Running head: Case History Research: An Introduction

Case History Research: An Introduction

Theodore C. Smith

ED 8113

Advanced Research Methods

7053 Enright Drive

Citrus Heights, CA 95621

Telephone: 916-726-4961

Email: tcsmith@surewest.net

Instructor: Larry Reynolds, Ph. D.

Abstract

Case studies are a form of qualitative research and have specific characteristics regarding purpose, design, data collection, analysis, validity, reliability, generalizability, presentation, advantages, and disadvantages. At least two sets of criteria exist for judging adequacy of case study reports—one for assessing suitability for use in classroom discussion and one for assessing suitability for use in making decisions about programs. This analysis suggests that (1) some case studies suitable for use in the classroom may lack the rigor required for making critical decisions and (2) some case studies that include comprehensive assessments and extensive documentation may be overly complex and difficult to use to stimulate discussion in educational settings. Thus, the critical factor regarding suitability may depend on intended use.

CONTENTS

Index of Tables

Index of Figures

Introduction

The case study is a form of qualitative research that is widely used. This paper outlines the characteristics of case study methodology, discusses two sets of criteria for assessing whether a given case study is "good," and uses these two sets of criteria in a limited review of two quite different case study reports.

The first section includes a definition of case study methodology, describing pertinent characteristics, purpose, design, data collection, analysis, validity, reliability, generalizability, presentation, advantages, and disadvantages. The second section compares and contrasts two sets of criteria for judging suitability of use for (a) classroom discussion and (b) decision-making. The third section examines two case study reports using these criteria and discusses appropriate application of such criteria. The final section summarizes the results of this study and presents recommendations for future research.

Overview of Case Study Methodology

Definition and Characteristics

Case studies are a type of qualitative research—the study of (a) an event that occurred in its natural context, (b) how it was interpreted and reacted to by participants, and (c) consequences and outcomes (Gall, Gall, & Borg, 2003). The purpose of case study research is to shed light on a given phenomenon. Virtually any phenomenon may be studied via a case study. Stake (2000) notes that, "As a form of research, case study is defined by interest in individual cases, not by the methods of inquiry used" (p. 435, as quoted by Gall et al, p. 435).

Some researchers study one case because of intrinsic interests while others study a suite of similar cases to test generalizability of patterns. Program Evaluation and Methodology Division (1990; hereafter referred to as PEMD), a unit with the U. S. General Accounting Office (GAO), notes that "conventional wisdom" holds that case studies are "always subjective and nongeneralizable" (p. 10). However, PEMD also suggest steps that can be taken to generalize case studies when desired. Epistemologically most researchers approach their case studies with an interpretive stance, intent on studying the meanings of the phenomenon to the actors involved in them, although some researchers adopt a positivist orientation, believing that observed physical and social phenomena are reality and, if unbiased, are independent of those who observe it (Gall et al, pp. 627 & 632).

Case study research has four characteristics. It involves:

1. The study of phenomena by focusing on specific instances (the case);
2. An in depth study of each case;
3. The study of a phenomenon in its natural context; and,
4. The study of the emic perspective—the research participants' perception and understanding of their social reality—of case study participants (Gall et al, p. 436).

Once the phenomenon is clarified, the researcher may select a case (an instance of the phenomenon) for intensive study. Any phenomenon has many aspects. Thus the researcher needs to select a focus—the aspect or aspects on which data collection and analysis will concentrate. The notion of focus tends to make the case study more manageable and meaningful (Gall et al, p. 437).

The focus in some case studies is divided into units (e.g., separate presentations of four similar instances), followed by a discussion of common themes. However, some studies are

holistic—they cannot be broken down into smaller cases. That is, simply documenting events in individual parts of an organization may not be informative of the overall entity, processes, or outcomes. For example, were one to separately examine the individual organizational units (finance, human resources, shipping, administration, assembly line, etc.) of a manufacturer and ignore the interactions among those units, one might gain an incomplete understanding of key factors that limit, affect, or enhance production and other outcomes.

The natural context of the phenomenon is a key component in case studies. Such studies typically involve fieldwork, but even fieldwork is not a component "the goal still is to learn about the phenomenon from the perspective of those in the field" (Gall et al, p. 438). Honan and Rule (2002) suggest that a case study is a "narrative description of a problem or dilemma written from the point of view of one or more protagonists" (p. 13). However, case studies may involve multiple points of view (e.g., a case that focuses on a situational conflict). In fact, Wrage (1994) suggests that cases which are well-suited for case-method teaching are (a) ambiguous, full of conflict, and leave important issues unresolved and (b) include varied and dynamic perspectives of participants.

Studying a phenomenon in context enables clarification of boundaries between the phenomenon and the context. For example, how does the phenomenon relate to and yet distinct from previous experiences; future plans; external activities, engagements, and roles; and the current setting being investigated? Case studies explore how participants react to events and outcomes. In an effort to increase understanding, emic researchers typically ask participants how they felt, what they thought, and why they reacted they way they did while etic researchers attempt to view phenomena and responses as objective, impartial outsiders.

Yin (1981) notes that a common misconception is that case studies are limited to ethnographies or participant observation; however case studies may involve collection of quantitative and/or qualitative data (pp. 58-59). He suggests that a case study is similar to an experiment, a history, or a simulation, each of which may involve collection of quantitative and or qualitative data. "As a research strategy, the distinguishing characteristic of the case study is that it attempts to examine: (a) a contemporary phenomenon in its real-life context, especially when (b) the boundaries between the phenomenon and context are not clearly evident" (p. 59). In contrast, he asserts that (a) experiments deliberately divorce a given phenomenon from its context while (b) histories involve studies of past events rather than real-time, direct observations. Furthermore, because context is a part of all case studies, "there will always be too many 'variables' for the number of observations made" (Yin, p. 59).

Purposes of Case Studies

Yin (1959) indicates that case studies may be (a) exploratory, (b) descriptive, or (c) explanatory and are best suited to studies that involve or produce explanations rather than descriptions of individual, non-repetitive incidents. Gall et al report that researchers use case studies to:

1. Describe and conceptualize a phenomenon;
2. Explore what the phenomenon means to participants;
3. Identify contextual factors that influence participants;
4. Document, explore, explain, or evaluate events and outcomes; and/or
5. Document, explore, explain, or evaluate the new and unusual (pp. 439-440).

PEMD (1990) notes that single incidents or anecdotes are memorable and convincing but not generalizable—it does not reveal whether the phenomenon is an isolated incident or

common, widespread occurrence (p. 8). Most GAO case studies either (a) illustrate findings established through other types of research or (b) constitute an in-depth study of a case of unique interest. PEMD recognize six types of case studies that might be appropriate for government use (Table 1). In essence, the sixth is a meta-case study—a process of combining and analyzing results from multiple case studies.

Table 1

Types of Case Studies Commonly Developed by U. S. General Accounting Office

Type	Description
Illustrative	A descriptive study that is intended to add realism and in-depth examples to information about a program or policy.
Exploratory	A descriptive study focused on generating hypotheses for later investigation.
Critical instance	Examines a single instance of unique interest or serves as a critical test of an assertion about a program, problem, or strategy.
Program implementation	Investigates operations, often at several sites and often normatively.
Program effects	Examines causality and usually involves multi-site, multi-method assessments.
Cumulative	Assimilates findings from several case studies to answer an evaluation question, whether descriptive, normative, or cause-and-effect.

Source: Modified after PEMD (1990, pp. 9-10).

Case studies typically involve a "thick description"—statements that recreate a situation and as much of its context as possible, accompanied by meanings and intentions. Thick descriptions use constructs that bring order to descriptive data (e.g., historical framing, context). Researchers may add depth to descriptions by identifying recurring instances and patterns (defined themes) within the case. Such themes enable identification of systematic relationships between variations (relational patterns), causality (causal patterns), and formative evaluations.

However, although thick descriptions are emphasized, case studies ideally are brief, well written, and to the point (Center for Public Health Practice, n. d.; Honan and Rule, 2002).

Design

The design of each case study is dependent upon (a) the phenomenon being studied and (b) the interpretation of the researcher who develops the design. Throughout any given case study, the researcher's beliefs, past experiences, biases, and interpretations influence the questions asked, the data gathered, any design adjustments made while the study is being conducted, and any analysis made during or following the study. In some case studies, the researcher is a participant. Maxwell (1998) identified five basic components (or sets of issues) of a case study, influenced by 14 factors (Figure 1). The key, Gall et al suggest, is threefold:

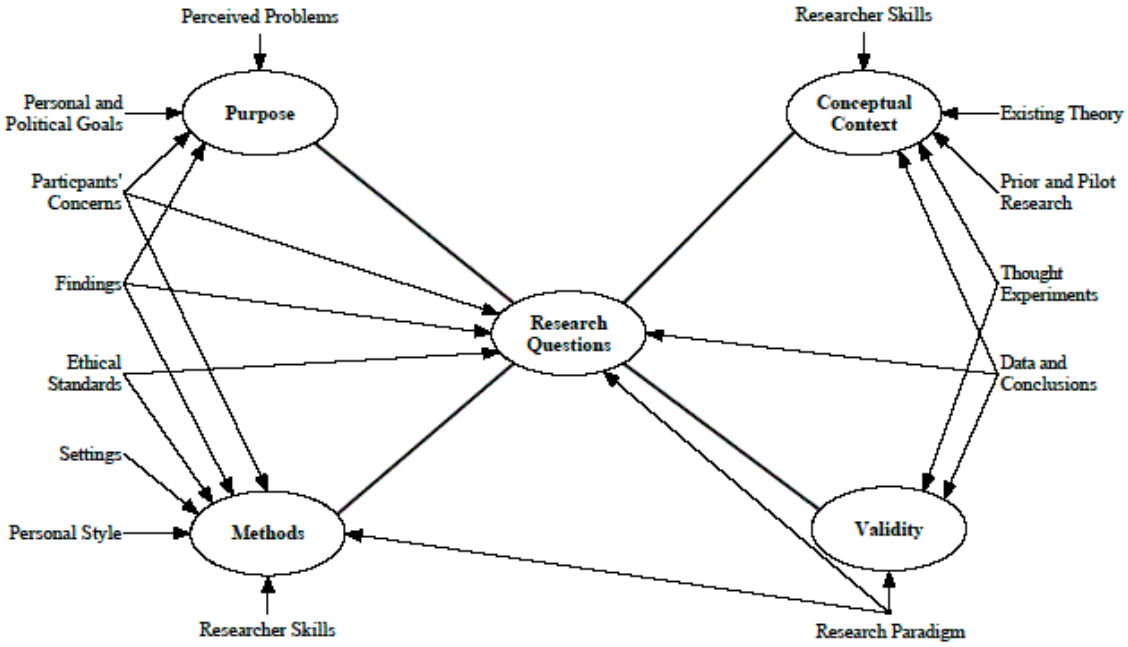

Figure 1. Components and contextual features of qualitative research design, from Gall, Gall, and Borg (2003, p. 442, after Maxwell, 1998, p. 73).

"[a] respecting the unique features of the case, [b] keeping interpretation at the center of the case study, and [c] creating a research design that has coherence and value" (pp. 442-443).

Design of a case study involves identifying a research problem that is worthy of study, formulating appropriate research questions, and selecting a case (or cases) for study. Gall et al suggest that following identification of a problem that is suitable for a case study, the researcher needs to translate the problem into questions or objectives. What question or questions does the researcher want to answer or, alternatively, what behaviors, events, impacts, or other phenomena does the researcher want to better understand?

A given case may be selected because it (a) is typical, (b) reflects the phenomenon of interest, (c) is unique and of special interest, (d) is politically important, or (e) other appropriate reason. PEMD (1990) suggests three general bases for selecting instances: (a) convenience, (b) purpose, and (c) probability (p. 25; Table 2), stressing the importance of selecting (a) a truly representative sample and (b) the correct basis:

> A good case study will use a basis for instance selection that is appropriate for the question to be answered. Using the wrong basis for selecting an instance is a fatal error in case study designs, as in all designs. Such a case study is a not-good case study, and it is irredeemably flawed despite any methodological virtues it may have in terms of data collection, analysis, and reporting (PEMD, p. 25).

Some cases may involve groups while others involve only one or more individuals. In some instances, it may be appropriate to study entire groups (or all the individuals therein), while in other instances the study may be limited to a small sample (Gall et al, p. 444).

Data Collection

In-depth case studies involve the collect of a substantial amount of data (words, images, objects, and, sometime, quantitative data), often over an extended time and via several methods. For example, case studies may involve taped interviews that are subsequently transcribed and

Table 2

Guidelines for Instance Selection in GAO Case Studies

Selection basis	When to use and what questions it can answer
Convenience	Convenient for data collection. What is happening and why?
Purpose	
Bracketing	What is happening at the extremes? What explains various differences?
Best cases	What accounts for an effective program?
Worst cases	Why isn't the program working?
Cluster	How do different types of programs compare with each other?
Representative	In instances chosen to represent important variations, what is the program like and why?
Typical	In a typical site, what is happening and why?
Special interest	In this particular circumstance, what is happening and why?
Probability	What is happening in the program as a whole, and why?

Source: Modified after PEMD (1990, p. 27).

analyzed to permit development of a coherent narrative. Interviewers engage in an interactive process of reviewing and clarifying the transcripts, narratives, and final product.

Miles (1979) criticizes qualitative research as being burdensome. "Qualitative data tend to overload the researcher badly at almost every point: the sheer range of phenomena to be observed, the recorded volume of notes, the time required for write-up, coding, and analysis can all become overwhelming" (p. 590). In describing the four-year case study of six public schools that he and his colleagues conducted, he details the numerous hours required to write notes, transcribe tapes, review and correct typing, and code data. He notes that such activities often

absorbed so much time that little was left to devote to data analysis. He asserts that, because of the prevailing view that all data are critical, while attempting to record everything

> . . . data analysis becomes somewhat meaningless, and fieldworkers get progressively less interested in doing it. Instead of really understanding and thinking about the site, the field worker risks being run by the site. And the richer and fuller the contacts with the site, the more reluctant the fieldworker is to miss a new and crucial site event (p. 593).

Yin (1981) responded by noting that Miles, while candid, contained few suggestions for overcoming the research issues raised. Thus Miles "leaves the reader with a sense that qualitative analysis—and its implicit companion, the case study—cannot yet be regarded a rational, much less scientific venture (Yin, p. 58). Yin suggests three ways that the problems reported by Miles can be minimized or avoided:

1. Avoid spending an inordinate amount of time writing and polishing narrative accounts unless publication of the narrative is required; instead, take notes and organize any narratives around specific questions or propositions.
2. Tabulate meaningful, rather than all, events. This requires prior identification of the primary questions of the study. New conceptualizations, if any, developed during the study will require new questions be posed and, if needed, new data gathered.
3. With respect to developing explanations, Yin suggests that researchers continuously make decisions about relevance of various data, similar to the method exercised by a police detective while investigating a case. Some clues must be pursued vigorously while other data can be regarded as irrelevant. However, during such a process, one also should pursue plausible alternative explanations (Yin, pp. 60-61).

Gall et al believe that data collection is emergent in case study research—that is, the researcher probably will learn something from the data collected that will influence or determine

subsequent data collection activities (p. 449). Thus, the researcher needs to analyze the data while the research is being conducted. Gall et al offer two strategies to facilitate this process:

1. Use standard forms to (a) record details of data gathered and summarize what was learned from each interview or field observation and (b) briefly summarize the content of each relevant document (after Miles and Huberman, 1994).
2. Think the whole research project through from start to finish prior to beginning fieldwork (after Wolcott, 1994, p. 404); consider the final form of the presentation and the relative emphasis of (a) the thick description and (b) the analysis and interpretation. Such an approach helps identify the types of data to collect and the level of detail.

Lincoln and Guba (1985) offer four criteria for determining when it is appropriate to end collecting data:

1. Data sources are exhausted (yield little new data that are relevant).
2. Categories are saturated; that is, the effort to gather data greatly exceeds the value of data gained.
3. Data are sufficiently consistent that regularities emerge.
4. New data are "far removed from the central core of viable categories that have emerged" and does not contribute new viable categories (as reported in Gall et al, p. 452).

Analysis

Gall et al (1990) focus on analyzing the verbal data gathered during a case study while acknowledging that images (e.g., drawings, film, photographs) also may be important sources of data. They identify three types of analysis (interpretational, structural, and reflective) as appropriate for analyzing case study data.

Interpretational Analysis

Interpretational analysis involves examining data to identify constructs, themes, and patterns useful for describing a given phenomenon. The analysis may be performed manually or by computer. The first step in the process is to assemble all of the data, converting images and other materials into descriptive forms. Next, the resultant database is split into meaningful segments or units. Such segments typically include one meaningful item (e.g., a question and a response) that is comprehensible even if read outside the context of the study. Segments may be of any length—a sentence, paragraph, or several pages of text. During this process, the location (e.g., page and line numbers) of each segment are recorded.

Next, the researcher develops meaningful categories—constructs that refer to specific phenomena mentioned—to encompass and summarize the data, along with subcategories, if warranted. Gall et al suggest using a list of categories that other researchers have developed. If categories need to be developed, the researcher need to define the category and specify guidelines for determining whether segments are or are not instances of the category.

Miles (1979) described a case wherein the number of categories for key actors being tabulated expanded from about 100 to 202 as analysis proceeded, yielding a "meaningless" (p. 594) coding task, causing field researchers and coders to hate their jobs (they eventually refused to code the data), and delaying project completion. In the end, these 202 codes were reduced to 23 "major 'themes'" (p. 594). Yin (1981) warns against using categories that are too small or too numerous, noting that doing either tends to overwhelm those doing the coding. Instead he suggests that data be reviewed prior to formal coding to scale the categories down to a manageable number of meaningful events (pp. 60-61).

The process of developing categories requires knowledge of grounded theory construction, in particular key processes of theoretical sensitivity, theoretical sampling, and theoretical saturation, the goal being to produce testable outcomes (Taber, 2000, p. 469). Taber defines theoretical sensitivity as a process (outlined in Figure 2) whereby one begins research with an open mind so that observations are colored as little as possible by expectations. This process ideally leads to an emergent fit—one where the category is adjusted to fit the data rather than the data being adjusted to fit the category. The approach uses multiple slices of data as a means of triangulation. A researcher's "theoretical sensitivity" during analysis leads to hunches that suggest the next stage of data collection. And "theoretical saturation" is reached when further collection and analysis of data does not significantly change the model being developed (Taber, pp. 470-471; similar to Lincoln and Guba's (1985) criteria for deciding when sufficient data are gathered).

After meaningful categories are developed, each segment is examined and coded to reflect whether it reflects none, one, or more instances of any categories in the system. Some coding systems also record the degree to which the instance is present (e.g., not present, present to a moderate degree, present to a high degree). If some categories are ambiguous or contain information that cannot be coded, the coding system may need to be revised and the data re-coded.

Following coding, the researcher typically groups all of the segments tagged with specific codes, reviewing the content to determine whether specific constructs are reasonable, supported by the data, or overlap with other constructs. If a given construct proves unreasonable or overlaps with another, redefinition or development of new categories may be necessary. This

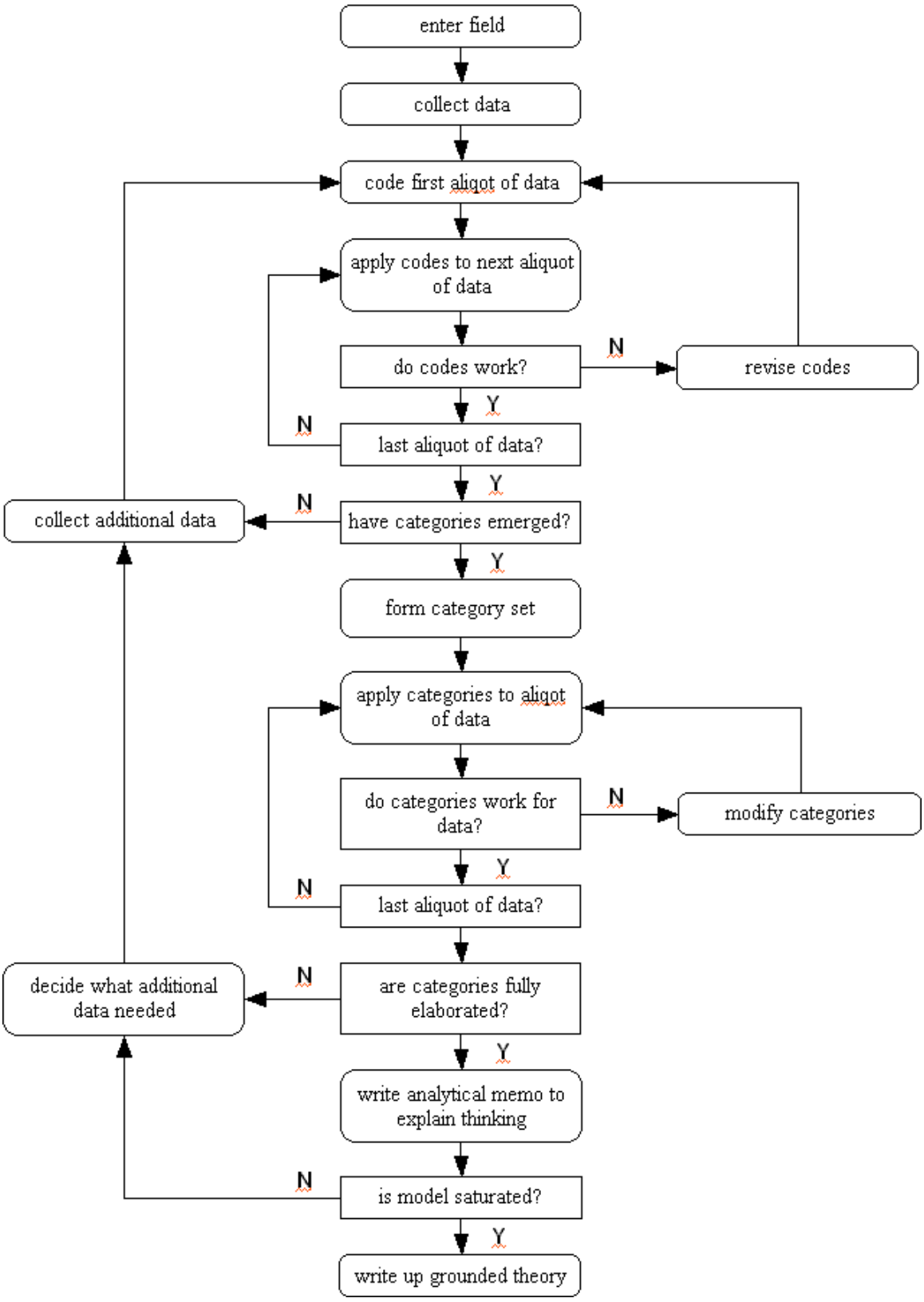

Figure 2. Schematic showing process for development of grounded theory (from Taber, 2000, p. 472).

continued or iterative process of comparing and revising categories until satisfactory closure is achieved is called "constant comparison" (Gall et al, p. 456).

One of the goals of the research is to discover recurring constructs and themes, either within a case or across multiple cases. Data from the latter also may be analyzed to discover relational or causal patterns. Yin (1981) notes, however, that to warrant cross-case tabulations requires a large enough number of cases (p. 62).

Structural Analysis

Structural analysis attempts to identify inherent patterns in discourse, events, text, or other phenomena (Gall et al, p. 457). For example, whereas interpretational analysis focuses on content or meaning within a conversation, structural analysis might note the sequence of speakers; count the number of words uttered; document repetition, use of language, use of jargon, movement patterns, or sequences in stories, textbooks, or curriculum. Structural analysts might attempt to document changes in such patterns over time—a longitudinal analysis.

Reflective Analysis

While interpretational and structural analyses involve explicit procedures to portray or evaluate the phenomenon being studied, reflective analysis relies greatly on the intuition and judgment of the researcher. Reflective analyses involve "a decision by the researcher to rely on her own intuition and personal judgment to analyze the data rather than on technical procedures involving an explicit category classification system" (Gall et al, p. 459).

Reflective analyses sometimes involve teams of researchers in a dynamic process to produce rich interpretations of data, bringing together "different kinds of knowledge, experience, and beliefs to forge new meanings through the process of joint inquiry . . ." (p. 459). Reflective processes typically produce thick descriptions but also may lead to discovery of constructs,

patterns, and themes. Reflective analyses may also aid educators and policy makers in examining (a) the purposes and features of educational programs, products, and methods and (b) strengths and weaknesses.

Reflective analyses are largely subjective, lack specified procedures, and typically benefit greatly from apprenticeship with an experienced researcher. Guidelines for hermeneutic research may be applicable. The analyses typically require examination of all data collected to identify salient features, followed by development of an understanding of such features by themselves and in context with each other and the environment. Gall et al note, "An interpretation or criticism that fits some of the data should not be contradicted by other data" (p. 460).

Hermeneutics, as the term is used in philosophy and social science, refers to the study of the process by which individuals arrive at the meaning of a given text, custom, or social myth. "Hermeneutic theorists claim that there is no objective reality and, therefore, no possibility of developing correct knowledge about reality. Instead, we develop interpretations of the world" (Gall et al, p. 506). Following this line of reasoning, authors express their interpretation and each reader comprehends the resultant text through his or her own interpretive process; thus authors and readers are "entangled in interpretive processes" and "There is no way of knowing the world objectively outside the interpretive act" (p. 506).

One of the principles of hermeneutics is the hermeneutic circle—"a continuous process of alternating between interpreting the meaning of an individual part of the text and the text as a whole" (p. 506). Interpretation of parts helps interpret the whole, and interpretation of the whole helps interpret the parts. The hermeneutic concept of a fused horizon recognizes that, although participants typically argue their own opinions and prejudices, they may be capable of (a) perceiving and translating the opinions and prejudices of other participants, (b) achieving a

common language, and, ideally, (c) producing "one human community of thought and action" (a fused horizon; Gall et al, pp. 506-507).

Validity and Reliability

Because researchers hold different views of the nature of reality and scientific inquiry, they also have different views regarding criteria to assess validity and reliability of case studies. Researchers who subscribe to a positivist philosophy claim that it is possible to obtain objective knowledge about the world. Gall et al suggest that Yin is a typical positivist case study researcher and that he uses three validity criteria and one reliability criterion:

1. Construct validity is the extent to which a measure used in a case study correctly operationalizes the concepts being studied.

2. Internal validity is the extent to which the researcher has demonstrated a causal relationship between X and Y by showing that other plausible factors could not have cased Y. The criterion of internal validity is not applicable to descriptive case study research because it does not seek to identify causal patterns in phenomena.

3. External validity is the extent to which the findings of a case study can be generalized to similar cases.

4. Reliability is the extent to which other researchers would arrive at similar results if the studied the same case using exactly the same procedures as the first researcher (Gall et al, 2003, p. 460, after Yin, 1994).

Regarding item 2 above, PEMD (1990) notes that a valid construct or measure:

> . . . reflects what it claims to reflect and not something else. For example, whether or not there are active opposition parties may be a more valid measure of whether a country is a democracy than how many people vote in an election. A valid cause-and-effect design—that is, one with internal validity—rules out alternative explanations of results by comparing what happened with an intervention to what happened in the absence of the intervention. For example, in a study of the effects of an employment training program, greater employment of participants after the training than before must be shown to be due to the training and not simply to better economic conditions, which also could increase employment (p. 63).

The PEMD illustration demonstrates that case studies that seek to identify cause and effect relationships need to examine the environmental factors that may not be evident in data gathered and/or may not be known to all participants.

A strong chain of evidence tends to strengthen the overall validity of a study. Audit trails—documentation that describes sources, when and how data were collected, how data were reduced, analyzed, reconstructed, and synthesized (and by whom), intermediate products (including notes and other materials), descriptions of intentions and dispositions, and instrument development information—help make the chain of evidence explicit. If voluminous, these materials do not necessarily need to be included in the final product; instead, representative samples may suffice. The materials need to be kept for a number of years in case other researchers wish to examine them.

Some cases are designed as experiments to test the effects of an intervention. There are procedures for testing the internal validity of causes in such case studies. These procedures require a set of theoretical propositions to be tested against the data gathered. If patterns are discovered that correspond to predictions drawn from the theoretical propositions, the causal inference is strengthened.

Researchers that subscribe to a constructivist, reflexive, or postmodern views generally hold that notions of validity and reliability do not apply to case study data and interpretations. Instead criteria such as authenticity, credibility, plausibility, and reliability are used, or alternative formulations of validity (Gall et al, p. 462). Validity assessments involve various combinations of (a) analyses of researcher positioning, (b) assessments of contextual completeness, (c) coding checks, (d) long-term observations, (e) member checking, (f) outlier

analysis, (g) representativeness checks, (h) triangulation, (i) usefulness assessments, and (j) use of specific reporting styles.

Researchers ideally demonstrate sensitivity in relating to the case being studied, present data in ways that are perceived as credible and authentic, and set the data within a context. Useful findings enlighten or liberate individuals or groups. Triangulation attempts to demonstrate credibility through corroboration. Member checking is a process wherein individuals review statements contained in the researchers report for accuracy and completeness. Outlier analysis involves identifying outliers and verifying presence of another factor or absence of a key factor found in mainstream cases.

Generalizability

Generalizability is considered an important goal in quantitative research. However, some case studies (e.g., an atypical case) probably are not generalizable although they still may be worthy of investigation. A case study that involves a random sample of a given population (e.g., students in a specific class) may yield generalizable information. Researchers can help readers determine the generalizability by (a) including a thick description of involved participants and contexts, (b) stating whether the case is representative of a general phenomenon, and, if a multiple-case design is used, (c) conducting a cross-case analysis to determine whether there is generalizability within the cases studied (Gall et al, p. 466).

PEMD (1990) agrees with Gall et al that because a case study typically examines a single instance, "it cannot contribute directly to the testing of general propositions" (p. 86). Instead "the product is a sharpened understanding of what might be important to look at further in similar situations . . . and can contribute powerfully to the invention of hypotheses" (p. 86).

Presentation of Results

Once all of the data are collected, the researcher may decide to focus on historical data, one or more specific aspects, or one or more specific events. Gall et al cite an example (Rossi, 1993) wherein the researcher excluded one participant's data from the report because of "unusual circumstances pertaining to the individual" (Gall et al, p. 467). This illustrates that decisions about what to include in a case study may be made late in the process. Ideally one should winnow the data and include only cases and aspects that have the greatest bearing on questions of the research interest, while including enough information to reveal sufficient context (Wolcott, 2001, p. 44, as cited in Gall et al, p. 467).

Yin (1981) states that "the typical case study report is a lengthy narrative that follows no predictable structure and is hard to write and hard to read" (p. 64). He suggests that such problems can be avoided via a clear conceptual framework. He also suggests presenting data as a list of answers to open-ended questions, making the results easier for readers to skim. Obviously organization is key to effective presentation of data.

Gall et al (2003) indicate that published case studies typically are presented as either reflective or analytic reports, although occasionally these two styles are mixed. If the styles are mixed, one style typically is dominant.

Reflective reporting uses literary devices to bring the case to life and reveals the researcher's "voice" and conveys the researcher's point of view. Gall et al indicate that reflective reports typically weave the case study data into a story. Events may be presented in chronological order, focused on a key event, recounted by several participants, or reported as a "typical" day in the life of a participant (p. 467). The organization of the report highlights what the researcher has learned. Occasionally researchers who have a postmodern perspective may use

more innovative or dramatic methods (e.g., a poem, drama, comedy, satire) to convey findings, partly through dynamic interaction with the audience that creates a shared emotional experience.

Researchers who (1) adopt a positivist and postpositivist perspective and (2) emphasize interpretation or structural analysis tend to use an analytic reporting style similar to that used in reports of quantitative research. Analytic reporting is characterized by an objective style and a conventional organization (e.g., introduction, literature review, methodology, results, and discussion; Gall et al, p. 470).

The organization of multiple-case studies tends to highlight different aspects. If, for example, presenting each case by itself with constructs, themes, and patterns tends to assist readers in gaining a holistic understanding but inhibits cross-case analysis. When constructs, themes, and patterns and illustrated by multiple examples, readers tend not to gain a holistic understanding of each case. Gall et al suggest that combining the approaches works well.

Advantages and Disadvantages

Gall et al (2003) believe that one of the advantages of case studies is that the researcher, through a thick description, can bring a case to life, providing readers with a better basis for developing theories, designing interventions, or taking specific action than provided by pure statistical analyses and sparse descriptions. Case studies also reveal the researcher's perspective, enabling readers to determine whether their own perspective is similar or differs.

Case studies are well-suited for investigating unusual phenomena, outliers, and specified individuals or situations. The emergent quality of case studies enables researchers to reframe research questions (or frame new ones); modify, adopt, or abandon data-collection methods; and change the focus of the case as data are collected and analyzed (p. 472).

The principal disadvantages are (a) the difficulty of generalizing findings for other situations and (b) ethical issues that may arise. For example, it may prove difficult to disguise an organization or the identity of the individuals studied. Case studies tend to be labor intensive and require considerable language skill to (a) identify constructs, themes, and patterns and (b) write an effective report.

Characteristics of a Good Case Study

An English proverb states "Beauty is in the eye of the beholder" means that "beauty is in the mind that observes it" (BookBrowse.com, 2004). The adage has application in professional circles in that beauty may equate to adequacy or suitability. In essence, the assessment of what is "good" may depend upon the purpose or intended use. There are at least two sets of criteria for discerning what constitutes a good case study—one by Honan and Rule (2002b) and one by PEMD (1990).

Honan and Rule Criteria

In their guidebook that describes how college faculty can as instructional tools, Honan and Rule (2002b) identify the characteristics of effective ("good") and ineffective ("bad") cases. Assuming that a given case matches the instructor's pedagogical goals and facilitates learning (a key requirement of an effective case), such cases exhibit some or all of the characteristics shown in Table 3.

As Honan and Rule suggest, writing the report requires considerable skill. Miles (1979) described the process of analyzing the data in his case study during writing as "intuitive, primitive, and unmanageable in any rational sense" (p. 597) even though his team developed some "rules of thumb" as aids. He notes:

Table 3

Characteristics of Relatively Good and Bad Case Studies for Instructional Purposes

Good	Bad
Is clearly written	Is poorly written (e.g., unorganized or unclear)
Has a clear chronology	Chronology is vague
Is focused on one or two core issues	Focus is vague
Has a compelling story line	Is not compelling
Presents one or more sides or perspectives in a reasonably balanced fashion	Presents a one-sided or skewed portrayal of events
Reflects challenges, problems or dilemmas that are faced by a large number of practitioners and important to a specific field	Lacks an obvious problem or dilemma
Links to broader problems of practice and one or more wider contexts	Is too idiosyncratic and, therefore, is inapplicable to other contexts
Sustains discussion; presents a puzzle that lingers with participants long after the discussion is over	Fails to sustain readers' interest or participants' discussion

Source: Extracted from and modified after Honan and Rule (2002b, pp. 14-15)

> While one can remember occasional use of the "rules of thumb" (e.g., a pause to search for negative evidence), the analysis process is more memorable for its moments of sheer despair in the face of the mass of data, alternating with moments of achieved clarity, followed by second-guessing skepticism ("Would someone else come to the same conclusion?") (p. 597).

However, in sprite of these difficulties, "The final cases had a strong ring of truth, as well as being fascinating accounts" (p. 597).

PEMD Criteria

PEMD (1990) also provides guidelines for reviewing case study reports, focusing on strengths and limitations from an applied perspective (Table 4). PEMD concludes that studies

Table 4

GOA Guidelines for Reviewing Case Study Reports

1. Are the evaluation questions stated clearly and explicitly?
2. Is the case study application clearly described? Is it appropriate?
3. Was the time span of the study long enough to address the core issues fairly?
4. Is the basis for case selection presented? Is it appropriate for the purpose of the case study?
5. Are the methods of data collection presented? Are they appropriate for the purpose of the case study?
6. If more than one investigator collected the data, how were the other evaluators selected, trained, and supervised?
7. Are information sources described clearly and fully? Are they appropriate?
8. Are the procedures for the formation of the database described?
9. Are the techniques of data gathering and data processing explicitly described?
10. Were there interpretation differences, and if so how were they resolved?
11. If other studies, investigations, or experiments relevant to the issue are available, have their results been presented and reconciled with the case study findings?
12. Are methodological strengths and limitations identified clearly?
13. Are the arguments for various resolutions of the evaluation question presented?
14. Are the arguments against various resolutions of the issue presented?
15. Does the case study identify the factors explaining the phenomena that were observed and state clearly whether the identification of these factors was based on insight and recognition or on quantitative techniques?
16. What is known about the competence and impartiality of the investigators?
17. Are comments on the draft report available?
18. Is there adequate information for judging generalizability?

Source: PEMD (1990, pp. 113-118; condensed and renumbered)

which fail to meet these standards "have questionable merit" for decision-making purposes (p. 113).

Review of Two Case Studies

Introduction

Comparing and contrasting the Honan and Rule (2002) and PEMD (1990) criteria suggests that determination of whether a case study is "good" or "bad" may depend on intended use. Alternatively, some mix of criteria may be appropriate for some uses.

Clearly, when an instructor needs a tool to stimulate discussion, an engaging case study that addresses situations that have wide applicability is helpful. However, if the objective is to evaluate effectiveness, identify a problem, or address some other specific phenomenon, wide applicability is not necessarily a critical issue. Similarly, readability or ease of understanding may be critical in a classroom setting but not as important (although very helpful) in a setting where experts who are capable of dissecting the report are evaluating the results.

In contrast, if the solution of a problem has great social, financial, or political consequences, the PEMD criteria have greater, if not essential, importance—any one criterion that is not met may represent a fatal flaw. But even a flawed a case study may represent an opportunity for classroom discussion, especially if the flaw is exposed and discussed during the class.

The PEMD criteria are easily converted to a checklist (in fact, PEMD presents the checklist in somewhat expanded form in an appendix). The Honan and Rule criteria also can be converted into a list of questions (Table 5) for use in a checklist.

Table 5

Questions, Suitable for Checklist Use, Developed from Honan and Rule Criteria

1. Is the report clearly written?
2. Does the report contain a clear chronology of events?
3. Is the focus of the study clear?
4. Does the report have a compelling story line?
5. Does the report present relevant sides/perspectives in a reasonably balanced fashion?
6. Are the challenges, problems or dilemmas clearly described in the report?
7. Are the challenges, problems or dilemmas described in the report ones that are faced by a large number of practitioners and important to a specific field?
8. Would the report stimulate discussion? Does it present a puzzle that will linger with participants long after the discussion is over?

Use of These Checklists in Reviewing Two Case Studies

Do these two distinctly different checklists yield different results regarding identification of "good" and "bad" case studies? As an initial test, two distinctly different case studies were reviewed.

Overview of Mallon (2002)

The first is a case study by Mallon (2002), contained in Honan and Rule (2002a). [The two Honan and Rule (2002a; 2002b) publications are companion products. Honan and Rule (2002b) is a guide to use of case studies in higher education instruction, while Honan and Rule (2002a) contains several (presumably good) case studies for use in a course.]

The Mallon case is a 24-page examination of a dilemma faced by Olivet College—whether to reinstitute a tenure system for faculty. Mallon provides context via a historical

summary of events that led to replacement of tenure with a system of five-year contracts in 1974. In 1994, recognizing that the performance evaluation system was not working, the college's president suggested that he might entertain a return of the tenure system. The description contained in the case provides some information that permits comparison of pre-1974 conditions with 1994 conditions.

Mallon's primary focus is on events that occurred between 1994 and 1999, when some administrators and faculty began to promote the tenure concept as means to boost faculty morale and an incentive to encourage improvement. Quotes and descriptions of attitudes contained in Mallon make it clear that various individuals (the president, deans, board, and faculty) held pro and con positions of the proposal for a variety of reasons. The president gave faculty the responsibility to (a) explore the problems faced by the college and (b) develop a proposal for workable tenure and review system. Early on in the process, many faculty realized that tenure represented a threat to their employment—that only those holding terminal degrees would be eligible, that the "up or out" provisions of a tenure system meant that more than half the faculty would be dismissed unless they earned a doctorate in five to eight years. Administrators refused to grandfather any exiting faculty into tenure—all would face tenure reviews. Mallon reports that the institution made little progress on a study of the issue, in part because more critical issues (financial stability, curriculum revision) and administrative changes (hiring of a new president and dean) demanded attention.

Mallon contains five exhibits, including the five-year tenure policy, a proposed phase-in plan for tenure, and three policy documents that describe the institution's vision, mission, a description of individual faculty and staff responsibilities. No statistical data are included in the

study (but it does not appear that surveys or other instruments were used during the study period).

While Mallon does a good job of describing context, meaning, and a range of opinions held by various participants, the study is entirely qualitative—there are no data to support a finding regarding how widely held these opinions are or whether some faculty and administrators have other reasons for supporting or opposing tenure, or for delaying the decision-making process. It is clear that the process of studying and negotiating the issue helped identify and clarify faculty responsibilities, but it is not clear whether a tenure system would help the college improve performance.

Overview of Fitzelle (1998)

The second case study is a 218-page dissertation by Fitzelle (1998) that describes governance at a specific university (Cornell), partly to determine whether the balance of power is shifting from faculty to administrators as asserted by faculty at many institutions. Fitzelle notes that the assertion is based on perceptions and lacks specificity (it is aimed at universities in general, not a specific institution). Thus, he is not certain whether a power shift is, in fact, occurring and, if so, whether it is occurring at all levels within an organization.

Fitzelle focuses on clarifying elements of governance and exploring faculty perceptions. He describes four specific objectives:

1. To specify how faculty governance is differentiated in its facets and dimensions so that a sharper definition of the decisions in which faculty and administrators have more or less influence becomes clear.
2. Measure certain predictors of faculty governance in the exploration of possible determinants.

3. Specify where faculty perceive lines of authority over key decisions lie.
4. Contrast the degree of influence that different subgroups have (p. 6).

He also examines the impact that years of employment, rank, productivity, age, gender, and position type (assistant, associate, or full professor) have on perceptions.

Fitzelle used a mixed methodology for his study. He conducted surveys to gather statistical data, and supplemented the surveys with structured interviews. He also discusses the limitations regarding generalizability, recognizing that the results may not be applicable to other institutions or even to Cornell as perceptions change over time. His study includes clearly stated hypotheses, developed based on a fairly thorough literature review, that he attempted to prove or disprove.

Fitzelle frames the investigation by discussing how university governance has evolved in the United States since 1636 and at Cornell since its founding. The description identifies key changes in governance and reactions by specific entities. For example, after armed students seized one building on campus, governance failed to resolve the crisis. That prompted the faculty to create an academic senate, but Cornell trustees refused to codify faculty governance power. Although the senate persisted, its composition caused it to focus more on social issues than governance, and those involved dwindled. Subsequent changes sought an effective governance formula or structure. Fitzelle's descriptions are sufficiently thick for readers to identify the tensions and differing views among various interested parties.

Fitzelle details his methods, including descriptions of survey instruments and sources for specific questions, scaling, sampling methods, survey and interview procedures, reliability, and validity, including appropriate qualifications. He also describes the variables, their relationship to each other, analytical procedures, participation rates, reasons some faculty gave for not

participating, and coding procedures for comments faculty made on the survey. Survey instruments, interview guide, analysis of gender differences, and other support documents are included as appendices.

Data are presented and compared with results of prior surveys (conducted in 1975, 1989, 1994, and other years). In general, faculty was perceived as being less empowered at the university level over time, with little or no perceived change in involvement in governance at lower levels (e.g., department) over time. Fitzelle also found that faculty tended "to report a limited role of their own influence and view of administration as autocratic when asked general questions, [but] when asked about specific decision areas, they demonstrated a more balanced view toward sharing power with administration. Statistically some predictors (funding source of specific programs, level of position held) influenced results related to specific factors. Three factors—involvement in (a) control of academics, (b) institutional decision-making, and (c) resource allocation—help predict variance in faculty perceptions.

Checklist Results

Each checklist was (a) simplified to fit on a single page and (b) used to evaluate the two case studies (Tables 6 and 7). In general, the length, complexity, and academic nature of the Fitzelle study may detract from its usefulness as a tool for simulating discussion of some concepts (e.g., whether self governance is desirable at a university); however, it rich data and analysis enables identification and discussion of concepts that some might feel are more obscure (e.g., how personal roles and experiences impact perception) but may be worthwhile topics for classroom discussion. In contrast, the Mallon study lacked the depth of data and, without such data, may prove less useful or reliable as a decision-making tool. However, Mallon appears to

Table 6

Analysis of Two Case Studies Using PEMD (1990)

	Mallon			Fitzelle		
Question (after PEMD, 1990)	Y	N	?	Y	N	?
1. Evaluation questions stated clearly and explicitly?		X		X		
2. Case study application clearly described & appropriate?	X			X		
3. Study time span of study long enough?			?	X		
4. Basis for case selection presented & appropriate?	X					X
5. Methods of data collection presented & appropriate?		X		X		
6. Adequate training, selection, & supervision of evaluators?		X				n/a
7. Information sources appropriate & described clearly & fully?		X		X		
8. Procedures for formation of the database described?		X		X		
9. Data gathering and data processing techniques described?		X		X		
10. Interpretation differences, if any, resolved?			n/a	X		
11. Results of other relevant studies/investigations/ experiments presented & reconciled?		X		X		
12. Methodological strengths & limitations clearly identified?		X		X		
13. Arguments for various resolutions of evaluation question presented?	X			X		
14. Arguments against various resolutions of issue presented?	X					n/a
15. Observed phenomena explained and techniques identified?		X				n/a
16. Investigators competent & impartial?	X			X		
17. Comments on the draft report available?		X			X	
18. Adequate information for judging generalizability?		X		X		
TOTALS	5	11	2	13	1	4

Table 7

Analysis of Two Case Studies Using a Checklist Based on Honan and Rule (2002b)

	Mallon			Fitzelle		
Question (after Honan and Rule)	Y	N	?	Y	N	?
1. Clearly written?	X			X		
2. Clear chronology of events?	X			X		
3. Clear focus?	X				X	
4. Compelling story line?	X				X	
5. Balanced presentation of relevant sides/perspectives?	X			X		
6. Challenges/problems/dilemmas clearly described?	X			X		
7. Includes challenges/problems/dilemmas faced by many practitioners and important to a specific field?	X			X		
8. Will case stimulate discussion/present a puzzle that will linger with participants?	X				X	
TOTALS	8	0	0	5	3	0

offer a compact, readable, case that enables students to identify and debate important, clear issues.

The differences in the two rating scales (reflected in the totals and the qualitative review comments) suggest a difficulty in satisfying both sets of criteria—one (PEMD) focused on quality of data and analysis, and the other focuses on writing quality and ability to quickly describe issues and convey various points of view. This is not all that surprising—each represents a different type of thick description. Whereas each might be included in assigned reading for a course, one might direct students to use a different approach—e.g., read Mallon and skim Fitzelle. However, short case study reports like Mallon need to include sufficient data to

remain credible and minimize side-discussions regarding validity and reliability of analysis, findings, and recommendations.

Summary, Conclusions, and Recommendations

Case studies are a widely used form of qualitative research and differ from other forms of qualitative methodologies. Case studies have specific characteristics, purpose, design, data collection, analysis, validity, reliability, generalizability, presentation, advantages, and disadvantages.

At least two sets of criteria exist for judging adequacy of a case study report. One set focuses on suitability of the report use for use in a classroom discussion. The other focuses on suitability of the report for use in making decisions about programs.

The limited analysis contained herein suggests:

1. Some case studies that are suitable for use in the classroom may lack the rigor required for making critical decisions.
2. Some case studies that contain comprehensive assessments and extensive documentation may be overly complex and difficult to use to stimulate discussion in educational settings.

Thus, as the English proverb suggests, for case studies, beauty is in the eye of the beholder—or, more appropriately beauty (what makes a good case study) depends on the intended use by specific users.

As for recommendations, this effort represents a very limited effort to identify criteria regarding what constitutes an effective case study report (and case study). Other lists of criteria may exist. Future research might (a) attempt to develop a unified set of criteria or (b) use criteria similar to those contained herein to review a broader spectrum of case study reports.

References

BookBrowse.com. (2004, April 14). *B I I T Eye O T B*: author. Retrieved June 13, 2004, from http://www.bookbrowse.com/wordplay/2004-04-14.cfm

Center for Public Health Practice. (n. d.). *Turing ideas into case studies*. Retrieved May 25, 2004, from http://ist-socrates.berkeley.edu/~cphpweb/turningideas.html

Fitzelle, G. T. (1998). *Faculty governance in higher education: A case study of its definition and determinants.* Unpublished Ph. D. dissertation, Cornell University.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*. Boston, MA: Allyn and Bacon.

Honan, J. P., & Rule, C. S. (Eds.). (2002a). *Casebook I: Faculty employment policies* (1st ed.). San Francisco, CA: Jossey-Bass.

Honan, J. P., & Rule, C. S. (2002b). *Using cases in higher education: A guide for faculty and administrators* (1st ed.). San Francisco: Jossey-Bass.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications. [As cited by Gall et al, 2003.]

Mallon, W. T. (2002). Olivet College: Aligning faculty employment policies with an evolving mission. In J. P. Honan & C. S. Rule (Eds.), *Casebook I: Faculty employment policies* (1st ed., pp. 77-100). San Francisco: Jossey-Bass.

Maxwell, J. A. (1998). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 69-100). Thousand Oaks, CA: Sage Publications.

Miles, M. B. (1979). Qualitative data as an attractive nuisance: The problem of analysis. *Administrative Science Quarterly, 24*(4), 590-601.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications. [As cited in Gall et al, 2003.]

Program Evaluation and Methodology Division [PEMD]. (1990, November). *Case Study Evaluations, GAO/PEMD-91-10.1.9*. Washington, DC: United States General Accounting Office. Retrieved May 17, 2004, from http://www.gao.gov/policy/10_1_9.pdf

Rossi, M. A. (1993). *The California Mathematics Project: Empowering elementary teachers to be leaders and change agents in mathematics reform.* Unpublished Ph. D. dissertation, University of Oregon, Eugene, OR.

Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. pp. 435-454). Thousand Oaks, Calif.: Sage Publications [As cited in Gall et al].

Taber, K. S. (2000, May). Case studies and generalizability: Grounded theory and research in science education. *International Journal of Science Education 22*(5), 469-487.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage Publications. [As cited in Gall et al, 2003.]

Wolcott, H. F. (2001). *Writing up qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications. [As cited in Gall et al, 2003.]

Wrage, S. D. (1994, Spring). Best case analysis: What makes a good case and where to find the one you need [Electronic version]. *International Studies Notes, 19*(2). Retrieved May 25, 2004, from http://csf.colorado.edu/isa/isn/19-2/Wrage.htm

Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly, 26*(1), 58-65.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.