeted.

When it comes to design modifications, it is of course essential that the evaluator be closely involved in decisionmaking. But close contact among the evaluator, the project director, and other project staff is essential throughout the life of the project. In particular, some project directors tend to see the summative evaluation as an add-on, that is, something to be done—perhaps by a contractor—after the project has been completed. But the quality of the evaluation is dependent on record keeping and data collection during the life of the project, which should be closely monitored by the evaluator.

In the next chapter, we illustrate some of the issues related to designing an evaluation, using the hypothetical example provided in Chapter 2.

References

- Greene, J.C., Caracelli, V.J., and Graham, W.F. (1989). Toward a Conceptual Framework for Mixed Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, Vol. II, No. 3.
- Miles, M.B., and Huberman, A.M. (1994). *Qualitative Data Analysis*, 2nd Ed. Newbury Park, CA: Sage
- National Science Foundation. (1993). *User-Friendly Handbook for Project Evaluation: Science, Mathematics and Technology Education*. NSF 93-152. Arlington, VA: NSF.
- Patton, M.Q. 1990. *Qualitative Evaluation and Research Methods*, 2nd Ed. Newbury Park, CA: Sage.

EVALUATION DESIGN FOR THE HYPOTHETICAL PROJECT

Step 1. Develop Evaluation Questions

As soon as the Center was notified of the grant award for the project described in Chapter 2, staff met with the evaluation specialist to discuss the focus, timing, and tentative cost allocation for the two evaluations. They agreed that although the summative evaluation was 2 years away, plans for both evaluations had to be drawn up at this time because of the need to identify stakeholders and to determine evaluation questions. The evaluation specialist requested that the faculty members named in the proposal as technical and subject matter resource persons be included in the planning stage. Following the procedure outlined in Chapter 5, the evaluation questions were specified through the use of Worksheets 1 through 5.

As the staff went through the process of developing the evaluation questions, they realized that they needed to become more knowledgeable about the characteristics of the participating institutions and especially the way courses in elementary preservice education were organized. For example, how many levels and sections were there for each course, and how were students and faculty assigned? How much autonomy did faculty have with respect to course content, examinations, etc.? What were library and other material resources? The evaluator and her staff spent time reviewing catalogues and other available documents to learn more about each campus and to identify knowledgeable informants familiar with issues of interest in planning the evaluation. The evaluator also visited three of the seven branch campuses and held informal conversations with department chairs and faculty to understand the institutional context and issues that the evaluation questions and the data collection needed to take into account.

During these campus visits, the evaluator discovered that interest and participation in the project varied considerably, as did the extent to which deans and department chairs encouraged and facilitated faculty participation. Questions to explore these issues systematically were therefore added to the formative evaluation.

The questions initially selected by the evaluation team for the formative and summative evaluation are shown in Exhibits 13 and 14.

Exhibit 13.		
Goals, stakeholders, and evaluation questions for a formative evaluation		
Project goal (implementation-related)	Evaluation questions	Stakeholders
1. To attract faculty and administrators' interest and support for project participation by eligible faculty members	Did all campuses participate? If not, what were the reasons? How was the program publicized? In what way did local administrators encourage (or discourage) participation by eligible faculty members? Were there incentives or rewards for participation? Did applicants and nonapplicants, and program completers and dropouts, differ with respect to personal and work-related characteristics (age, highest degree obtained, ethnicity, years of teaching experience, etc.)	granting agency, center administrators, project staff
2. To offer a state of the art faculty development program to improve the preparation of future teachers for elementary mathematics instruction	Were the workshops organized and staffed as planned? Were needed materials available? Were the workshops of high quality (accuracy of information, depth of coverage etc.)?	granting agency, project sponsor (center administrators), other administrators, project staff
3. To provide participants with knowledge concerning new concepts, methods, and standards in elementary math education	Was the full range of topics included in the design actually covered? Was there evidence of an increase in knowledge as a result of project participation?	center administrators, project staff
4. To provide followup and encourage networking through frequent contact among participants during the academic year	Did participants exchange information about their use of new instructional approaches? By e-mail or in other ways?	project staff
5. To identify problems in carrying out the project during year 1 for the purpose of making changes during year 2	Did problems arise? Are workshops too few, too many? Should workshop format, content, staffing be modified? Is communication adequate? Was summer session useful?	granting agency, center administrators, campus administrators, project staff, participants

Exhibit 14. Goals, stakeholders, and evaluation questions for a summative evaluation		
Project goal (outcome)	Evaluation questions	Stakeholders
1. Changes in instructional practices by participating faculty members	Did faculty who have experienced the professional development change their instructional practices? Did this vary by teachers' or by students' characteristics? Did faculty members use the information regarding new standards, materials, and practices? What obstacles prevented implementing changes? What factors facilitated change?	granting agency, project sponsor (center), campus administrators, project staff
2. Acquisition of knowledge and changes in instructional practices by other (nonparticipating) faculty members	Did participants share knowledge acquired through the project with other faculty? Was it done formally (e.g., at faculty meetings) or informally?	granting agency, project sponsor (center), campus administrators, project staff
3. Institution-wide change in curriculum and administrative practices	Were changes made in curriculum? Examinations and other requirements? Expenditures for library and other resource materials (computers)?	granting agency, project sponsor (center), campus administrators, project staff, and campus faculty participants
4. Positive effects on career plans of students taught by participating teachers	Did students become more interested in classwork? More active participants? Did they express interest in teaching math after graduation? Did they plan to use new concepts and techniques?	granting agency, project sponsor (center), campus administrators, project staff, and campus faculty participants

Step 2. Determine Appropriate Data Sources and Data Collection Approaches to Obtain Answers to the Final Set of Evaluation Questions

This step consisted of grouping the questions that survived the prioritizing process in step 1, defining measurable objectives, and determining the best source for obtaining the information needed and the best method for collecting that information. For some questions, the choice was simple. If the project reimburses participants for travel and other attendance-related expenses, reimbursement records kept in the project office would yield information about how many participants attended each of the workshops. For most questions,

however, there might be more choices and more opportunity to take advantage of the mixed method approach. To ascertain the extent of participants' learning and skill enhancement, the source might be participants, or workshop observers, or workshop instructors and other staff. If the choice is made to rely on information provided by the participants themselves, data could be obtained in many different ways: through tests (possibly before and after the completion of the workshop series), work samples, narratives supplied by participants, self-administered questionnaires, indepth interviews, or focus group sessions. The choice should be made on the basis of methodological (which method will give us the "best" data?) and pragmatic (which method will strengthen the evaluation's credibility with stakeholders? which method can the budget accommodate?) considerations.

Source and method choices for obtaining the answers to all questions in Exhibits 13 and 14 are shown in Exhibits 15 and 16. Examining these exhibits, it becomes clear that data collection from one source can answer a number of questions. The evaluation design begins to take shape; technical issues, such as sampling decisions, number of times data should be collected, and timing of the data collections, need to be addressed at this point. Exhibit 17 summarizes the data collection plan created by the evaluation specialist and her staff for both evaluations.

The **formative** evaluation must be completed before the end of the first year to provide useful inputs for the year 2 activities. Data to be collected for this evaluation include

- Relevant information in existing records;
- Frequent interviews with project director and staff;
- Short self-administered questionnaires to be completed by participants at the conclusion of each workshop; and
- Reports from the two to four staff observers who observed the 11 workshop sessions.

In addition, the 25 year 1 participants will be assigned to one of three focus groups to be convened twice (during month 5 and after the year 1 summer session) to assess the program experience, suggest program modifications, and discuss interest in instructional innovation on their home campus.

Exhibit 15.		
Evaluation questions, data sources, and data collection methods for a formative evaluation		
Question	Source of information	Data collection methods
1. Did all campuses participate? If not, what were the reasons? How was the program publicized? In what way did local administrators encourage (or discourage) participation by eligible faculty members? Were there incentives or rewards for participation? Did applicants and nonapplicants, and program completers and dropouts, differ with respect to personal and work-related characteristics (age, highest degree obtained, ethnicity, years of teaching experience, etc.)?	project records, project director, roster of eligible applicants on each campus, campus participants	record review; interview with project director; rosters of eligible applicants on each campus (including personal characteristics, length of service, etc.), participant focus groups
2. Were the workshops organized and staffed as planned? Were needed materials available? Were the workshops of high quality (accuracy of information, depth of coverage etc.)?	project records, correspondence, comparing grant proposal and agenda of workshops	project director, document review, other staff interviews,
3. Was the full range of topics included in the design actually covered? Was there evidence of an increase in knowledge as a result of project participation?	project director and staff, participants, observers,	participant questionnaire, observer notes, observer focus group, participant focus group, work samples
4. Did participants exchange information about their use of new instructional approaches? By e-mail or in other ways?	participants, analysis of messages of listserv	participant focus group
5. Did problems arise? Are workshops too few, too many? Should workshop format, content, staffing be modified? Is communication adequate? Was summer session useful?	project director, staff, observers, participants	interview with project director and staff, focus group interview with observers, focus group with participants

Exhibit 16.		
Evaluation questions, data sources, and data collection methods for summative evaluation		
Question	Source of information	Data collection methods
1. Did faculty who have experienced the professional development change their instructional practices? Did this vary by teachers' or by students' characteristics? Do they use the information regarding new standards, materials, and practices? What obstacles prevented implementing changes? What factors facilitated change?	participants, classroom observers, department chair	focus group with participants, reports of classroom observers, interview with department chair
2. Did participants share knowledge acquired through the project with other faculty? Was it done formally (e.g., at faculty meetings) or informally?	participants, other faculty, classroom observers, department chair	focus groups with participants, interviews with nonparticipants, reports of classroom observers (nonparticipants' classrooms), interview with department chair
3. Were changes made in curriculum? Examinations and other requirements? Expenditures for library and other resource materials (computers)?	participants, department chair, dean, budgets and other documents	focus groups with participants, interview with department chair and dean, document review
4. Did students become more interested in classwork? More active participants? Did they express interest in teaching math after graduation? Did they plan to use new concepts and techniques?	students, participants	self-administered questionnaire to be completed by students, focus group with participants

Exhibit 17. First data collection plan		
Method	Sampling plan	Timing of activity
Formative evaluation		
Interview with project director; record review	Not applicable	Once a month during year/during month 1; update if necessary
Interview with other staff	No sampling proposed	At the end of months 3, 6, 10
Workshop observations	No sampling proposed	Two observers at each workshop and summer session
Participant questionnaire	No sampling proposed	Brief questionnaire to be completed at the end of every workshop
Focus group for participants	No sampling proposed	The year 1 participants (n=25) will be assigned to one of three focus groups that meet during month 5 of the school year and after summer session.
Focus group for observers	No sampling proposed	One meeting for all workshop observers during month 11
Summative evaluation		
Classroom observations	Purposive selection: 1 participant per campus; 2 classrooms for each participant; 1 classroom for 2 nonparticipants in each branch campus	Two observations for participants each year (classroom months 4 and 8); one observation for nonparticipants; for 2-year project; a total of 96 observations (two observers at all times)
Focus group with participants	No sampling proposed	The year 2 participants (n=25) will be assigned to one of three focus groups that meet during month 5 of school year and after summer session.
Focus group with classroom observers	No sampling proposed	One focus group with all classroom observers (4-8)
Interview with 2 (nonparticipant) faculty members at all institutions	Random select if more than 2 faculty members in a department	One interview during year 2
Interview with department chairperson at all campuses	Not applicable	Personal interview during year 2
Interview with dean at 8 campuses	Not applicable	During year 2
Interview with all year 1 participants	Not applicable	Towards the end of year 2
Student questionnaires	25% sample of students in all participants' and nonparticipants' classes	Questionnaires to be completed during year 1 and 2
Interview with project director and staff	No sampling proposed	One interview towards end of year 2
Record review	No sampling proposed	During year 1 and year 2

The **summative** evaluation will use relevant data from the formative evaluation; in addition, the following data will be collected:

- During years 1 and 2, teams of two classroom observers will visit a sample of participants' and nonparticipants' classrooms. There will be focus group meetings with these observers at the end of school years 1 and 2 (four to eight staff members are likely to be involved in conducting the 48 scheduled observations each year).
- During year 2 and after the year 2 summer session, focus group meetings will be held with the 25 year 2 participants.
- At the end of year 2, all year 1 participants will be interviewed.
- Interviews will be conducted with nonparticipant faculty members, department chairs and deans at each campus, the project director, and project staff.
- Student surveys will be conducted.

Step 3. Reality Testing and Design Modifications: Staff Needs, Costs, Time Frame Within Which All Tasks (Data Collection, Data Analysis, and Report Writing) Must Be Completed

The evaluation specialist converted the data collection plan (Exhibit 17) into a timeline, showing for each month of the 2 1/2-year life of the project data collection, data analysis, and report-writing activities. Staff requirements and costs for these activities were also computed. She also contacted the chairperson of the department of elementary education at each campus to obtain clearance for the planned classroom observations and data collection from students (undergraduates) during years 1 and 2. This exercise showed a need to fine tune data collection during year 2 so that data analysis could begin by month 18; it also suggested that the scheduled data collection activities and associated data reduction and analysis costs would exceed the evaluation budget by \$10,000. Conversations with campus administrators had raised questions about the feasibility of on-campus data collection from students. The administrators also questioned the need for the large number of scheduled classroom observations. The evaluation staff felt that these observations were an essential component of the evaluation, but they decided to survey students only once (at the end of year 2). They plan to incorporate question about impact on students in the focus group discussions with

participating faculty members after the summer session at the end of year 1. Exhibit 18 shows the final data collection plan for this hypothetical project. It also illustrates how quantitative and qualitative data have been mixed.

Exhibit 18. Final data collection plan			
Activity	Type of method*	Scheduled collection date	Number of cases
1. Interview with project director	Q ₂	Once a month during year 1; twice during year 2 (months 18 and 23)	1
2. Interview with project staff	Q ₂	At the end of months 3, 6, 10 (year 1); at the end of month 23 (year 2)	4 interviews with 4 persons = 16 interviews
3. Record review	Q ₂	Month 1 plus updates as needed	Not applicable
4. Workshop observations	Q ₂	Each workshop including summer	2 observers, 11 observations = 22 observations
5. Participants' evaluation of each workshop	Q ₁	At the end of each workshop and summer	25 participants in 11 workshops = 275 questionnaires
6. Participants' focus groups	Q ₂	Months 5,10,17,22	12 focus groups for 7-8 participants
7. Workshop observer focus groups	Q ₂	Month 10	1 meeting for 2-4 observers
8. Classroom observations	Q ₂	Months 4, 8, 16, 20	2 observers 4 times in 8 classrooms = 64 observations
9. Classroom observations (nonparticipant classrooms)	Q ₂	Months 8 and 16	2 observers twice in 8 classrooms= 32 observations
10. Classroom observers focus group	Q ₂	Months 10 and 22	2 meetings with all classroom observers (4-8)
11. Interviews with department chairs at 8 branch campuses	Q ₂	Months 9 and 21	16 interviews
12. Interviews with all year 1 participants	Q ₂	Month 21	25 interviews
13. Interviews with deans at 7 branch campuses	Q ₂	Month 21	7 interviews
14. Interviews with 2 nonparticipant faculty members at each campus	Q ₂	Month 21	16 interviews
15. Student survey	Q ₁	Month 20	600 self-administered questionnaires
16. Document review	Q ₂	Months 3 and 22	Not applicable

*Q₁ = quantitative; Q₂ = qualitative.

It should be noted that due chiefly to budgetary constraints, the priorities that the final evaluation plan did not provide for the systematic collection of some information that might have been of importance for the overall assessment of the project and recommendations for replication. For example, there is no provision to examine systematically (by using trained workshop observers, as is done during year 1) the extent to which the year 2 workshops were modified as a result of the formative evaluation. This does not mean, however, that an evaluation question that did not survive the prioritization process cannot be explored in conjunction with the data collection tools specified in Exhibit 17. Thus, the question of workshop modifications and their effectiveness can be explored in the interviews scheduled with project staff and the self-administered questionnaires and focus groups for year 2 participants. Furthermore, informal interaction between the evaluation staff, the project staff, participants, and others involved in the project can yield valuable information to enrich the evaluation.

Experienced evaluators know that, in hindsight, the prioritization process is often imperfect. And during the life of any project, it is likely that unanticipated events will affect project outcomes. Given the flexible nature of qualitative data collection tools, to some extent the need for additional information can be accommodated in mixed method designs by including narrative and anecdotal material. Some of the ways in which such material can be incorporated in reaching conclusions and recommendations will be discussed in Chapter 7 of this handbook.

Chapter 7

REPORTING THE RESULTS OF MIXED METHOD EVALUATIONS

Preparing a report provides evaluators the opportunity to depict in detail the rich information obtained from the various study activities.

The final task the evaluator is required to perform is to summarize what the team has done, what has been learned, and how others might benefit from this project's experience. As a rule, NSF grantees are expected to submit a final report when the evaluation has been completed. For the evaluator, this is seen as the primary reporting task, which provides the opportunity to depict in detail the rich qualitative and quantitative information obtained from the various study activities.

In addition to the contracting agency, most evaluations have other audiences as well, such as previously identified stakeholders, other policymakers, and researchers. For these audiences, whose interest may be limited to a few of the topics covered in the full report, shorter summaries, oral briefings, conference presentations, or workshops may be more appropriate. Oral briefings allow the sharing of key findings and recommendations with those decisionmakers who lack the time to carefully review a voluminous report. In addition, conference presentations and workshops can be used to focus on special themes or to tailor messages to the interests and background of a specific audience.

In preparing the final report and other products that communicate the results of the evaluation, the evaluator must consider the following questions:

- How should the communication be best tailored to meet the needs and interests of a given audience?
- How should the comprehensive final report be organized? How should the findings based on qualitative and quantitative methods be integrated?
- Does the report distinguish between conclusions based on robust data and those that are more speculative?

- Where findings are reported, especially those likely to be considered sensitive, have appropriate steps been taken to make sure that promises of confidentiality are met?

This chapter deals primarily with these questions. More extensive coverage of the general topic of reporting and communicating evaluation results can be found in the earlier *User-Friendly Handbook for Project Evaluation* (NSF, 1993).

Ascertaining the Interests and Needs of the Audience

The diversity of audiences for which the findings are likely to be of interest is illustrated for the hypothetical project in Exhibit 19. As shown, in addition to NSF, the immediate audience for the evaluation might include top-level administrators at the major state university, staff at the Center for Educational Innovation, and the undergraduate faculty who are targeted to participate in these or similar workshops. Two other indirect audiences might be policymakers at other 4-year institutions interested in developing similar preservice programs and other researchers. Each of these potential audiences might be interested in different aspects of the evaluation's findings. Not all data collected in a mixed method evaluation will be relevant to their interests. For example:

- The National Science Foundation staff interested in replication might want rich narrative detail in order to help other universities implement similar preservice programs. For this audience, the model would be a descriptive report that traces the flow of events over time, recounts how the preservice program was planned and implemented, identifies factors that facilitated or impeded the project's overall success, and recommends possible modifications.
- Top-level administrators at the university might be most interested in knowing whether the preservice program had its intended effect, i.e., to inform future decisions about funding levels and to optimize the allocation of scarce educational resources. For this audience, data from the summative evaluation are most pertinent.
- Staff at the Center for Educational Innovation might be interested in knowing which activities were most successful in improving the overall quality of their projects. In addition, the Center would likely want to use any positive findings to generate ongoing support for their program.

Exhibit 19. Matrix of stakeholders						
Potential audience for the study findings	Levels of audience involvement with the program	Intended impacts of the study				
		Assess success of the program	Facilitate decision-making	Generate support for the program	Revise current theories about preservice education	Inform best practices for preservice education programs
National Science Foundation	Direct	X	X		X	X
Top-level administrators at the major state university	Direct	X	X		X	
Staff at the Center for Educational Innovation	Direct	X	X	X	X	X
Undergraduate faculty targeted to participate in the workshops	Direct	X			X	
Policymakers at other 4-year institutions interested in developing similar preservice programs	Indirect	X	X			X
Other researchers	Indirect	X			X	X

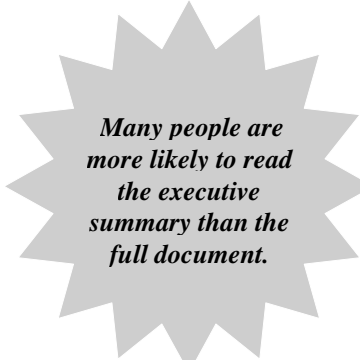
In this example, the evaluator would risk having the results ignored by some stakeholders and underutilized by others if only a single dissemination strategy was used. Even if a single report is developed for all stakeholders (which is usually the case), it is advisable to develop a dissemination strategy that recognizes the diverse informational needs of the audience and the limited time some readers might realistically be able to devote to digesting the results of the study. Such a strategy might include (1) preparing a concise executive summary of the evaluation's key findings (for the university's top-level administrators); (2) preparing a detailed report (for the Center for Educational Innovation and the National Science Foundation) that describes the history of the program, the range of activities offered to undergraduate faculty, and the impact of these activities on program participants and their students; and (3) conducting a series of briefings that are tailored to the interests of

specific stakeholders (e.g., university administrators might be briefed on the program's tangible benefits and costs). By referring back to the worksheets that were developed in planning the evaluation (see Chapter 5), the interests of specific stakeholders can be ascertained. However, rigid adherence to the original interests expressed by stakeholders is not always the best approach. This strategy may shortchange the audience if the evaluation—as is often the case—pointed to unanticipated developments. It should also be pointed out that while the final report usually is based largely on answers to summative evaluation questions, it is useful to summarize salient results of the formative evaluation as well, where these results provide important information for project replication.

Organizing and Consolidating the Final Report

Usually, the organization of mixed method reports follows the standard format, described in detail in the earlier NSF user-friendly handbook, that consists of five major sections:


- Background (the project's objectives and activities);
- Evaluation questions (meeting stakeholders' information needs);
- Methodology (data collection and analysis);
- Findings; and
- Conclusions (and recommendations).



Many people are more likely to read the executive summary than the full document.

In addition to the main body of the report, a short abstract and a one-to four-page executive summary should be prepared. The latter is especially important because many people are more likely to read the executive summary than the full document. The executive summary can help focus readers on the most significant aspects of the evaluation. It is desirable to keep the methodology section short and to include a technical appendix containing detailed information about the data collection and other methodological issues. All evaluation instruments and procedures should be contained in the appendix, where they are accessible to interested readers.

Regardless of the audience for which it is written, the final report must engage the reader and stimulate attention and interest. Descriptive narrative, anecdotes, personalized observations, and vignettes make for livelier reading than a long recitation of statistical



The purpose of the final report is not only to tell the story of the project, but also to assess in what ways it succeeded or failed in achieving its goals.

measures and indicators. One of the major virtues of the mixed method approach is the evaluator's ability to balance narrative and numerical reporting. This can be done in many ways: for example, by alternating descriptive material (obtained through qualitative techniques) and numerical material (obtained through quantitative techniques) when describing project activities, or by using qualitative information to illustrate, personalize, or explicate a statistical finding.

But—as discussed in the earlier chapters—the main virtue of using a mixed method approach is that it enlarges the scope of the analysis. And it is important to remember that the purpose of the final report is not only to tell the story of the project, its participants, and its activities, but also to assess in what ways it succeeded or failed in achieving its goals.

In preparing the findings section, which constitutes the heart of the report, it is important to balance and integrate the descriptive and evaluative reporting section. A well-written report should provide a concise context for understanding the conditions in which results were obtained and identifying specific factors (e.g., implementation strategies) that affected the results. According to Patton (1990),

Description is thus balanced by analysis and interpretation. Endless description becomes its own muddle. The purpose of analysis is to organize description so that it is manageable. Description is balanced by analysis and leads into interpretation. An interesting and reasonable report provides sufficient description to allow the reader to understand the basis for an interpretation, and sufficient interpretation to allow the reader to understand the description.

For the hypothetical project, most questions identified for the summative evaluation in Exhibit 16 can be explored through the joint use of qualitative and quantitative data, as shown in Exhibit 20. For example, to answer some of the questions pertaining to the impact of faculty training on their students' attitudes and behaviors, quantitative data (obtained from a student survey) are being used, together with qualitative information obtained through several techniques (classroom observations, faculty focus groups, interviews with knowledgeable informants.)

Exhibit 20. Example of an evaluation/methodology/matrix							
Project goals	Summative evaluation study questions	Data collection techniques (see codes below)					
		a	b	c	d	e	f
Changes in instructional practices by participating faculty members	Did the faculty who experienced the workshop training change their instructional practice?	X	X		X	X	
	Did the faculty who experienced the workshop training use the information regarding new standards, materials, and practices?	X	X		X	X	
	What practices prevented the faculty who experienced the workshop training from implementing the changes?	X	X				
Acquisition of knowledge and changes in instructional practices by other (nonparticipating) faculty members	Did participants share the knowledge acquired through the workshops with other faculty?	X	X			X	
	What methods did participants use to share the knowledge acquired through the workshops?	X					
Institution-wide changes in curriculum and administrative practices	Were changes made in curriculum?	X					X
	Were changes made in examinations and other requirements?	X					X
	Were changes made in expenditures for libraries and other resource materials?	X					X
Positive effects on career plans of students taught by participating teachers	Did students become more interested in their classwork?	X	X			X	
	Did students become more active participants?	X	X		X	X	
	Did students express interest in teaching math after graduation?					X	
	Did students plan to use new concepts and techniques?					X	
a = indepth interviews with knowledgeable informants d = classroom observations b = focus groups e = surveys of students c = observation of workshops f = documents							

Formulating Sound Conclusions and Recommendations

In the great majority of reporting activities, the evaluator will seek to include a conclusion and recommendations section in which the findings are summarized, broader judgments are made about the strengths and weaknesses of the project and its various features, and recommendations for future, perhaps improved, replications are presented. Like the executive summary, this section is widely read and may affect policymakers' and administrators' decisions with respect to future project support.

The report writer can include in this section only a limited amount of material, and should therefore select the most salient findings. But how should saliency be defined? Should the "strongest" findings be emphasized, i.e., those satisfying accepted criteria for soundness in the quantitative and qualitative traditions? Or should the writer present more sweeping conclusions, ones which may be based in part on impressionistic and anecdotal material?

It can be seen that the evaluator often faces a dilemma. On the one hand, it is extremely important that the bulk of evaluative statements made in this section can be supported by accurate and robust data and systematic analysis. As discussed in Chapter 1, some stakeholders may seek to discredit evaluations if they are not in accord with their expectations or preferences; such critics may question the conclusions and recommendations offered by the evaluator that seem to leap beyond the documented evidence. On the other hand, it is often beneficial to capture insights that result from immersion in a project, insights not provided by sticking only to results obtained through scientific documentation. The evaluator may have developed a strong, intuitive sense of how the project really worked out, what were its best or worst features, and what benefits accrued to participants or to institutions impacted by the project (for example, schools or school systems). Thus, there may be a need to stretch the data beyond their inherent limits, or to make statements for which the only supporting data are anecdotal.

We have several suggestions for dealing with these issues:

- Distinguish carefully between conclusions that are based on "hard" data and those that are more speculative. The best strategy is to start the conclusions section with material that has undergone thorough verification and to place the more subjective speculations toward the end of the section.
- Provide full documentation for all findings where available. Data collection instruments, descriptions of the study subjects, specific procedures followed for data collection, survey

response rates, refusal rates for personal interviews and focus group participation, access problems, etc., should all be discussed in an appendix. If problems were encountered that may have affected the findings, possible biases and how the evaluator sought to correct them should be discussed.

- Use the recommendations section to express views based on the total project experience. Of course, references to data should be included whenever possible. For example, a recommendation in the report for the hypothetical project might include the following phrase: "Future programs should provide career-related incentives for faculty participation, as was suggested by several participants." But the evaluator should also feel free to offer creative suggestions that do not necessarily rely on the systematic data collection.

Maintaining Confidentiality

All research involving human subjects entails possible risks for participants and usually requires informed consent on their part. To obtain this consent, researchers usually assure participants that their identity will not be revealed when the research is reported and that all information obtained through surveys, focus groups, personal interviews, and observations will be handled confidentially. Participants are assured that the purpose of the study is not to make judgments about their performance or behavior, but simply to improve knowledge about a project's effectiveness and improve future activities.

In quantitative studies, reporting procedures have been developed to minimize the risk that the actions and responses of participants can be associated with a specific individual; usually results are reported for groupings only and, as a rule, only for groupings that include a minimum number of subjects.

In studies that use qualitative methods, it may be more difficult to report all findings in ways that make it impossible to identify a participant. The number of respondents is often quite small, especially if one is looking at respondents with characteristics that are of special interest in the analysis (for example, older teachers, or teachers who hold a graduate degree). Thus, even if a finding does not name the respondent, it may be possible for someone (a colleague, an administrator) to identify a respondent who made a critical or disparaging comment in an interview.

Of course, not all persons who are interviewed in the course of an evaluation can be anonymous: the name of those persons who have a unique or high status role (the project director, a college dean, or a school superintendent) are known, and anonymity should not be promised. The issue is of importance to more vulnerable persons, usually those in subordinate positions (teachers, counselors, or students) who may experience negative consequences if their behavior and opinions become known.

It is in the interest of the evaluator to obtain informed consent from participants by assuring them that their participation is risk-free; they will be more willing to participate and will speak more openly. But in the opinion of experienced qualitative researchers, it is often impossible to fulfill promises of anonymity when qualitative methods are used:

Confidentiality and anonymity are usually promised—sometimes very superficially—in initial agreements with respondents. For example, unless the researcher explains very clearly what a feedback case will look like, people may not realize that they will not be anonymous at all to other people within the setting who read the case (Miles and Huberman, 1994).

The evaluator may also find it difficult to balance the need to convey contextual information that will provide vivid descriptive information and the need to protect the identity of informants. But if participants have been promised anonymity, it behooves the evaluator to take every precaution so that informants cannot be linked to any of the information they provided.

In practice, the decision of how and when to attribute findings to a site or respondent is generally made on a case-by-case basis. The following example provides a range of options for revealing and disclosing the source of information received during an interview conducted for the hypothetical project:

- **Attribute the information to a specific respondent within an individual site:** “The dean at Lakewood College indicated that there was no need for curriculum changes at this time.”
- **Attribute the information to someone within a site:** “A respondent at Lakewood College indicated that there was no need for curriculum changes at this time.” In this example, the respondent's identity within the site is protected, i.e., the reader is only made aware that someone at a site expressed a preference for the status quo. Note that this option would not

be used if only one respondent at the school was in a position to make this statement.

- **Attribute the information to the respondent type without identifying the site:** “The dean at one of the participating colleges indicated that there was no need for curriculum changes at this time.” In this example, the reader is only made aware of the type of respondent that expressed a preference for the status quo.
- **Do not attribute the information to a specific respondent type or site:** “One of the study respondents indicated that there was no need for curriculum changes at this time.” In this example, the identity of the respondent is fully protected.

Each of these alternatives has consequences not only for protecting respondent anonymity, but also for the value of the information that is being conveyed. The first formulation discloses the identity of the respondent and should only be used if anonymity was not promised initially, or if the respondent agrees to be identified. The last alternative, while offering the best guarantee of anonymity, is so general that it weakens the impact of the finding. Depending on the direction taken by the analysis (were there important differences by site? by type of respondent?), it appears that either the second or third alternative 2 or 3 represents the best choice.

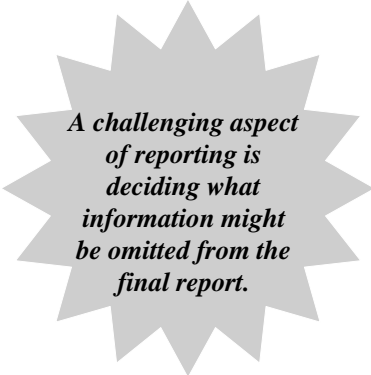
One common practice is to summarize key findings in chapters that provide cross-site analyses of controversial issues. This alternative is “directly parallel to the procedure used in surveys, in which the only published report is about the aggregate evidence” (Yin, 1990). Contextual information about individual sites can be provided separately, e.g., in other chapters or an appendix.

Tips for Writing Good Evaluation Reports

Start early. Although we usually think about report writing as the final step in the evaluation, a good deal of the work can (and often does) take place before the data are collected. For example, a background section can often be developed using material from the original proposal. While some aspects of the methodology may deviate from the original proposal as the study progresses, most of the background information (e.g., nature of the problem, project goals) will remain the same throughout the evaluation. In addition, the evaluation study questions section can often be written using material that was developed for the evaluation design. The

evaluation findings, conclusions, and recommendations generally need to wait for the end of the evaluation.

Because of the volume of written data that are collected on site, it is generally a good idea to organize study notes as soon after a site visit or interview as possible. These notes will often serve as a starting point for any individual case studies that might be included in the report. In addition, as emphasized in Chapter 4, preparing written text soon after the data collection activity will help to classify and display the data and reduce the overall volume of narrative data that will eventually need to be summarized and reported at the end of the study. Finally, preparing sections of the findings chapter during the data collection phase allows researchers to generate preliminary conclusions or identify potential trends that can be confirmed or refuted by additional data collection activities.



A challenging aspect of reporting is deciding what information might be omitted from the final report.

Make the report concise and readable. Because of the volume of material that is generally collected during mixed method evaluations, a challenging aspect of reporting is deciding what information might be omitted from the final report. As a rule, only a fraction of the tabulations prepared for survey analysis need to be displayed and discussed. Qualitative field work and data collection methods yield a large volume of narrative information, and evaluators who try to incorporate all of the qualitative data they collected into their report risk losing their audience. Conversely, by omitting too much, evaluators risk removing the context that helps readers attach meaning to any of the report's conclusions. One method for limiting the volume of information is to include only narrative that is tied to the evaluation questions. Regardless of how interesting an anecdote is, if the information does not relate to one of the evaluation questions, it probably does not belong in the report. As discussed previously, another method is to consider the likely information needs of your audience. Thinking about who is most likely to act upon the report's findings may help in the preparation of a useful and illuminating narrative (and in the discarding of anecdotes that are irrelevant to the needs of the reader).



New desktop publishing and software packages have made it easier to enhance papers and briefings with photographs, colorful graphics, and even cartoons.

The liberal use of qualitative information will enhance the overall tone of the report. In particular, lively quotes can highlight key points and break up the tedium of a technical summation of study findings. In addition, graphic displays and tables can be used to summarize significant trends that were uncovered during observations or interviews. Photographs are an effective tool to familiarize readers with the conditions (e.g., classroom size) within which a project is being implemented. New desktop publishing and software packages have made it easier to enhance papers and briefings with photographs, colorful graphics, and even cartoons. Quotes can be enlarged and italicized throughout the report to make

important points or to personalize study findings. Many of these suggestions hold true for oral presentations as well.

Solicit feedback from project staff and respondents. It is often useful to ask the project director and other staff members to review sections of the report that quote information they have contributed in interviews, focus groups, or informal conversations. This review is useful for correcting omissions and misinterpretations and may elicit new details or insights that staff members failed to share during the data collection period. The early review may also avoid angry denials after the report becomes public, although it is no guarantee that controversy and demands for changes will not follow publication. However, the objectivity of the evaluation is best served if overall findings, conclusions and recommendations are not shared with the project staff before the draft is circulated to all stakeholders.

In general, the same approach is suggested for obtaining feedback from respondents. It is essential to inform them of the inclusion of data with which they can be identified, and to honor requests for anonymity. The extent to which other portions of the write-up should be shared with respondents will depend on the nature of the project and the respondent population, but in general it is probably best to solicit feedback following dissemination of the report to all stakeholders.

References

- Miles, M.B., and Huberman, A.M. (1994). *Qualitative Data Analysis*, 2nd Ed. Newbury Park, CA: Sage.
- National Science Foundation. (1993). *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering, and Technical Education*. NSF 93-152. Washington, DC: NSF.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Method*, 2nd Ed. Newbury Park, CA: Sage.
- Yin, R.K. (1989). *Case Study Research: Design and Method*. Newbury Park, CA: Sage.

PART IV.

SUPPLEMENTARY

MATERIALS

Chapter 8 ANNOTATED BIBLIOGRAPHY

In selecting books and major articles for inclusion in this short bibliography, an effort was made to incorporate those useful for principal investigators (PIs) and project directors (PDs) who want to find information relevant to the tasks they will face, and which this brief handbook could not cover in depth. Thus, we have not included all books that experts in qualitative research and mixed method evaluations would consider to be of major importance. Instead, we have included primarily reference materials that NSF/EHR grantees should find most useful. Included are many of those already listed in the references to Chapters 1 through 7.

Some of these publications are heavier on theory, others deal primarily with practice and specific techniques used in qualitative data collection and analysis. However, with few exceptions, all the publications selected for this bibliography contain a great deal of technical information and hands-on advice.

Denzin, Norman K., and Lincoln, Yvonna S. (Eds.). (1994). *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.

This formidable volume (643 pages set in small type) consists of 36 chapters written by experts on their respective topics, all of whom are passionate advocates of the qualitative method in social and educational research. The volume covers historical and philosophical perspectives, as well as detailed research methods. Extensive coverage is given to data collection and data analysis, and to the “art of interpretation” of findings obtained through qualitative research. Most of the chapters assume that the qualitative researcher functions in an academic setting and uses qualitative methods exclusively; the use of quantitative methods in conjunction with qualitative approaches and constraints that apply to evaluation research are seldom considered. However, two chapters—“Designing Funded Qualitative Research,” by Janice M. Morse, and “Qualitative Program Evaluation,” by Jennifer C. Greene—contain a great deal of material of interest to PIs and PDs. But PIs and PDs will also benefit from consulting other chapters, in particular “Interviewing,” by Andrea Fontana and James H. Frey, and “Data Management and Analysis Methods,” by A. Michael Huberman and Matthew B. Miles.

The Joint Committee on Standards for Educational Evaluation. (1994). *How to Assess Evaluations of Educational Programs*, 2nd Ed. Thousand Oaks, CA: Sage.

This new edition of the widely accepted Standards for Educational Evaluation is endorsed by professional associations in the field of education. The volume defines 30 standards for program evaluation, with examples of their application, and incorporates standards for quantitative as well as qualitative evaluation methods. The Standards are categorized into four groups: utility, feasibility, propriety, and accuracy. The Standards are intended to assist legislators, funding agencies, educational administrators, and evaluators. They are not a substitute for texts in technical areas such as research design or data collection and analysis. Instead they provide a framework and guidelines for the practice of responsible and high-quality evaluations. For readers of this handbook, the section on Accuracy Standards, which includes discussions of quantitative and qualitative analysis, justified conclusions, and impartial reporting, is especially useful.

Patton, Michael Quinn. (1990). *Qualitative Evaluation and Research Methods*, 2nd Ed. Newbury Park, CA: Sage.

This is a well-written book with many practical suggestions, examples, and illustrations. The first part covers, in jargon-free language, the conceptual and theoretical issues in the use of qualitative methods; for practitioners the second and third parts, dealing with design, data collection, analysis, and interpretation, are especially useful. Patton consistently emphasizes a pragmatic approach: he stresses the need for flexibility, common sense, and the choice of methods best suited to produce the needed information. The last two chapters, "Analysis, Interpretation and Reporting" and "Enhancing the Quality and Credibility of Qualitative Analysis," are especially useful for PIs and PDs of federally funded research. They stress the need for utilization-focused evaluation and the evaluator's responsibility for providing data and interpretations, which specific audiences will find credible and persuasive.

Marshall, Catherine, and Rossman, Gretchen B. (1995). *Designing Qualitative Research*, 2nd Ed. Thousand Oaks, CA: Sage.

This small book (178 pages) does not deal specifically with the performance of evaluations; it is primarily written for graduate students to provide a practical guide for the writing of research proposals based on qualitative methods. However, most of the material presented is relevant and appropriate for project evaluation. In succinct and clear language, the book discusses the main ingredients of a sound research

project: framing evaluation questions; designing the research; data collection methods; and strategies, data management, and analysis. The chapter on data collection methods is comprehensive and includes some of the less widely used techniques (such as films and videos, unobtrusive measures, and projective techniques) that may be of interest for the evaluation of some projects. There are also useful tables (e.g., identifying the strengths and weaknesses of various methods for specific purposes; managing time and resources), as well as a series of vignettes throughout the text illustrating specific strategies used by qualitative researchers.

Lofland, John, and Lofland, Lyn H. (1995). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd Ed. Belmont, CA: Wadsworth.

As the title indicates, this book is designed as a guide to field studies, using as their main data collection techniques participant observation and intensive interviews. The authors' vast experience and knowledge in these areas results in a thoughtful presentation of both technical topics (such as the best approach to compiling field notes) and nontechnical issues, which may be equally important in the conduct of qualitative research. The chapters that discuss gaining access to informants, maintaining access for the duration of the study, and dealing with issues of confidentiality and ethical concerns are especially helpful for PIs and PDs who seek to collect qualitative material. Also useful is Chapter 5, "Logging Data," which deals with all aspects of the interviewing process and includes examples of question formulation, the use of interview guides, and the write-up of data.

Miles, Matthew B., and Huberman, A. Michael. (1994). *Qualitative Data Analysis - An Expanded Sourcebook*, 2nd Ed. Thousand Oaks, CA: Sage.

Although this book is not specifically oriented to evaluation research, it is an excellent tool for evaluators because, in the authors' words, "this is a book for practicing researchers in all fields whose work involves the struggle with actual qualitative data analysis issues." It has the further advantage that many examples are drawn from the field of education. Because analysis cannot be separated from research design issues, the book takes the reader through the sequence of steps that lay the groundwork for sound analysis, including a detailed discussion of focusing and bounding the collection of data, as well as management issues bearing on analysis. The subsequent discussion of analysis methods is very systematic, relying heavily on data displays, matrices, and examples to arrive at meaningful descriptions, explanations, and

the drawing and verifying of conclusions. An appendix covers choice of software for qualitative data analysis. Readers will find this a very comprehensive and useful resource for the performance of qualitative data reduction and analysis.

New Directions for Program Evaluation, Vols. 35, 60, 61. A quarterly publication of the American Evaluation Association, published by Jossey-Bass, Inc., San Francisco, CA.

Almost every issue of this journal contains material of interest to those who want to learn about evaluation, but the three issues described here are especially relevant to the use of qualitative methods in evaluation research. Vol. 35 (Fall 1987), *Multiple Methods in Program Evaluation*, edited by Melvin M. Mark and R. Lance Shotland, contains several articles discussing the combined use of quantitative and qualitative methods in evaluation designs. Vol. 60 (Winter 1993), *Program Evaluation: A Pluralistic Enterprise*, edited by Lee Sechrest, includes the article "Critical Multiplism: A Research Strategy and its Attendant Tactics," by William R. Shadish, in which the author provides a clear discussion of the advantages of combining several methods in reaching valid findings. In Vol. 61 (Spring 1994), *The Qualitative-Quantitative Debate*, edited by Charles S. Reichardt and Sharon F. Rallis, several of the contributors take a historical perspective in discussing the long-standing antagonism between qualitative and quantitative researchers in evaluation. Others look for ways of integrating the two perspectives. The contributions by several experienced nonacademic program and project evaluators (Rossi, Datta, Yin) are especially interesting.

Greene, Jennifer C., Caracelli, Valerie J., and Graham, Wendy F. (1989). "Toward a Conceptual Framework for Mixed-Method Evaluation Designs" in *Educational Evaluation and Policy Analysis*, Vol. II, No. 3.

In this article, a framework for the design and implementation of evaluations using a mixed method methodology is presented, based both on the theoretical literature and a review of 57 mixed method evaluations. The authors have identified five purposes for using mixed methods, and the recommended design characteristics for each of these purposes are presented.

Yin, Robert K. (1989). *Case Study Research: Design and Method*. Newbury Park, CA: Sage.

The author's background in experimental psychology may explain the emphasis in this book on the use of rigorous methods in the conduct

and analysis of case studies, thus minimizing what many believe is a spurious distinction between quantitative and qualitative studies. While arguing eloquently that case studies are an important tool when an investigator (or evaluator) has little control over events and when the focus is on a contemporary phenomenon within some real-life context, the author insists that case studies be designed and analyzed so as to provide generalizable findings. Although the focus is on design and analysis, data collection and report writing are also covered.

Krueger, Richard A. (1988). *Focus Groups: A Practical Guide for Applied Research*. Newbury Park, CA: Sage.

Krueger is well known as an expert on focus groups; the bulk of his experience and the examples cited in his book are derived from market research. This is a useful book for the inexperienced evaluator who needs step-by-step advice on selecting focus group participants, the process of conducting focus groups, and analyzing and reporting results. The author writes clearly and avoids social science jargon, while discussing the complex problems that focus group leaders need to be aware of. This book is best used in conjunction with some of the other references cited here, such as the *Handbook of Qualitative Research* (Ch. 22) and *Focus Groups: Theory and Practice*.

Stewart, David W., and Shamdasani, Prem N. (1990). *Focus Groups: Theory and Practice*. Newbury Park, CA: Sage.

This book differs from many others published in recent years that address primarily techniques of recruiting participants and the actual conduct of focus group sessions. Instead, these authors pay considerable attention to the fact that focus groups are by definition an exercise in group dynamics. This must be taken into account when interpreting the results and attempting to draw conclusions that might be applicable to a larger population. However, the book also covers very adequately practical issues such as recruitment of participants, the role of the moderator, and appropriate techniques for data analysis.

Weiss, Robert S. (1994). *Learning from Strangers - The Art and Method of Qualitative Interview Studies*. New York: The Free Press.

After explaining the different functions of quantitative and qualitative interviews in the conduct of social science research studies, the author discusses in considerable detail the various steps of the qualitative interview process. Based largely on his own extensive experience in planning and carrying out studies based on qualitative interviews, he discusses respondent selection and recruitment, preparing for the

interview (which includes such topics as pros and cons of taping, the use of interview guides, interview length, etc.), the interviewing relationship, issues in interviewing (including confidentiality and validity of the information provided by respondents), data analysis, and report writing. There are lengthy excerpts from actual interviews that illustrate the topics under discussion. This is a clearly written, very useful guide, especially for newcomers to this data collection method.

Wolcott, Harry F. (1994). *Transforming Qualitative Data: Description, Analysis and Interpretation*. Thousand Oaks, CA: Sage.

This book is written by an anthropologist who has done fieldwork for studies focused on education issues in a variety of cultural settings; his emphasis throughout is “on what one does with data rather than on collecting it.” His frank and meticulous description of the ways in which he assembled his data, interacted with informants, and reached new insights based on the gradual accumulation of field experiences makes interesting reading. It also points to the pitfalls in the interpretation of qualitative data, which he sees as the most difficult task for the qualitative researcher.

U.S. General Accounting Office. (1990). *Case Study Evaluations*. Transfer Paper 10.1.9. issued by the Program Evaluation and Methodology Division. Washington, DC: GAO.

This paper presents an evaluation perspective on case studies, defines them, and determines their appropriateness in terms of the type of evaluation question posed. Unlike the traditional, academic definition of the case study, which calls for long-term participation by the evaluator or researcher in the site to be studied, the GAO sees a wide range of shorter term applications for case study methods in evaluation. These include their use in conjunction with other methods for illustrative and exploratory purposes, as well as for the assessment of program implementation and program effects. Appendix 1 includes a very useful discussion dealing with the adaptation of the case study method for evaluation and the modifications and compromises that evaluators—unlike researchers who adopt traditional field work methods—are required to make.

Chapter 9 GLOSSARY

Accuracy:	The extent to which an evaluation is truthful or valid in what it says about a program, project, or material.
Achievement:	Performance as determined by some type of assessment or testing.
Affective:	Consists of emotions, feelings, and attitudes.
Anonymity: (provision for)	Evaluator action to ensure that the identity of subjects cannot be ascertained during the course of a study, in study reports, or in any other way.
Assessment:	Often used as a synonym for evaluation. The term is sometimes recommended for restriction to processes that are focused on quantitative and/or testing approaches
Attitude:	A person's mental set toward another person, thing, or state.
Attrition:	Loss of subjects from the defined sample during the course of a longitudinal study.
Audience(s):	Consumers of the evaluation; those who will or should read or hear of the evaluation, either during or at the end of the evaluation process. Includes those persons who will be guided by the evaluation in making decisions and all others who have a stake in the evaluation (see stakeholders).
Authentic assessment:	Alternative to traditional testing, using indicators of student task performance.
Background:	The contextual information that describes the reasons for the project, including its goals, objectives, and stakeholders' information needs.
Baseline:	Facts about the condition or performance of subjects prior to treatment or intervention.

Behavioral objectives:	Specifically stated terms of attainment to be checked by observation, or test/measurement.
Bias:	A consistent alignment with one point of view.
Case study:	An intensive, detailed description and analysis of a single project, program, or instructional material in the context of its environment.
Checklist approach:	Checklists are the principal instrument for practical evaluation, especially for investigating the thoroughness of implementation.
Client:	The person or group or agency that commissioned the evaluation.
Coding:	To translate a given set of data or items into descriptive or analytic categories to be used for data labeling and retrieval.
Cohort:	A term used to designate one group among many in a study. For example, “the first cohort” may be the first group to have participated in a training program.
Component:	A physically or temporally discrete part of a whole. It is any segment that can be combined with others to make a whole.
Conceptual scheme:	A set of concepts that generate hypotheses and simplify description.
Conclusions (of an evaluation):	Final judgments and recommendations.
Content analysis:	A process using a parsimonious classification system to determine the characteristics of a body of material or practices.
Context (of an evaluation):	The combination of factors accompanying the study that may have influenced its results, including geographic location, timing, political and social climate, economic conditions, and other relevant professional activities in progress at the same time.
Criterion, criteria:	A criterion (variable) is whatever is used to measure a successful or unsuccessful outcome, e.g., grade point average.
Criterion-referenced test:	Tests whose scores are interpreted by referral to well-defined domains of content or behaviors, rather than by referral to the performance of some comparable group of people.
Cross-case analysis:	Grouping data from different persons to common questions or analyzing different perspectives on issues under study.

Cross-sectional study:	A cross-section is a random sample of a population, and a cross-sectional study examines this sample at one point in time. Successive cross-sectional studies can be used as a substitute for a longitudinal study. For example, examining today's first year students and today's graduating seniors may enable the evaluator to infer that the college experience has produced or can be expected to accompany the difference between them. The cross-sectional study substitutes today's seniors for a population that cannot be studied until 4 years later.
Data display:	A compact form of organizing the available information (for example, graphs, charts, matrices).
Data reduction:	Process of selecting, focusing, simplifying, abstracting, and transforming data collected in written field notes or transcriptions.
Delivery system:	The link between the product or service and the immediate consumer (the recipient population).
Descriptive data:	Information and findings expressed in words, unlike statistical data, which are expressed in numbers.
Design:	The process of stipulating the investigatory procedures to be followed in doing a specific evaluation.
Dissemination:	The process of communicating information to specific audiences for the purpose of extending knowledge and, in some cases, with a view to modifying policies and practices.
Document:	Any written or recorded material not specifically prepared for the evaluation.
Effectiveness:	Refers to the conclusion of a goal achievement evaluation. "Success" is its rough equivalent.
Elite interviewers:	Well-qualified and especially trained persons who can successfully interact with high-level interviewees and are knowledgeable about the issues included in the evaluation.
Ethnography:	Descriptive anthropology. Ethnographic program evaluation methods often focus on a program's culture.
Executive summary:	A nontechnical summary statement designed to provide a quick overview of the full-length report on which it is based.
External evaluation:	Evaluation conducted by an evaluator from outside the organization within which the object of the study is housed.

Field notes:	Observer's detailed description of what has been observed.
Focus group:	A group selected for its relevance to an evaluation that is engaged by a trained facilitator in a series of discussions designed for sharing insights, ideas, and observations on a topic of concern to the evaluation.
Formative evaluation:	Evaluation designed and used to improve an intervention, especially when it is still being developed.
Hypothesis testing:	The standard model of the classical approach to scientific research in which a hypothesis is formulated before the experiment to test its truth.
Impact evaluation:	An evaluation focused on outcomes or payoff.
Implementation evaluation:	Assessing program delivery (a subset of formative evaluation).
Indepth interview:	A guided conversation between a skilled interviewer and an interviewee that seeks to maximize opportunities for the expression of a respondent's feelings and ideas through the use of open-ended questions and a loosely structured interview guide.
Informed consent:	Agreement by the participants in an evaluation to the use, in specified ways for stated purposes, of their names and/or confidential information they supplied.
Instrument:	An assessment device (test, questionnaire, protocol, etc.) adopted, adapted, or constructed for the purpose of the evaluation.
Internal evaluator:	A staff member or unit from the organization within which the object of the evaluation is housed.
Intervention:	Project feature or innovation subject to evaluation.
Intra-case analysis:	Writing a case study for each person or unit studied.
Key informant:	Person with background, knowledge, or special skills relevant to topics examined by the evaluation.
Longitudinal study:	An investigation or study in which a particular individual or group of individuals is followed over a substantial period of time to discover changes that may be attributable to the influence of the treatment, or to maturation, or the environment. (See also cross-sectional study.)
Matrix:	An arrangement of rows and columns used to display multi-dimensional information.

Measurement:	Determination of the magnitude of a quantity.
Mixed method evaluation:	An evaluation for which the design includes the use of both quantitative and qualitative methods for data collection and data analysis.
Moderator:	Focus group leader; often called a facilitator.
Nonparticipant observer:	A person whose role is clearly defined to project participants and project personnel as an outside observer or onlooker.
Norm-referenced tests:	Tests that measure the <i>relative</i> performance of the individual or group by comparison with the performance of other individuals or groups taking the same test.
Objective:	A specific description of an intended outcome.
Observation:	The process of direct sensory inspection involving trained observers.
Ordered data:	Non-numeric data in ordered categories (for example, students' performance categorized as excellent, good, adequate, and poor).
Outcome:	Post-treatment or post-intervention effects.
Paradigm:	A general conception, model, or "worldview" that may be influential in shaping the development of a discipline or subdiscipline. (For example, "The classical, positivist social science paradigm in evaluation.")
Participant observer:	A person who becomes a member of the project (as participant or staff) in order to gain a fuller understanding of the setting and issues.
Performance evaluation:	A method of assessing what skills students or other project participants have acquired by examining how they accomplish complex tasks or the products they have created (e.g., poetry, artwork).
Planning evaluation:	Evaluation planning is necessary before a program begins, both to get baseline data and to evaluate the program plan, at least for evaluability. Planning avoids designing a program that cannot be evaluated.
Population:	All persons in a particular group.
Prompt:	Reminders used by interviewers to obtain complete answers.
Purposive sampling:	Creating samples by selecting information-rich cases from which one can learn a great deal about issues of central importance to the purpose of the evaluation.

Qualitative evaluation:	The approach to evaluation that is primarily descriptive and interpretative.
Quantitative evaluation:	The approach to evaluation involving the use of numerical measurement and data analysis based on statistical methods.
Random sampling:	Drawing a number of items of any sort from a larger group or population so that every individual item has a specified probability of being chosen.
Recommendations:	Suggestions for specific actions derived from analytic approaches to the program components.
Sample:	A part of a population.
Secondary data analysis:	A reanalysis of data using the same or other appropriate procedures to verify the accuracy of the results of the initial analysis or for answering different questions.
Self-administered instrument:	A questionnaire or report completed by a study participant without the assistance of an interviewer.
Stakeholder:	A stakeholder is one who has credibility, power, or other capital invested in a project and thus can be held to be to some degree at risk with it.
Standardized tests:	Tests that have standardized instructions for administration, use, scoring, and interpretation with standard printed forms and content. They are usually norm-referenced tests but can also be criterion referenced.
Strategy:	A systematic plan of action to reach predefined goals.
Structured interview:	An interview in which the interviewer asks questions from a detailed guide that contains the questions to be asked and the specific areas for probing.
Summary:	A short restatement of the main points of a report.
Summative evaluation:	Evaluation designed to present conclusions about the merit or worth of an intervention and recommendations about whether it should be retained, altered, or eliminated.
Transportable:	An intervention that can be replicated in a different site.

- Triangulation:** In an evaluation, triangulation is an attempt to get a fix on a phenomenon or measurement by approaching it via several (three or more) independent routes. This effort provides redundant measurement.
- Utility:** The extent to which an evaluation produces and disseminates reports that inform relevant audiences and have beneficial impact on their work.
- Utilization of (evaluations):** Use and impact are terms used as substitutes for utilization. Sometimes seen as the equivalent of implementation, but this applies only to evaluations that contain recommendations.
- Validity:** The soundness of the inferences made from the results of a data-gathering process.
- Verification:** Revisiting the data as many times as necessary to cross-check or confirm the conclusions that were drawn.
- Sources:**
- Jaeger, R.M. (1990). *Statistics: A Spectator Sport*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluations (1981). *Standards for Evaluation of Educational Programs, Projects and Materials*. New York: McGraw Hill.
- Scriven, M. (1991). *Evaluation Thesaurus*. 4th Ed. Newbury Park, CA: Sage.
- Authors of Chapters 1-7.