Quality and Quantity, 17 (1983) 281–339
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Drawing Valid Meaning from Qualitative Data: Some Techniques of Data Reduction and Display *

A. MICHAEL HUBERMAN
University of Geneva

MATTHEW B. MILES
Center for Policy Research

Introduction

There are clear signs these days that qualitative research in education has come of age. Its methodology receives regular reviews and updating in the mainstream journals and anthologies (e.g., Wilson, 1977; Smith, 1978) and is frequently taught in graduate schools alongside basic statistics courses. RFPs (requests for research proposals) now call routinely for the buttressing of surveys by field studies or other "softer" data-collection modes. Ethnographic jargon—"thick description," "grounded theory," "triangulation"—can now be heard on the lips of the most rigorous psychometricians without the slightest note of derision. In fact, many of these people have undergone a process of mellowing, if not downright conversion, in their assessments of what naturalistic research can do, even within the limits of existing methodological canons (e.g., Snow, 1974; Cronbach, 1975; Campbell, 1975; Cook and Campbell, 1979). It may be recalled that not too long ago, Campbell and Stanley (1963) assimilated field studies to one group, that of post-test-only design, and dismissed this model as so lacking in control and randomization procedures as "to be of almost no scientific value" (p. 176).

Another good sign is that qualitative researchers have begun to move away from a largely defensive posture (e.g., spending much of their time pitting the strengths of field studies against the weaknesses of survey approaches, or dwelling obsessively on safeguards against measurement error in field studies in order to show that qualitative research can be even *more* rigorous than correlational and experimental studies). Interest now has

* Prepared for the Symposium on Advances in the Analysis of Qualitative Data, American Educational Research Association Meeting, March 1982.

0033-5177/83/$03.00

shifted, happily, to a more self-aware, deliberate study of how qualitative research is actually done—what a "qualitative design" looks like; how it affects data collection; how these data are aggregated, partitioned and analyzed; and how findings can be reported in a more easily digestible form than in the past. Judging from the demand, there is clearly a need for this kind of work, even though it often does not go beyond descriptive reporting and procedure-exchanging.

Such a demand, however, gives rise to some cause for alarm. It suggests that a lot of qualitative researchers are experiencing difficulty in the field—and that there is an insufficient corpus of reliable, valid or even minimally agreed-on procedures to rely upon. This is a curious state of affairs, since field studies have been in existence longer than experimental studies, and social anthropologists have written extensively about their experiences in the field, some with warnings and advice built directly into their subtitles (e.g., Wax, 1971).

Part of the problem seems to be that these writings are uneven, to put it charitably. There is typically a greater wealth of advice than is needed concerning gaining and maintaining access, and on avoiding obvious sources of bias, but a decided poverty concerning issues of data reduction and analysis (Sieber, 1976). Some sociologists and ethnographers have even hesitated to become involved in these issues, on the grounds that an unequivocal determination of the validity of their findings is not possible (e.g., Becker, 1958; Bruyn, 1966; Lofland, 1971). There has also been a somewhat magical belief in "bracketing", and in the intuitive insights generated by the experienced ethnographer as s/he progressively discerns clear classifications and an overarching pattern from the welter of field data, and does this in ways that are presumably irreducible or even incommunicable.

It should also be noted that many qualitative researchers have, on epistemological grounds, simply refused to enter this arena at all, which is, incidentally, another reason why basic methodological canons for qualitative research are so lacking. For many social phenomenologists and ethnomethodologists, and even for some social interactionists, there is no innate social reality to account for, and thus there is no need to evolve a robust set of methodological canons to help explicate its laws (see Dreitzel, 1970, pp. v–xvii). The social processes which can be determined are judged to be ephemeral, continuously fluid, or to have no existence independent of social actors' ways of accounting for and describing them. On this basis, whatever "causes" we may consider it possible to abstract from observed social relationships would be simply the fruit of our fertile imaginations. As Wittgenstein put it in the *Tractatus*, "causes are superstitions" [1], and the debate between causal idealists, such as Wittgenstein, and causal realists, who work in the logical positivist tradition, has continued to rage and shows no signs of subsiding in our own or our grandchildren's lifetimes (see Kim, 1981).

Finally, most of the serviceable work concerning validation and analysis issues done by social and cultural anthropologists does not clearly apply to the settings that educational field researchers operate in, or to the kinds of research they are now trying to do. Field research in exotic cultures does not transpose directly to field research in one's own culture. Studies of simple or primitive cultures are not necessarily conducted in the same way as studies of complex, multicultural social groups. Even more crucially, if it is desired to make stronger inferences or to generate policy, the familiar single case-study must give way to multisite case studies—which, speaking methodologically, represent a virtually new species of social-scientific research.

Scope of This Paper

Over the past several years, both of us have found it necessary to contend with many of the above shortcomings of qualitative research methodology, most notably the lack of general procedural canons or specific decision rules for analyzing and verifying data. Initially, we considered this state of affairs as creative and challenging; it enabled us to break new ground, to try out homegrown techniques and to mix psychometric with phenomenological procedures. However, we felt progressively less optimistic as we came to grips with problems of imprecise measurement, weak generalizability, vulnerability to several sources of bias, data overload with a high dross rate, underattention to manipulable variables in favor of contextual features, and labor-intensiveness. Writing later about this initial work revealed our ambivalence; Miles (1979) presented an article entitled "Qualitative data as an attractive nuisance", and Huberman (1981a) wrote one entitled "Splendors and miseries of qualitative research".

The present paper draws on this experience in two new studies, one just being completed and one in progress, illustrating the dilemmas faced and the solutions found. The aim is to be concrete and self-examining, and to stimulate similar activity among other qualitative analysts.

The first project considered is a multisite field study of the dissemination of educational innovations (Huberman and Miles, 1982), carried out as part of the *Study of Dissemination Efforts Supporting School Improvement* (Crandall et al., 1982). In the field study, we tried to learn from our previous errors by devising a battery of procedures for data collection and analysis that could overcome some of the more acute problems. Essentially, we aimed for more extensive data-processing in the initial stages and for greater homogeneity in the modes of data collection and reporting among field researchers. For example, we used a guiding conceptual model, a finite and focused set of research questions, a preliminary list of sensitizing codes, ongoing procedures for comparing data sets, and a common structured format for writing-up site reports. Further details are given shortly.

We then embarked on a follow-up study (Miles and Huberman, 1983) whose chief objective was to refine and make explicit the data-analysis methods used in the first. The subsequent study had four parts. First, it involved feeding back to informants at the twelve field sites (a) summaries of the principal findings, (b) an explanatory path model (we called it a "causal network") with an accompanying text explicating the key variables that had been identified at the site, and (c) a set of predictions concerning what we thought was likely to have happened by midway in the following year, in the light of the findings. We wanted to investigate whether our methods of analysis had, in fact, yielded findings that were plausible to respondents at the field sites (in general, it appeared that this was indeed the case).

The second task in the follow-up study was a substantive exploration of three topics, through secondary analysis of the original data. These topics were: users' and administrators' motives in adopting innovations, including career-related and other nonimmediate agendas; the process of user practice mastery; and the micropolitics of institutionalization.

The next—and core—component of the study focused on methods of analysis. The intention was to become vigorously self-aware of what we were doing. To this end, a monitoring mechanism was established that would register and document each of the steps taken in the course of the secondary analysis, from the initial formulation of a research question to the final write-up. Included in this methodological audit were the specific analysis task, the step-by-step procedures used during the analysis and the rationale for each one, the confidence in the conclusions, and a review of the strengths and weaknesses of the analytic procedures used. In the course of use, this self-auditing form has already undergone three iterations. It will probably be reshaped a few more times until it can serve as a workable device for other qualitative analysts to use as they work, and for secondary and meta-analysts to use in verifying reported findings (e.g., Guba, 1981). The current version of the documentation form is attached (see Appendix I).

The three tasks mentioned above lead into the fourth, the preparation of a methodological handbook for the analysis of qualitative data. It is hoped that the process of scrutinizing the methods used to analyze and report nonnumerical data sets will yield a corpus of procedures that are both manageable and robust. In addition to our own self-auditing exercises, we plan to draw on promising work done along the same lines by other qualitative methodologists (e.g., Guba and Lincoln, 1981; Firestone and

Dawson, 1981; Stearns et al., 1980; Smith, 1978; Glaser, 1978).

In preparing the handbook, we focused more narrowly on five analytic tasks which seem to bedevil analysts the most: coding data, integrating qualitative and quantitative data sets, data reduction, data display, and conclusion-drawing and verification. In the remainder of this article, we examine two of these, data reduction and data display, drawing on the techniques used in the dissemination study mentioned earlier, and on what has been learnt to date from using the self-auditing documentation form [2]. First, a short overview of the issues is given.

THE DATA-REDUCTION PROBLEM

Every qualitative analyst encounters the problem of data overload. The fact that field-study data are not usually translated immediately into numeric or alphanumeric form means that a lot of words accumulate in the course of data collection, increased still further by the analyst's concern for following serendipitous leads, confirming hunches and resolving puzzles—all of which may or may not be fruitful, but will certainly add bulk to the corpus of field notes. Typically, veterans begin reducing their data set early in the course of data collection, often via coding or interim site summaries, but they may still be left with 500-odd pages of transcribed field notes that have to be reduced just to allow preliminary analysis. This is, in effect, the nub of the problem: qualitative data need to be reduced for any analysis to occur, and the choice of a reduction strategy or heuristic will determine what kind of analysis is possible and will thus foreclose other kinds. Reducing data implies aggregating and partitioning them according to some decision rules that may be, at best, tentative or intuitive, but that always have important consequences. At worst, this procedure may lead to hasty, unfounded conclusions, or to shapeless data from which it is impossible to extract any meaning—something which may be discovered weeks later, when there is no time to go back over the 500 pages of field notes and reduce the material differently.

Reduction not only allows analysis, it *is* analysis, in that clusters and partitions will necessarily follow the analyst's evolving sense of how the data come together and how they address the research questions s/he wishes to answer. This leads to two general questions, of (1) how qualitative data can be reduced without unduly distorting or oversimplifying them, (2) and which reductive methods still leave room for a wide range of alternative analytic approaches, including a realistic possibility of reducing the same data set in ways other than those tried initially.

One part of the answer, we feel, lies in the way in which qualitative data are displayed. Displays in matrix or graphic form are a major improvement over narrative text, with several important functions. First, they help a reader make sense of a large data set, which might otherwise be spread out over hundreds of dispersed pages, and retrieved only selectively by the analyst. Such charts can also do the basic work of laying out main findings for the reader, leaving the text to provide illustrations and qualifications. However, displays are helpful in other ways as well.

First, displays can facilitate cross-site analysis. Using common formatting techniques in reporting findings at the site level makes it possible to consider sites economically side by side. A common graph, for example, already "standardizes" the data to a common "metric" that allows overlaying of one set of results on another, to reveal common trends and "distances" between site configurations. In other words, displays reduce batches of data in ways that enhance their comparability, and allow hypothesis generation for further analysis. Graphic or matrix displays are a way of getting the trees located in the forest in such a way as to see not only what the forest (or parts of it) looks like, but also how it would look if the trees were moved around. Looking at data in this way, as illustrated below, generates analysis: it actually tells the qualitative researcher what the next, most likely analysis step should be, and how s/he will know whether it has worked—whether it can be verified or corroborated. The result is something like "grounded data-manipulation": continuing iterative cycles of data reduction, display, analysis, new reductions, new displays and new analyses, until a full and coherent set of meanings has been generated from the data.

Data Base

Before illustrating some of the data-reducing and data-displaying devices used in the dissemination field study mentioned above, we now review the data base very briefly.

Table I shows the principal characteristics of the sample. Twelve field sites were drawn from 146 sites comprising a larger survey sample (Crandall et al., 1982), and two main dissemination strategies were chosen for attention: that of the *National Diffusion Network* (NDN), an outside-in delivery strategy, through which a nationwide pool of innovations is made available locally with assistance from the developer; and Title IV-C, a homegrown strategy supporting the local development of projects and their subsequent statewide diffusion. The regional and setting characteristics of the sample are shown,

TABLE I

Characteristics of Field-Study Sample

Site	Program sponsorship	U.S. Region	Setting	Year initiated	Status (as initially assessed)	Program type	Program name or initials *	Program content
Astoria	NDN	Southeast	Small city	1978	Expanding	Add-on	EPSF	Early childhood
Banestown	NDN	Southeast	Rural	1979	Expanding	Pull-out	Catch-up	Reading/mathematics
Burton	NDN	Midwest	Suburban	1979	Expanding	Add-on	IPLE	Law and government
Calston	NDN	Midwest	Center city	1978	Ongoing	Drop-in	Matteson 4D	Reading
Carson	IV-C	Plains	Rural	1977	Expanding	Add-on	IPA	Individualized educational planning
Dun Hollow	IV-C	Northeast	Urban sprawl	1977	Dwindling	Add-on	Eskimo studies	Social studies
Lido	NDN	Northeast	Rural	1976	Dwindling	Add-on	KARE	Environment
Masepa	NDN	Plains	Rural	1978	Ongoing	Drop-in	ECRI	Language/arts
Perry-Parkdale	NDN	Midwest	Suburban	1977	Ongoing	Subsystem	EBCE	Career education
Plummet	IV-C	Southwest	Center city	1976	Ongoing	Subsystem	Bentley Center	Alternative school
Proville	IV-C	Southwest	Urban sprawl	1977	Dwindling	Pull-out	CEP	Vocational education
Tindale	IV-C	Midwest	Urban sprawl	1976	Ongoing	Drop-in	Tindale reading model	Reading

* IV-C program names are pseudonyms, to avoid identifying specific sites.

TABLE II

Breakdown by Site of Data Collection and Methods

Site	Number of site visits	Total days on site	Number of phone contacts		Number of formal or informal interviews	Number of observations	Number of documents collected	Number of pages of transcribed field notes
			Interviews	Updating				
Astoria	7	$4\frac{1}{2}$	8	9	20	3	5	112
Banestown	4	9		6	47	7	8	408 *
Burton	3	6		3	31	8	20	148
Calston	3	7	3	5	16	2	2	134
Carson	3	8		15	70	17	81	364
Dun Hollow	3	6		8	27	6	20	170
Lido	3	5		2	22	5	13	92
Masepa	3	9		7	46	10	6	370
Perry-Parkdale	3	7		5	60	11	49	301
Plummet	4	11	1	6	24	8	22	231
Proville	4	8	1	5	28	0	10	158
Tindale	4	9		8	49	8	23	225

* Combined notes of two researchers.

together with the dates of initial implementation and current status, in Table I. The program types specify the forms of implementation of the projects on a local basis: substitution of new materials (drop-in), enrichment of the regular curriculum (add-on), provision of a locale for tutorial or project work (pull-out), or constitution of a new, independent program or institution (subsystem). The final two columns in Table I list the twelve project acronyms and their areas of focus.

Table II shows a breakdown by field site of the data-collection procedure [3]. In the course of the school year, a site was typically visited three or four times, for two or three days, with interim contact by telephone. The number of visits and total days on site varied with the proximity of the site, the number of informants, the complexity of the program, and the difficulty of eliciting credible accounts from site informants. Data were collected chiefly by means of interviews—usually multiple interviews with key informants, and single interviews with more-peripheral actors—using a common, semi-structured schedule across sites which covered the principal research questions. There were also informal talks (e.g., in empty classrooms, cars, cafes) that yielded valuable background information. Observations were typically unstructured, although each field researcher usually had specific things s/he was looking for. Similarly, whenever a document appeared significant, we asked to look at and/or copy it. Each document retrieved was abstracted on a document-analysis form, and the interview and observation notes were dictated in narrative form together with any pertinent analytical or methodological notes, and then transcribed. Finally, about midway in the school year, site-specific raw questionnaire and interview data became available, and were used as a verification device, as a source of new leads to follow and of puzzles to solve.

Data Reduction

Data reduction actually occurs throughout the entirety of any project involving qualitative data: during basic project design, during data collection itself, and during preliminary and final analysis.

DESIGN STAGE

Both of us having had problems in previous "inductive" efforts, we decided to begin with a relatively tight design. In part, this was making a virtue of necessity; we intended to interface with a survey-type data base (Crandall et al., 1982), and thus had to be clear where the same issues were being addressed in commensurable terms, in order to merge the two data sets appropriately. More generally, we found it essential to elaborate and obtain agreement among the field staff concerning an overarching conceptual framework, a set of research questions, and basic instrumentation. Doing this, of course, reduced the scope of the study substantially. For instance, the initial conceptual framework (see Appendix II) was longitudinal, considered particular contextual and school-specific variable sets, and construed the innovation process as a series of reciprocal interactions among users, innovation characteristics and institutional parameters—in short, a conflict model of school innovation, as opposed to a more rational or technological model.

From the conceptual framework, 34 research questions were generated, corresponding globally to the blocks of variables in the framework. Typically, a question had a descriptive, broad-scoped formulation. Here is an example: "What were the manifest and latent reasons of teachers and administrators that led to the choice of this new program?". There were then subquestions (e.g., one for latent reasons, with possible response domains including gains in power or prestige, career mobility, boredom reduction, peer contact, etc.); these were subsequently translated into probes on the interview schedule. Of course, no-one was asked about his/her latent motives, but researchers knew that they had to deliver an answer—from other informants, through informal chats, by well-supported inference—for all key informants at the site.

Finally, a semistructured, omnibus interview schedule keyed to the 34 research questions was generated, containing possible probes. Again, we opted for middle-range precision. Here are three sample items: "What were you doing before you came here?"; "What was this school like, before (the innovation) entered the picture?"; "When you first heard about (the innovation), what did you think it might be like to use it?". Interviewers were free to vary the order within the schedule or to put it aside when informants diverged in promising directions. Experience here was that field researchers reshaped the schedule according to their congenial interviewing styles.

Critics might argue that, by restricting the range of constructs, questions and instruments, other ways of looking at and capturing the phenomena under study were screened out. Of course: anticipatory data reduction was occurring. This should not cause undue concern. First, the framework, questions and probes were general and middle-range. Next, it would have been hidebound to ignore the value of existing empirical and conceptual work as an orienting frame. Third, we expected to make—and did make—changes in the conceptual framework as the data proved it incomplete, unbalanced, or, in some parts, overbuilt. We began with conceptual consensus among a team with various social-scientific persuasions (psychology, social psychology, education, anthropology), and then allowed each discipline to inform the others in the course of data collection about more compelling or promising ways to look at the phenomena all were studying. Furthermore, the field-study instrumentation was designed to be redesigned as a function of emerging trends, thereby remaining data-sensitive and nonfrozen.

In summary, the data-reduction issue in the gearing-up phase was addressed by specifying the boundaries of data collection, and by standardizing the initial data-collection procedures to allow the production of reasonably comparable data sets across the twelve field sites.

FIELDWORK

In a nutshell, data reduction during fieldwork was dealt with by collecting less data than field studies usually do and by analyzing continuously the data collected. This ongoing analysis indicated where to focus and probe discriminatingly in future data collection, and gave a continuous reading of which research questions were still unanswered.

In conducting these cycles of data collection and analysis, nine devices were used: coding, policing, dictating field notes, "connoisseurship", progressive focusing and funneling, interim site summaries, external critiquing, memoing, and outlining. We now consider each one briefly.

Coding

We started with about 85 sensitizing codes, all derived from the 34 research questions. For example, a user's initial assessment of an innovation was coded as SIZUP, and changes made in the innovation during use were coded INNMOD. After the first month of fieldwork, the number of codes was expanded to 104, and these remained intact throughout data collection. All codes were defined operationally, and initial field notes were coded separately by two analysts, and then compared until adequate agreement was obtained.

With a view to subsequent analysis, we adopted three additional types of codes—actually, metacodes. One type, anticipating single-site analysis, signaled leitmotifs or recurrent patterns, together with puzzles or apparent contradictions. A second type built towards cross-site analysis by adding to these leitmotif codes a suffix (OS) to indicate that a field researcher had unearthed a pattern s/he had also found in notes from some other site. Finally, there was a code to identify what seemed to be a causal link (CL-EXPL) that could account for a recurrent pattern, and another (SITECL-EXPL) whenever site informants gave their own explanation of why things had turned out as they did.

Coding was onerous, but useful. Not only did it render the data appropriate for single-site analysis, but it also guided ongoing data collection. The strategy—sometimes obviated when there was a coding backlog—was to code the previous set of field notes before the next trip to the field sites. This line-by-line coding had the merits suggested by Glaser (1978): it revealed gaps and puzzles, identified core themes, illuminated theoretical components, uncovered potential sources of bias, and, overall, set the agenda for the next field visit. Had coding been done only at the end of data collection, the ongoing analysis would have been less sharp and would have resulted in data that were incomplete or equivocal, with no opportunity to obtain resolution.

Policing

Perhaps this term is too strong, but it implies more than "monitoring". First, the four field researchers worked in teams (with a senior researcher overseeing a junior researcher). Also, the two senior researchers read each others' field notes. The idea was not only to uncover instances of possible bias, but also for the researchers to keep one another on track. There was a real danger of an analyst's wandering off tangentially or rummaging around happily and blindly in his or her area of expertise. Of course, some of this is endemic to ethnographies. A field site is seamless; everything is related to everything else, and everything has meaning, from where people sit at lunch to how they introduce researchers to other informants. After an initial wave of site visits that accumulated every promising piece of data into an alarming amount of field notes, strict conventions were established governing the level of detail that could be tolerated. A datum would be registered only if it addressed either a specific research question or the relationship between two or more questions. Second readers then worked to keep their partners' accounts lean. The degree of leanness was hard to judge; it was estimated according to the fact that transcribed field notes under 200 pages were thin on core issues, and sometimes had gaps, whereas those over 300 yielded more information than could be processed.

There was also self-policing, as another vehicle for data reduction to facilitate subsequent analysis. Each fieldworker kept some form of book-keeping ledger, with an indication of which research questions had been answered satisfactorily for which informants. This helped (a) to avoid data-collection redundancy, and (b) to specify which data needed to be completed on the next site visit.

Dictating field notes

We chose, as a general rule, to dictate notes from interviews and observations, and not to take verbatim tapes. This reduced the data set substantially. It also brought to awareness self-reflective issues or questions that were unclear in the fieldworker's mind and that could be, literally, talked out and transcribed (such issues were marked off in the text with double parentheses). These issues or questions, moreover, were data-condensing; they usually united strands of information into meaningful or synthesizing units of analysis. In order to protect against selective note-taking, we had each member of a team taking separate notes on the first field visit, and then checked for agreement between the two transcripts.

Connoisseurship

It is hard to avoid immodesty here. We reasoned that people who knew their way around schools, could relate easily to school people, and who knew the research literature concerning knowledge dissemination and use would be less easily misled and distracted, and could keep the dross rate down—all ways of keeping the data base manageable. This proved true. We found, in fact, that many informants were laconic or elusive until they had decided that the fieldworkers were aware of school issues. The second-reader method helped to correct for arrogance.

Progressive focusing and funneling

Qualitative researchers are discovering that procedures for protecting reliability and heightening validity can be as rigorous—in their own terms—as the canons of classical test theory (see especially Guba and Lincoln, 1981). Since we have illustrated this thesis elsewhere (Huberman and Crandall, 1982), we shall be brief here. The main point of emphasis is that iterative procedures are needed: data are collected, coded, analyzed, and then new data collected as a function of that analysis—until, after several such cycles, a final account is at hand that is plausible, internally consistent and verified by recourse to multiple sources of corroboration. We drew mostly on two methodologies: (a) grounded theory/theoretical sensitivity (Glaser and Strauss, 1967; Glaser, 1978); and (b) investigative social research (Douglas, 1976). When combined, these constitute an approach which could be described as detective work done by intellectuals. Anthropologists have sometimes described similar techniques of analytic induction (e.g., Lindesmith, 1947; Turner, 1953), but this documentation is usually sketchy and sometimes defensive. More generally, philosophers of science (e.g., Hesse, 1974) have pointed out that the rules for verifying the results of analytic induction are nowhere nearly as well formulated as those of, say, propositional logic or statistical inference.

Our inductions had the desired effect of limiting and channeling the next round of data collection; we were examining progressively fewer elements in more detail, notably the elements we would be concentrating on during case-study write-ups and cross-site analyses (see below). So, essentially, the drafting stage was reached with less data—and more of it analyzed already—than we suspect qualitative researchers often have.

Interim site summaries

Here again, it seemed that every exercise obliging us to collate and compare what had been collected to date was a beneficial, if stressful, means of focusing down onto the essential questions to be addressed. Nonetheless, there was still room for new findings or insights to emerge that we had neither thought nor dreamed of.

These insights often emerged when the data were summarized and pooled, thereby guiding the next set of visits to the field sites. In these 20-page summaries, written about halfway through the school year, each researcher reviewed preliminary findings, audited the research questions not yet or not well addressed, and commented on the verisimilitude of his/her data. The exercise took precious time from ongoing data collection, but ultimately saved more time in subsequent analysis. By way of doing preparatory work for cross-site analysis, we added to this exercise another one: two external readers, one working on another part of the dissemination study and another with substantive expertise but no connection to the study, critiqued and synthesized the material from the interim site summaries.

Memoing

Following Glaser (1978), we wrote periodic memos to one another about issues uncovered at more than one site. For example, one memo concerning the career trajectories of users and administrators led to greater data collection on this theme, and ultimately to a key subsection of the cross-site analysis dealing with implementation motives. For the most part the memoing was episodic and superficial, but the device nonetheless has potential value as a cognitive activator, and as a stimulus to sharper focusing of data collection and to more-differentiated ongoing analysis.

Outlining

All the data-reduction methods used so far led naturally and rapidly to the specification of a standard writing outline for site-level case reports. The outline dealt systematically with each research question, and specified the data displays (see below) associated with each. The fact that these outlines were drafted and iterated to stability before final data collection had been completed also made for additional focusing; it guided the last part of the field work.

Data Displays

It has been argued that the several steps outlined above bring the analyst into the final write-up phase with leaner, better-analyzed data than usual. So far, so good. However, for each site about 250 pages of coded field notes still result, plus a few dozen document analysis forms and a large pile of raw questionnaire and interview forms.

The key question is that of how this material can be reduced quickly and without too much loss of relevant detail. Overall, the answer adopted here stressed standardized formatting, with a strong emphasis on graphic and matrix displays. We first summarize and illustrate these methods.

DISPLAYS FOR WITHIN-SITE ANALYSIS

As noted above, a detailed site-report outline was developed with standard formats for aggregating and partitioning blocks of codes that fell into a single research question or set of questions. This was the principal working tool for the twelve site reports. It had the merit of assuring cross-site comparability and the drawback of specifying in advance how data within a given research question were to be reduced and displayed. The drawback is that the site data could easily have been collapsed into other aggregates and partitions that were analytically satisfactory or compelling—but this would have varied among sites and reduced cross-site comparability. As it happened, the formats chosen were plausible and fruitful, probably because they were created near the end of data collection, and were more data-sensitive than an earlier iteration would have been. Data were displayed in two general modes: matrices and figures.

Matrices

Some of these were descriptive (e.g., a three-way display of types of assistance (ongoing, event-linked) by phase (initial, first year, second year) and source (building administrator, peers)), or could be organized around a checklist (e.g., a list of implementation-readiness criteria, or an index of institutionalization with judgments and qualifications in the cells). Table III displays part of a checklist matrix for judging preimplementation readiness, for early and later innovation users and administrators. The appeal of the matrix, beyond the fact that it allows an analyst to draw preliminary conclusions (such as the modest improvement over time shown in Table III), is that the twelve sites can readily be compared by pulling out the twelve sheets from the individual case reports, together with the two or three pages of text in each report that provide further context, qualifications, illustrations and a summary. Checklist matrices can also be keyed easily to survey

TABLE III

Excerpt from Checklist Matrix (Masepa Site)

Conditions	*Presence of supporting conditions*			
	Early users ($n = 2$) 1977–78	Later users ($n = 6$) 1978–80	Administrators 1977–78	Administrators 1978–80
Commitment	*Moderate – high* Bayies: "I was committed". Quint: "I don't know why I stayed with it".	*Low – moderate* "I had no choice" "I was excited about it". "Why not?".	*Building level: low* "I wasn't that sold on it". *Central office: high*	*Building level high* "Pushing it hard". *Central Office: high*
Understanding	*Poor* "Didn't know what was going on". "All very confusing".	*Poor – Fair* "Didn't know what it was all about". "Understandable". "Not bad, but not how it all fits together".	*Building level: poor* "Little jumbled". *Central office: poor – fair* "I don't know anything about reading".	*Building level: good* Trained and experienced *Central office: poor* Delegated to trainers
Materials	*Absent* "Nothing was ready". "Had to do it all myself".	*Excellent* "Emily Boud just marveled at that". "Everything was ready".	not applicable	not applicable
Preparatory training	*Poor* "I should have had more training". "They really only knew what they had done themselves".	*Present but insufficient* "Short and strenuous". "I was over-whelmed—too much thrown at me". "Better than I had thought before starting in. It worked".	*None*	*Building level: sufficient* Short, intensive training at demonstration site

data, which is indeed what was done with the readiness indices [4]. Five of the indices on the full readiness checklist replicate corresponding questionnaire items. The case-study material can thus provide a validity check, add new empirical factors, extend and differentiate the survey measures by giving varying weight to the factors, and illustrate the survey findings with citations and examples.

We also used progressive matrices, charting shifts over time in the levels of users' practice mastery, changes in each innovation, organizational changes, and sequences of technical assistance. As an illustration, Table IV shows an excerpt from a site report using a format for tracking changes in innovations over time. The list of program components is taken directly from the survey component of the larger study, so that, here again, the survey measures of fidelity of implementation can be matched with field-study measures. Note

TABLE IV

Illustration of Progressive Matrix (Carson Site): Changes in IPA (individualized planning) Innovation

Program components	First implementation 1977–78	Later implementation 1978–79	Later implementation 1979–80
Education plans	Sample forms developed	Plans kept in files; form simplified somewhat for next year	Monitoring of teachers' educational plans minimal; educational plans discussed at nine-weeks conference, rather than at start of year; some coordination with special-education profiles; little specification of "long-range goals" by elementary teachers; some plans minimal, nonresponsive to student interests or individual purposes; some teachers not using plan forms; material prepared on "how to write a plan"
Frequency of conferences		Fewer home conferences	Home conferences discontinued; directions for conferences standardized (manual): high school, reduced to two conferences including parent and two with student alone; elementary, reduced to two three-way conferences
Student profiles	Modified; color-coded	Addition of more affective testing (school attitude, self-	First full filling-out of interest inventory; dropped-student attitude interview; profile data incompletely filled in; profile used casually or partly by

TABLE IV (continued)

Program components	First implementation 1977–78	Later implementation 1978–79	Later implementation 1979–80
		concept, etc.)	teachers: ambiguity about which tests to be included for next year (in absence of validation requirement); use of profile standardized (manual); profile simplified, some information added: triplicate for parent, teacher, profile file (year-end changes); some coordination with special-education individual plans
Management team		Fewer meetings (less than weekly)	Fewer meetings (every three to four weeks)
Parent involvement	Clarification: parent should help with activities, not leave to teacher		"Family learning nights" held less frequently than planned; some parents do not take part in conferences
Administrative team (coordinator and aide)		Added full-time aide; field-trip coordination more centralized by coordinator	Coordinator does pull-out "batching" of students in special activities, more for elementary, especially by kindergarten and grades 1 and 2(?)
Formative evaluation			Some complaints about "too many questionnaires"; data returns usually incomplete
Teacher needs assessment (in-service)		Special in-service supplied for newcomers to program	Connection to individual plans sometimes vague or not visible to teachers; in-service reduced from hopes; in-service committee recommendations to be channeled through management team; orientation materials prepared (for teachers, students, parents)
Parent advisory committee		Committee not separately formed; used existing high-school Title I advisory committee	Committee met only twice

TABLE IV (continued)

Program components	First implementation 1977–78	Later implementation 1978–79	Later implementation 1979–80
Community component (citizens help with activities)		Community resource-book delayed; used little by teachers	Only 25–30% of parents help with activities
Other aspects			
Purposes of program		In high school, moved toward advisory, class-scheduling emphasis; at elementary level, "don't push into career side"	Closely tied to "advisor" system in high school; defined more as "enrichment", add-on, interest-focused program in elementary school
"Match-ups" of students and teachers	Some switching of advisees among teachers	Discontinuance of multigrade match-ups, especially in elementary schools	Students at high school who wish it may change advisors
Numbers involved	14 teachers, 4 administrators, 60 students	20 teachers, 4 administrators, 83(?) (120?) students	47 teachers, 686 students
Entry to program		Voluntary, but some pressure on teachers to take part	Required for all teachers (less motivation); required for all students; wider socio-economic range, lower-capability students included; resource teachers' participation dropped
Teacher time investment per student			Reduced by one-half to one-third (10–30 instead of 60 min per conference); less intense, more "diluted"
Student activities		Rarer to spend weekend time with students	Sometimes omitted by teachers; often carried out in "batch" or group form (e.g., unit on dinosaurs); more frequently done during school day (not nights, weekends)

TABLE V

Illustration of Progressive Matrix, Showing Effects of a Stimulus (Masepa Site): Ongoing Assistance of Various Types

Location	Users' assessment	Types of assistance provided	Short-run effects (users' "states")	Longer-run consequences	Researchers' explanations
Building administration	+ +	1. Authorizes changes	Relieves pressure, encourages	Users are helped administratively and substantively, feel obliged to do ECRI with minor adaptations	Administrative authority, servicing, availability and flexibility lead to sustained, faithful implementation of model
	+ + –	2. Eases schedules, assigns aides	Helps early implementation		
	+	3. Controls fidelity	Feeling policed		
		4. Consults, offers solutions	Feeling backed-up, substantially helped		
Central office administration	+	1. Promotes elementary curriculum for reading instruction	Pressures non-users	Program is perceived as supported, assisted, "protected" by central office	Central Office able to push program and answer requests, yet not perceived as main actor by users
	+ +	2. Answers building administration and trainers' requests	Building administrators have material, administrative support		
Helping teacher	+ +	1. Provides materials	Reduces effort, increases repertoire	New and experienced users receive systematic instruction, follow-up, materials; stay with program and are careful about making changes in it	Personalized in-service mechanism, with both training and assistance, allows for mastery and spread of ECRI in "faithful" format
	+	2. Demonstrates, models	Trains, facilitates use		
	+ +	3. Answers requests	Problems solved rapidly		
	+ +	4. Encourages	Maintains level of effort		
	+ –	5. Circulates, controls	Ambivalent: helped yet coerced		

User–Helping teacher meetings	+ +	1. Comparing practices with others	Encourages, regulates	Creates reference group, gives users a voice, solves ongoing problems and lowers anxiety	Multi-purpose forum which consolidates use and users, defuses opposition
	+	2. Debugging, complaining	Cathartic, solves short-run problems		
	+	3. Learning about new parts	Expands repertoire		
	+−	4. Encouragement	Gets through rough moments		
Teacher users in other schools; target school	+	1. Sharing materials	Increases stock	Increases commitment, regulates use (decreases deviance)	Additional source of assistance, which increases as number of users grows
	+	2. Exchanging tips, solutions	New ideas, practices: problems solved		
	+	3. Comparing, encouraging	Motivates, stimulates		
Trainers in target school, other schools	+ +	1. Tips for presentations	Facilitates practice	Reliable, unthreatening back-up provided in school	Elaborate and effective lateral network: trainers seen as peers
	+ +	2. Solution to short-term problems	Helps expand beyond core format		
	+	3. Encourages	Maintains effort		
	+	4. Serves as successful model	Stimulates		

[a] + +, very effective; +, effective; + −, mixed response; −, ineffective.

also the list of other aspects, indicating that the field researcher also found core components of local significance other than those highlighted by the program developer.

Another type of progressive matrix considered longitudinally the effects of a stimulus or sets of stimuli. Table V shows the locations, types, assessments and consequences of the assistance provided to users at a field site, together with an analytic commentary by the site researcher. At the site-report level, this helps the analysts to obtain an overall view of assistance and its impact without missing details, to contrast positively assessed with negatively assessed assistance, and to array these data according to the source or locus of aid.

Other matrices used were even more directly causal, displaying the connection between assorted states or processes, and outcomes. Table VI shows how data were displayed that bear on the question of what led to changes in the organizational structure and functioning of a site.

Still another type of summarizing matrix arrayed conceptually clustered data, usually within a set of research questions: Table VII is an illustration. It combines informant-given motives for adopting an innovation (as distinct from researcher-inferred motives), responses given or inferences made about career plans implicated in the new project, informants' judgments of the centrality or relative importance of the innovation in their daily life—how large it loomed—and their initial attitudes towards the project. Here again, motive categories and ordinal or dichotomous scales—involving data transformations on the part of the researcher—readied each of the twelve charts for cross-site analysis, while at the same time enabling the site analyst to explore the relationships among these variables.

Figures

In addition to matrices, we used commonly formatted graphic displays for within-site and cross-site analysis. These included:

profiles of the growth or decline of innovations over time as a function of the numbers of users and numbers of units concerned (we called these "growth gradients");

organization charts with entries to show authority lines, informal relationships, attitudes towards an innovation, year of adoption (for teachers), informants' age and relationship to the new practice (user, supporter or adversary, innovation champion); such charts were also used to map flows of assistance during innovation implementation, including the direction, type, and assessment by recipients (Fig. 1 shows an example);

event-state flowcharts, which assembled the key events during the life of the project, identified their system-state effects, and interrelated the entire set to show how these factors evolved mutually over time (an excerpt appears in Fig. 2); and

TABLE VI

Illustration of Causal Matrix (Carson Site): Innovation as a Force for Organizational Change

Strains or difficulties created	Underlying issues (as seen by author)	How coped with	How resolved: types of resulting change [a]
Conflicting expectations: should parents or teachers do activities?	Work load; parent–teacher role conflict	"Explaining" that teachers could not take primary responsibility for out-of-school activities	Increased use of "batched" activities, many set up by coordinator (P)
View that forms and procedures were "extra", overloading	Work load; autonomy, resistance to control	In-service assistance	Repeated revision and simplification of forms and procedures; production of operating manual; reduction of expectations (no home visits, fewer conferences) (P)
User uncertainty and resistance to use	Autonomy	In-service assistance; management-team interviews of all staff	See above; also, creation of in-service committee (S), with coordination through management team (P)
Extra time requirements	Work load	Initially, via volunteerism, high commitment	Use of substitutes (P); dismissal of school during conference days (P); reduction of expectations (above)
Program is complex, demanding, externally funded	Authority, coordination, accountability	Early creation of management team, addition of elementary teachers	Institutionalization of management team (S); heightened expectations for teacher-upward influence (C); lowered morale when expectations violated (C)
Enthusiasm of "advocate" teachers led to peer criticism	Autonomy	Quieting conflicts through informal discussion; informal coordination and referral	Norms supporting flexibility and colleague influence within schools, and cross-school interaction (C); increased linkage and closer interaction between schools (C); hobby day (P)

[a] S, structural change; P, procedural change; C, climate change.

TABLE VII

Illustration of Conceptually Clustered Matrix (Masepa Site): Motives and Attitudes of Users, Nonusers and Administrators

User	Motives	Career relevance	Centrality	Initial attitude towards program
Early users 1977–78				
R. Quint	*Self-improvement*: "To get better, I had to change"… "Maybe I wasn't teaching the best ways". *Pressure*: "They wanted us to do it". *Social influence*: "Everybody was saying what Gail's doing is great".	None—improvement of practice	*High*: "Biggest thing I've ever done that somebody else told me to do".	*Neutral*: "There wasn't any appeal. They said it worked so I was going to try it".
L. Bayeis	*Observation*: saw G. Norrist do it and "was impressed". *Fit to personal style*: "I like structure". *Practice improvement*: "Looking around for a different way to teach reading". *Novelty*: "You get tired of always doing the same old thing".	Vehicle to turnkey trainer role; also became Title I coordinator	*High*: "Most important thing I've been involved with".	*Favorable*
Second generation 1978–79				
F. Morelly	*Social influence*: heard from several friends about program.	None—possibly stabilizing job at the school	*High*: "This is the only new thing I've done since I've been out of	*Neutral*: apprehensive

	Opportunity, effort justification: "I took the training for recertification credit. After all that, I had to follow through". *Pressure*: "He [Van Whye] is the reason we do it here. He's so enthusiastic about it".		school...I had to invest so much".	
L. Brent	*Social opinion, influence*: "I heard how good it was". *Pressure*: "[Van Whye] was really sold on it. They really want it in". *Conformity*: most doing it or planning to in the school; "It's what's coming". *Self-improvement*: occasion to "keep growing".	None—possibly fear	*High*: "It's been a nightmare".	*Unfavorable* (once training began)
Recent users 1979–80				
V. Sharpert	*Obligation*: requirement to obtain teaching post; "I didn't have a choice". *Practice improvement*: complementing preservice training.	Route to teaching job in district	*High*: "My first job".	*Neutral*: apprehensive

TABLE VII (continued)

User	Motives	Career relevance	Centrality	Initial attitude towards program
A. Olkin	*Social influence*: "Heard it was good";..."a good friend liked it". *Pressure*: "Strongly encouraged" by Wrelling and Danloff. *Observation, modeling*: saw A. Norrist; "She really impressed me".	None—felt obligated by administration	*High*: "This was really *the* big one for me".	*Neutral*: mixed feelings
S. Sorels	*Observation*: "It was so good for my own kids... tremendous change in reading, spelling, work habits".	Route to full-time teaching position	*High*: "This was really a big step for me—a big move...[nothing else] as high as this in my career".	*Favorable*: "I was excited about it".
Nonuser				
C. Shinder	*Relative disadvantage*: "My program was better". *Poor fit with personal style*: "too scholastic...too programmed".	None	—	*Unfavorable*

Administrators				
K. Weelling (Principal)	*Met need*: "I was looking for a highly structured, skill-oriented reading program". *Novelty, promise of practical improvement*: intrigued by reading about mastery learning; wanted to see it in operation.	None at first—later, appreciates visibility	*High*: "Largest investment I've ever made".	*Neutral*, then *favorable*
J. Danloff (Curriculum Coordinator)	*Relative advantage, face validity of program*: "Well organized"; could be used for other subject matter. *Social influence*: "impressed" that outstanding teachers favored the program. *Practice improvement*: beginning teachers ill-prepared in reading; "We didn't know what to do with them... They just had to learn on the job".	Another in a series of implementations	*Moderate*: "It was one thing among a lot of things I was working on"	*Favorable*
W. Paisly (Assistant Superintendent)	*Social influence*: "Talked into it" by J. Danloff	None	*Low*: "It was no big deal...".	*Neutral*

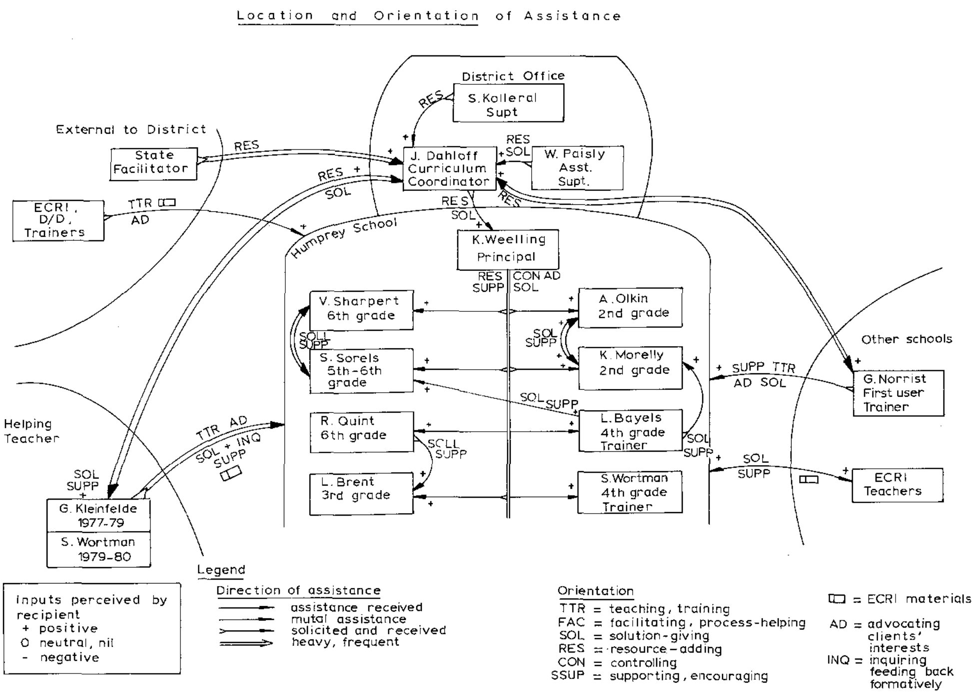

Fig. 1. Illustration of Organization Chart with Assistance Flows (Masepa Site).

causal networks that emerged from the flowcharts and represented the interrelationships among a set of core variables identified for all twelve sites.

Concerning this last device, the causal network, we were searching for a tool to summarize and interrelate the full set of data. The best vehicle seemed to be the themes, patterns, and contrasts across the twelve sites that the field researchers were coding progressively as they came across phenomena reported by others. Building from here, we worked up a set of some 30 core variables that were salient and influential at all sites. The list contained three sets of temporally related variables: antecedents, mediators and outcomes [5].

For each report, a field researcher drew a flowchart (see Fig. 3) mapping the progression and interrelationships of these variables. Hypothesized causal influence was represented by arrows between variables whose magnitude was set grossly at high, moderate or low. A two- or three-page text summarized the relationships, helping both to keep us honest and to aid reading by others. This device was an economical yet appropriately complex summarizing tool. It allowed validity checks via critique and correction by a second researcher, and via feedback at the field sites (informants could, and did,

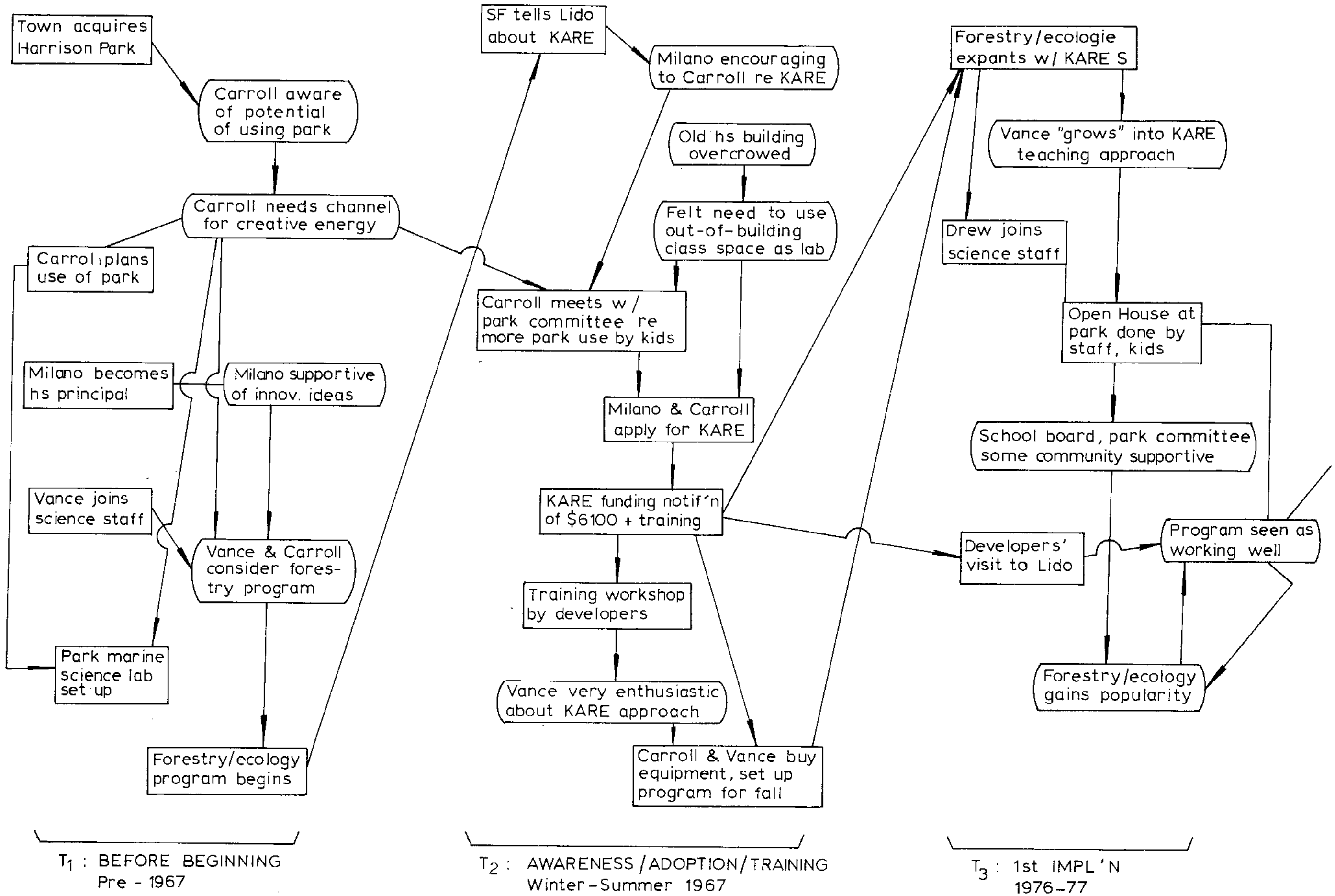

Fig. 2. Excerpt from Event-State Network (Lido Site). Boxes represent Specific Events; ovals represent System States.

TABLE VIII

Illustration of Predictor–Outcome Matrix: Contextual Features as Predictors of Organizational Change [a]

Site	Environmental pressure	Demographic changes	Needy student population	District innovative history: District	District innovative history: Building	Superintendent orientation (Style, attitude towards innovation)
Innovation in place, organizational changes						
Carson (IV-C)	Moderate	More professionals	—	Past, low; recent, high	Moderate–high	Reformist *
Masepa (NDN)	—	—	Many Title I children	High, proactive	Moderate	Superintendent, distant; asst. superintendent, active
Innovation in place, few organizational changes						
Plummet (IV-C)	High	More inner-city students	Hard-core offenders, dropouts	Moderate: need for improvement felt	(not applicable)	MD *
Perry-Parkdale (NDN)	Low–moderate	—	—	Moderate—high	One high; one moderate	Entrepreneurial *
Tindale (IV-C)	—	More low-ability students	Many low-ability students	Low	Low	MD *
Banestown (NDN)	Moderate	—	30% Title I population	Low	Low	Weak–distant *†
Astoria (NDN)	—	—	—	Low	Low	MD *

School board attitude	External fund-seeking history	Salience/ scale of funds	Motivation for adoption	Central office advocates	Climate of school
Supportive, progressive	Little done *	Low ($96 000)	Problem-solving	Present (superintendent)	Collaboration, cohesiveness, tolerance
Supportive, progressive	Successful *	Low or none (30–50 000) *	Problem-solving	Present	Collaboration, cohesiveness, tolerance
Supportive progressive	MD	High ($300 000) *	Problem-solving	Present	(not applicable)
Supportive (MD)	Active, opportunistic, successful	High ($300 000) *	Opportunism; some problem-solving	Present	One, collaboration; one, isolation
Supportive, middle-of-road	Selective, cautious about federal funds	Low–moderate ($87 000)	Problem-solving	Present	Isolation (MD)
Conflictful, progressive/ conservative mix	Little done currently *	Low–moderate ($5000–6000)	Problem-solving	Present	Isolation, some diversity, tolerance
(MD)	(ineligible for public funds except Title I)	None	Problem-solving	Present	Collaboration, cohesiveness

TABLE VIII (continued)

Site	Environmental pressure	Demographic changes	Needy student population	District innovative history		Superintendent orientation (Style, attitude towards innovation
				District	Building	
Limited use of innovation, no organizational changes						
Calston (NDN)	—	More mixed population, more Latin Americans	More students requiring individualization	Moderate	Moderate –low	MD
Lido (NDN)	—	—	—	Low	Low–moderate	"Lame duck", restrictive †
Burton (NDN)	—	—	—	Past, low; recent, moderate –high	Low	Active, change-oriented *
Dun Hollow (IV-C)	—	—	—	High	Moderate	MD *
Proville (IV-C)	Low–moderate	—	—	Moderate –high	MD	Power-oriented *

redraw, eliminate or add arrows and revise the narrative—although, gratifyingly, their changes were relatively minor). The display was grounded in the data, not spread over the site reports. The network display mode prepared and facilitated cross-site analysis. Finally, it served as a good analogue to the modeling done using the survey data.

Path-analytic devotees might point out that the above causal networks are nonrobust, contain only gross estimates of variable magnitudes, no path coefficients, are overidentified, etc. Perhaps the key here is not to compare these networks with path models, but to see them as outcrops of a progressive analysis during which relationships are deduced and mapped gradually, building from a field researcher's evolving sense of causal influence, but amenable to critical review by a second analyst. Note that this is essentially what happens during secondary analyses of survey-based data.

Buard attitude	External fund-seeking history	Salience/ scale of funds	Motivation for adoption	Central office advocates	Climate of school
MD	Active, successful	"High" (in-service, materials provided)	Problem-solving	Present	Isolation
Active, not always supportive (MD)	Selective, cautious about federal funds	High ($61 000)	Opportunism/ problem-solving mix	Absent	Isolation *
Supportive progressive	MD	High ($31 000)	Problem-solving	Present	Collaboration (MD)
Indifferent, traditional	MD	None	Opportunism	Absent	Isolation? * (MD)
Conservative (MD)	Active, opportunistic, successful	Low originally, then moderate–high	Opportunism	Present	MD

[a] Items marked with an asterisk (*) were inferred or estimated from the site report; MD denotes missing data; a dagger (†) denotes lack of salience in local context.

Displays for Cross-Site Analysis

Up to this point, the data have undergone two analysis and reduction cycles, the first during the iterative process of analytic induction in the data-collection phase, the second at the time of the site-specific write-ups. At this second stage, coded data have been assembled, analyzed in conformity with the set of research questions, reduced in a commonly formatted tabular or graphic form, reanalyzed and summarized in a few pages of accompanying text. We thus reach the stage of cross-site analysis with twelve sets of summarizing displays and a few pages of text for each. Of course, the cross-site analyst also has the case reports (ranging from 80 to 150 pages) for ready reference.

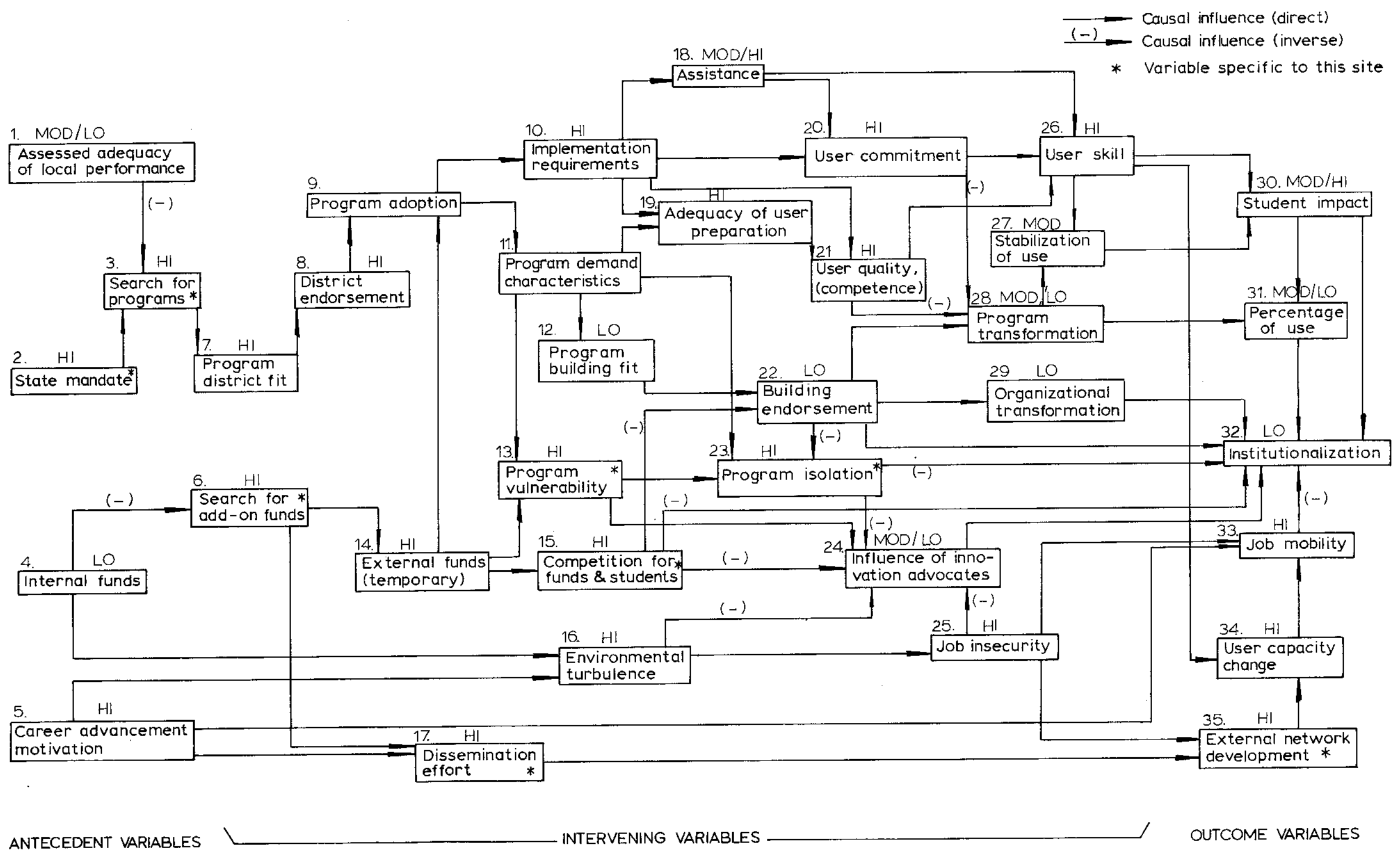

Fig. 3. Illustration of a Causal Network (Perry-Parkdale CARED Program).

We then used two display-analytic general procedures in virtually all sections of the cross-site analysis. First, a descriptive metamatrix was assembled that displayed the data from all twelve sites on a theme including one or more research questions. Sometimes this began mechanically by covering an office wall with twelve matrices or figures for a rapid scan. The descriptive metamatrix was then produced, usually with a blend of text excerpts, core sentences and descriptors for each site in each cell, sometimes with ratings (high, moderate, low, etc.) (for an example, see Table X).

Then began the next cycle of analysis and reduction. In many cases, this entailed regrouping the sites by magnitude of an outcome measure (e.g., more or less practice change), then arraying this scale against a series of predictors. Table VIII shows such a predictor–outcome matrix for analyzing the influence of contexual factors on the degree of organizational change. It is important to note that the analyst selects the predictors, drawing from those (a) already on a site-specific summary matrix or figure, (b) not on a summarizing display but identified in several site reports, or (c) abstracted from the metamatrix. From here on, the shape of the data usually guides subsequent analysis. In general, however, we tended to move towards an *R* analysis, by examining associations among variables and factoring them into fewer predictors, and/or towards a *Q* analysis, by looking for clusters or families of sites.

Obviously, it is not straightforward to claim that this last series of analytic procedures will produce valid meaning. For quantitative data, first- and second-order factoring, choice and formulation of predictors, configural scoring and cluster analysis are all data manipulations for which there are agreed-on algorithms. If it is claimed to perform similar operations without numbers, or with nothing more than dummy variables and two- or three-point ordinal scales, how plausible is the claim, and how credible are the results? Or are we talking about analogues of statistical analysis, of the type that Barton and Lazarsfeld (1955) initially called "quasistatistics" and that presumably conform to "the logical structure of quantitative research" (p. 348)—whatever that is? Or perhaps these are just metaphors that qualitative analysts use to express the types of bivariate and multivariate operations engaged in by any inductive analyst when making attributions and estimates —mentally registering the numbers of times things happen together or apart, or happen only when still other things are present, by processing information cognitively in a sort of mental covariance analysis. Cognitive psychologists (e.g., Kelley, 1973; Nisbett and Ross, 1980) are beginning to understand both the power and the shortcomings of these intuitive procedures; this can help to determine how robust they are on logical or statistical grounds.

These are important questions that need answering, and our current work, most notably the self-documenting log, is designed to obtain some preliminary closure. At this stage descriptive answers are needed. What do qualitative researchers actually *do* when they reduce and analyze data? For example, how did Stearns et al. (1980) progress from case studies to "propositions" to a "site factor matrix", and what kind of matrix is being considered? How did Stake and Easley (1978) reduce a gargantuan data set to a small number of "issues" and "problems", and what is the veridical status, inferentially speaking, of those issues and problems? The only way to understand such processes is to exhibit them for analysis and critique, and possibly for replication—rather than alluding to them generally, or concentrating only on final products.

Cross-Site Analysis Methods: A Case Account

We here offer an account of two of our cross-site analyses, drawing on the documentation data. The first deals with early implementation dynamics, and the second with changes in an innovation.

EARLY IMPLEMENTATION

The cross-site analytic question had two parts: (1) What was early implementation like? (2) What accounted for smooth and rough experiences during initial use? A key predictor for the second question was obviously the degree of preparedness, which is shown for one of the field sites in Table III.

We began with the site-level summary charts. For this question, we had two sets of charts: estimates of the presence or absence of requisite start-up (readiness) conditions (such as those in Table III), and another summary chart showing users' initial feelings and concerns, their degree of understanding of the project, what they were spending most of their time on, and what problems they were having. From the readiness charts, the data were simply standardized to an ordinal scale running from "factor not in place" to "factor fully in place" (the same scale, incidentally, as used in the survey data). From the initial-experience chart, the relative smoothness of early experience was assessed, which sometimes entailed making a determination not made in the site report, but which was typically derivable from the list of concerns and problems (a second analyst corroborated these judgments). Using a regression "logic", we then constructed the left-hand part of Table IX up through the column labeled "Training".

The display helped to visualize how good overall preparedness was (not too good), and which factors were most or least in place. It also allowed a visual estimate of the relationship between smoothness of initial use, program sponsorship (NDN versus IV-C), and readiness. Clearly, the next step was to go into further detail.

The between-program sponsor differences are marked: all five IV-C sites were rough starters. In order to examine the smoothness–readiness relationship, weighted scores were assigned and computed (see note c to Table IX). The scores also had another weighted estimate added: (F) means that the site researcher noted that the readiness component markedly facilitated early use by its presence (or hindered early use by its absence); (B) signifies a barrier. Where this datum was absent, it was inferred from the section of the site report concerning early experience, which sometimes involved a second-order estimate, reviewed by a second analyst. The preparedness scores and medians for each smoothness group are listed in Table IX. There is a linear relationship, but the range is modest and there are exceptions in the rough-starting sites. By returning to the display and the twelve short sections in the twelve cases, reasons for these exceptions were found. For example, the high score for the Carson site is a function of the fact that the commitment items may be overweighting the total preparedness score. We also derived scores for rough starters and very rough starters, and found the first group median to be 13 and the second 11, which strengthens the linearity of the relationship between preparedness and smoothness of early use. We further considered two more variables (last two columns) in order to determine whether these two conditions on the initial-readiness chart were actually met during early use.

It should be mentioned here that we always kept our words and numbers together, for two reasons. First, our previous experience has been that condensing qualitative data (words, plus verbal estimates of magnitude or valence, such as more–less, good–bad) into numerical variables, and then putting aside the field-study data and manipulating these values, is likely to lead to either banal or mystifying results, largely because the variables had not been created in the first place to be measured and transformed in this way. It would be better to put more energy into survey measures which are designed for such manipulations than to transform them in this way.

Secondly, it was important to be able to shift rapidly back down into the raw data set when equivocal results were encountered—back to the site-specific charts and text, and, if necessary, back to the coded field notes. If only numbers were available, it would have been necessary to resolve the problem with more numbers, and it would have taken a week to get back into the untransformed raw data.

However, the analysis is not yet complete. The display in Table IX indicates that some requisite conditions matter more than others (for example, consider those underlined, with (B) or (F) appended; it can be seen that user commitment is more important than user understanding); but also, only about one-third (42) of the total pool of readiness indices (132) are underlined. It can also be seen that being well prepared does not help smoothness of initial use as much as being poorly prepared hinders it. For some sites (e.g., Astoria), few or none of the factors are underlined. Thus, there must

TABLE IX

Predictor–Outcome Metamatrix Used in Analysis: Degree of Preparedness as Related to Ease of Early Implementation at Field Sites (First Generation of Users) [a]

Ease of early use at sites [b]	Commitment			Understanding		
	Users	Building principal	Central office administration	Users	Building principal	Central office administration
Smooth						
Astoria (NDN)	*f*	*f*	*f*	*f*	*f*	*f*
Burton (NDN)	*p*	*p*	<u>*f*(F)</u>	*p*	*a*	<u>*f*(F)</u>
Mostly smooth						
Lido (NDN)	<u>*f*(F)</u>	*f*	MD	*p*	*p*	MD
Mixed [d]						
Calston (NDN)	*f*	*f*	<u>*f*(F)</u>	*p*	*p*	*f*
Perry-Parkdale (NDN)	<u>*f*(F)</u>	*a*	<u>*f*(F)</u>	*f*	*a*	*f*
Rough						
Banestown (NDN)	<u>*f*(F)</u>	*p*	<u>*f*(F)</u>	*p*	*a*	*p*
Masepa (NDN)	*f*	*p*	<u>*f*(F)</u>	<u>*a*(B)</u>	*a*	*p*
Carson (IV-C)	<u>*f*(F)</u>	<u>*f*(F)</u>	*f*	<u>*p*(B)</u>	*p*	*p*
Dun Hollow (IV-C)	*p*	*p*	*p*	*f*	*f*	*f*
Plummet (IV-C)	*f*	*f*	*f*	*p*	<u>*p*(B)</u>	*p*
Proville (IV-C)	<u>*p*(B)</u>	<u>*p*(B)</u>	<u>*f*(F)</u>	*p*	*p*	*p*
Tindale (IV-C)	<u>*a*(B)</u>	<u>*f*(F)</u>	<u>*f*(F/B)</u>	*p*	*f*	<u>*p*(B)</u>

[a] Underlined entries signify that the field researcher estimated the corresponding factor to be decisive in affecting ease of early use: <u>(F)</u> indicates facilitation, <u>(B)</u> a barrier to successful early use; *f* indicates factor "fully in place", *p* indicates factor "partly in place", *a* indicates factor largely absent or missing; MD denotes missing data.

have been something else also influencing smoothness of early use. This called for a wider hunt for predictors, so we returned to the twelve brief sections concerning early experience, notably to the summary table of user concerns/behavior.

To summarize the results of that search, five more variables were derived at a low level of inference: the actual degree of practice change; the degree of latitude to make changes; the size or scale of innovation; the goodness of actual organizational fit; and the degree of user constraint to adopt. We then prepared a new array and found that these variables, especially the first three, indicated not only more about readiness, but more than readiness.

The smooth sites were small-scale, with a high latitude for making changes

Resources/ materials	Skills	Training	Preparedness score [c]	Group median	Ongoing aid/in-service	Building level-support
f	*f*	*p*	19	18	*p*	*p*
f(F)	*p*	*f*	17		*p*	*p*
f(F)	*p*	*f*	17	17	*a*	*f*(F)
a(B)	*p*(F)	*f*	16	15.5	*p*(B)	*f*
f	*p*(B)	*f*	15		*p*	*p*
p(B)	*p*(B)	*p*(B)	12	13	*p*(F)	*f*
a(B)	*a*(B)	*p*	8		*f*(F)	*p*
p	*p*(B)	*p*(B)	16		*f*(F)	*f*
p(B)	*f*(F)	*p*	14		*p*	*p*
p(B)	*p*(F)	*a*	14		*p*	*f*
p	*p*	*p*	10		*p*	*p*
f(F/B)	*p*	*f*	13		*f*(F/B)	*f*

[b] Field researcher's judgments from users' responses and/or from observation of practice in use.
[c] Computed in the following way: $f = 2$, $p = 1$, $a = 0$, (F) = +1, and (B) = −1 points.
[d] Smooth for some users, rough for others.

and minor-to-moderate practice change. They had used this latitude to derive a small innovation from the large-sized version they started with, so readiness was easy to ensure and initial use was smooth. The rough-starters involved major practice changes in moderate-to-large-scale projects, with varying amounts of latitude. Thus readiness was far more problematic, and initial use was difficult. We reached this conclusion by moving back and forth between the cross-site display and the site-specific displays and summaries, and by doing some more of the underlining and counting done for the readiness scale. In the process, we happened on a grounded second-order variable–administrative latitude—of the sort that is less likely to emerge in survey analysis when the stage is reached of recoding and rescaling derived variables, further and further away from the original data. This variable also served to strengthen a subsequent section of the cross-site analysis, as follows.

TRANSFORMATIONS IN INNOVATIONS

In the conceptual framework (Appendix II), we had postulated a cycle of transformations during which changes would ensue in an innovation, in the users' daily practice, and in the working arrangements of the school. In accordance with conflict theories of change (e.g., Schelling, 1963) and with empirical work on implementation (Berman and McLaughlin, 1975), we searched for reciprocal influences among users, innovation demands or characteristics, and institutions (as it turned out, organizational changes were not abundant, largely because most of the projects themselves were defined as institutionally modest or classroom-bound—if not at the start, then at the end of initial implementation).

As in the preceding section, the cross-site analysis of changes in innovations proceeded by stages, shuttling back and forth between the summary tables and text in the case reports and the emerging cross-site analytic findings. We first wanted to know, for the twelve field sites, the nature of such changes, how they evolved over time, the degree of change, and whether there were between-program-sponsor (NDN versus IV-C) differences. Then we would turn to a consideration of how and why such changes in the innovations occurred, and determine whether there were families of sites having similar profiles.

The point of departure was, again, the twelve summary charts from the case reports, of the form of Table IV. Note that there is a lot of text, and no specification of the type or nature of the changes made—a deliberate decision, so as not to predefine categories that we wished to be grounded empirically. Scanning the twelve charts, we sought indications of what was being done to the innovations. At the broadest level, we found that the entries in all cells for all twelve cases fell into three categories: people were reducing an innovation, adding to it, or reconfiguring it, i.e., reorganizing its parts and/or folding in segments according to their preimplementation repertoires. This led to the first descriptive metamatrix, shown as Table X (for simplicity, only the IV-C portion is presented here).

Once again, the matrix involved major data reduction (the three types of change, and the estimates of extent and of importance). It allowed us to see in one place the data to be analyzed, and to generate hunches to guide that analysis. The first feature noted—through what we called "squint analysis" —was that the NDN projects involved making a lot of reductions and making them early, quite unlike the IV-C profiles. In fact, we were fascinated to see that many of these reductions were made before actual implementation began. We also saw that some sites continued to reduce an innovation over time, whereas others stabilized, and yet others reinserted components they had discarded previously.

To estimate the degrees of change, we simply counted the numbers of components which had not been used or had been changed, using an instrument from the survey which listed the components and showed changes which the developer deemed unacceptable. We did this for two periods (cutting the total time in half), constructed a simple ordinal scale (see note c to Table X), and produced the penultimate column of the table. This indicates that more than one-half (7) of the sites had changed between one-third and two-thirds of the components considered essential by the developer to unacceptable versions. Three more sites show such a degree of change in more than two-thirds of the key components. Only the two remaining sites were faithful implementations, so the degree of change was far-reaching [6].

However, the significance of these changes varied. It seemed that some of the components within a project were rather trivial ("teacher visits student in detention") and others more weighty ("special education teacher teams with regular teacher to develop, update and implement individualized educational plans"). Thus we needed an index not only of the amount of change, but also of the significance of changes made in the innovation, which is shown in the last column of Table X. This index was derived by judging and weighting each component changed as minor, moderate or significant, and then taking the median. This estimate was a discretionary one on the part of the analyst—and one not done independently by the two analysts, although it well could have been. This point was duly noted in reviewing the confidence that could be placed in the analysis. We then examined the degrees of significant change (five significant-change sites, four moderate, and three minor), and began to examine the text content in the site reports for the significant and minor-change cases in order to determine how significant-change sites differed from minor-change ones, and obtain details of how users and administrators were justifying these changes. On this last score, we found six distinct motives that exhausted the full set (this is another instance showing our belief that qualitative cross-site analysis has to be exhaustive, rather than probabilistic; *all* the cases must be categorized, and usually all analyses and interpretations must be made using the full set —a more stringent condition than for many kinds of statistical analysis).

The reasons given by users and administrators for making changes were the starting point in our search for predictors. If, for instance, teachers across the twelve sites were saying simply that they did not like a new practice, we considered this as a possible instance of poor user fit. We began

TABLE X

Excerpt from Descriptive Metamatrix Used in Analysis (IV-C Sites): Types, Evolution and Extents of Changes in Innovation [a]

Site	Initial implementation (6–9 months)			Second year		
	Additions	Reductions	Reconfigurations	Additions	Reductions	Reconfigurations
Carson	Full-scale implementation (by volunteers)			More testing; more centralization; adding staff (aide); adding in-service training	Fewer conferences; fewer meetings; less use of community	Simplification of educational plans, using existing mechanisms (parent meetings)
Dun Hollow	More lessons, activities, materials	—	—	Still more instructional supplements, e.g., vocabulary cards and sentences	—	—
Plummet	—	Lowered academic content	—	More staff; more counseling; more teaming	Dropping conventional curricula	Organizing program
Proville	—	Simplified student selection; not recruiting staff; no training; poor student–job match; deleting jobs	—	Partial reinstatement of first-year reductions	—	—
Tindale	—	Omitting small sections of units; selective use of materials; uneven testing	—	Partial reinstatement of first year reductions	More widespread discarding of small sections; elimination of one unit	Shifting sequences

[a] Blank cells indicate no changes (initial implementation), or no further changes (later implementation).

Third year			Extent and evolution of changes [b]	Importance of changes [c]
Additions	Reductions	Reconfigurations		
Program mandated for all	Less monitoring; educational plans more desultory; fewer meetings, conferences; evaluations incomplete; in-service incomplete	"Batching" of students	Low–moderate	Significant
Supplementary pictures	Not using approved resource personnel	Rearranging instructional sequences (to make them more logical)	Low–moderate	Minor
Same trend continued in second year			Low–High	Significant
Further reinstatement of first-year reductions	—	—	High–moderate	Significant
—	—	More widespread shifts in sequences	Low throughout	Minor

[b] Low, less than one-third of key components; moderate, one-third to two-thirds of key components; high, more than two-thirds of key components shifting into "unacceptable" or "absent" category.
[c] Researcher's estimates.

TABLE XI

Predictor–Outcome Matrix used in Analysis: Effects of Five Predictors on Changes in Innovations [a]

Site	User fit [b]	Anticipated classroom change [b]	Anticipated organizational change [b]	Organizational fit	Demand characteristics [c]
Significant changes					
Astoria (NDN)	Good	High	High	Poor	Strong
Burton (NDN)	Fair	High	High	Poor	Strong
Carson (IV-C)	Fair–good	High	High	Fair–good	Moderate–strong
Plummet (IV-C)	Good	High	High	Good	Strong
Proville (IV-C)	Good	Low [c]	Moderate	Fair	Moderate
Moderate changes					
Banestown (NDN)	Good	Low	Low–moderate	Good–fair	Small–moderate
Calston (NDN)	Good	Moderate	Low	Good	Small–moderate
Lido (NDN)	Fair	Moderate	Low	Good	Small–moderate
Perry-Parkdale (NDN)	Fair–good	Low [d]	Moderate	Good	Moderate–strong
Minor changes					
Masepa (NDN)	Fair–Poor	High	Moderate	Fair	Moderate–strong
Dun Hollow (IV-C)	Poor	Low	Low	Good	Small–moderate
Tindale (IV-C)	Fair	Moderate	Low	Good	Moderate

[a] Underlining indicates strong predictors of ensuing change.
[b] From informants' responses.
[c] Researchers' estimates.
[d] Program with few classroom-level components.

with six predictors similar to those for which we already had codes, then added two predictors that were congruent with the reasons given for making changes, and had also emerged earlier in the analysis. This exhausted the set. Table XI is a display which allowed examination of five of these variables, with sites arrayed by degree of significance of change.

Without going into detail, note that: (a) some of these variables covary fairly well with the significance of change, while others do not; (b) there are relatively few underlines to denote causal influence, and no clear pattern except for the user-fit variable at minor-change sites; and (c) there are a fair number of second-order researcher estimates that required data transformations from site-report tables or text. Again, the second analyst verified the first; the first analyst also re-did the same exercise from the beginning two days later, which yielded only one discrepancy in the 60 cells.

A similar analysis (not detailed here) was made for the three remaining variables: implementation readiness, administrative latitude to make changes, and initial (pre-implementation) scale or scope. For the middle variable, administrative latitude, a significant result was found. The chart indicated that latitude was high in all cases except those where minor changes had occurred in the innovation. This suggests that users were making more significant changes whenever they had permission to do so—and thereby, of course, as was seen in Table X, generally reducing the degree of practice change. Greater latitude meant less relative innovativeness. Moreover, this variable resulted in underlines in nine of the twelve cases, making it a far more significant predictor, in the eyes of field researchers and analysts, of innovation changes than the other seven variables examined. Finally, of these nine sites, the chart showed that eight were NDN projects. Since Table X had indicated that modifications in NDN innovations came early, often prior to actual implementation, it appeared as if administrators were giving early and wide latitude to users who were asking for it—possibly as a precondition of use in their classrooms. If this was so, we had a bargaining-exchange paradigm (e.g., Elmore, 1978) with strong potential explanatory power.

The enchantment of qualitative data analysis in cases like this is that it is easy to return to the raw data rapidly—provided, of course, that the corresponding sections are readily accessible, and that those sections are already summarizations of the raw data set. It is then possible to search in the innovative-change sections for data indicating how and when modifications were made in the innovations. In this instance, we were returning to about five pages of text and two tables for each of the twelve sites, which was manageable. The documentation form shows that this next step took about an hour. This could not have been done as quickly had we had, say, 45 cases, but it is fair to argue that if there were 45 cases, either survey-type analysis should be done in the first place, or a family of 12–15 sites should be selected and a qualitative analysis done on those.

In at least four of the eight NDN cases, we found precisely the scenario we hypothesized might be present—overt bargaining for latitude to make preimplementation changes. There was also covert bargaining at all sites, which often appeared in the form of informants' accounts of fine-tuning of a project, or "making it easier to do in the school" by reducing an innovation's disruptive impact. We also noted a bargaining chain stretching from the original developers, who accepted local adaptations in order to get their projects installed, to central office administrators haggling with principals, then to principals negotiating with teachers. Further, these deals differed across levels and sites, ranging from unlimited discretionary power given to teachers to make changes, to strict policing of users.

Where to go from here? We saw emerging trends but could not yet plot them. We appeared to have an economical construct (negotiated change) uniting several strands, but it was not yet clear how this construct interacted over time with the other seven predictors. We saw, if very dimly, various scenarios.

We also had data from elsewhere in the cross-site analysis that could provide alternative interpretations. For example, we had found earlier that approximately one-half of the users had adopted an innovation because of administrative pressure: given the choice, perhaps they would have imple-

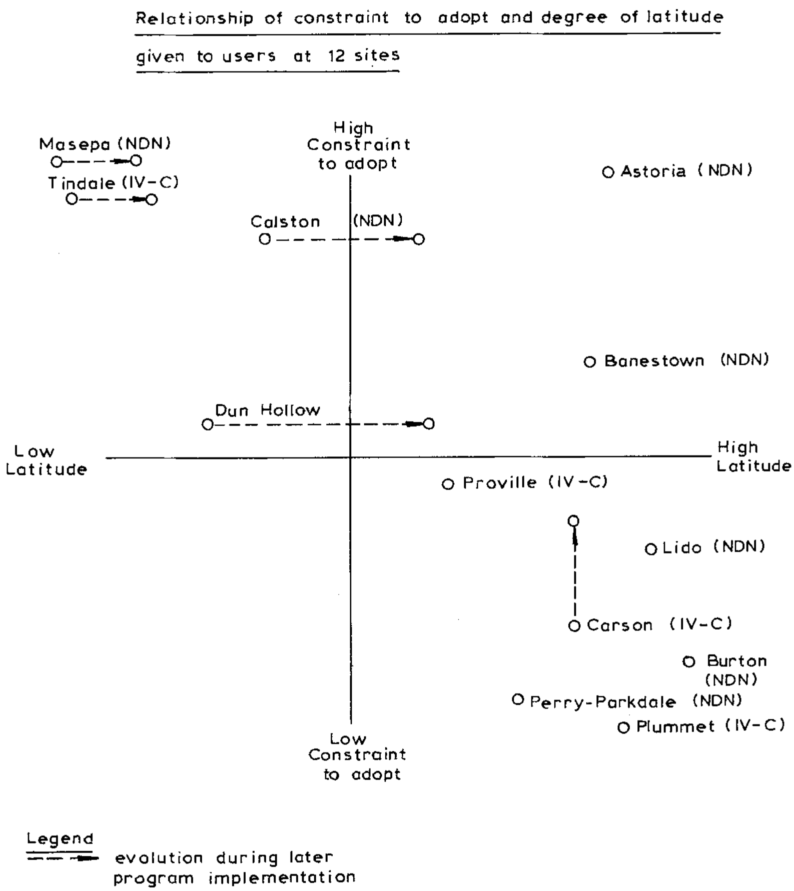

Fig. 4. Circumplex Plot used in the Analysis.

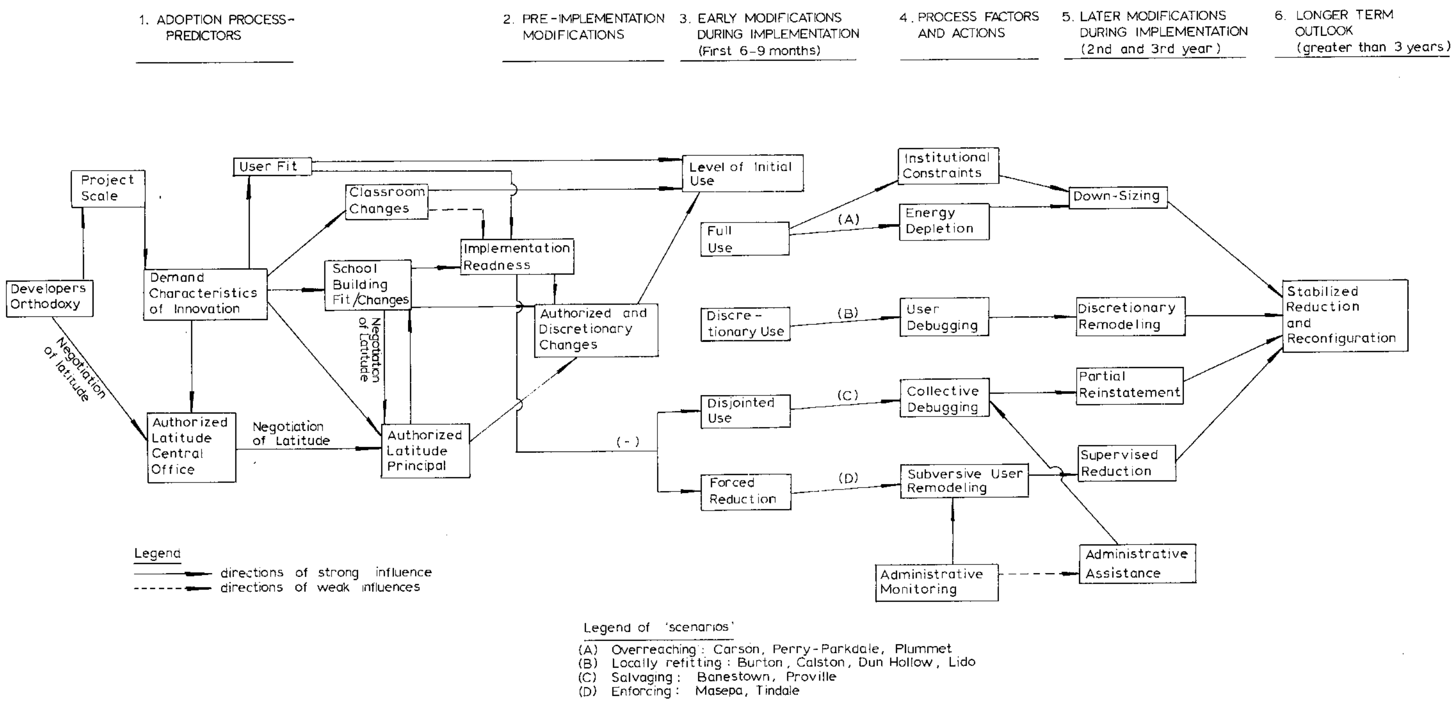

Fig. 5. Final Model of the Process of Change in the Innovation.

mented less of it, or none at all. Perhaps administrators were letting users modify projects as a reward or palliative after having pressured them to adopt. We needed to check this hypothesis before pursuing the logic being evolved within the eight-predictor data set under analysis. Here again, we made the same two moves as stressed throughout this paper, namely, setting up the data in ways that allowed us (1) to see clustering and covariation, and (2) to gauge, if only primitively, the extent or significance of these relationships. In this instance, we plotted constraint versus latitude in circumplex form, in order to investigate whether there was bunching in the appropriate quadrants (high constraint–high latitude, low constraint–low latitude). Figure 4 shows that there was not, although there was a drift in this direction during subsequent implementation.

The next step was to search for a condensed, integrated way of segmenting the emerging bargains, scenarios and interactions among the predictors. To save time, we simply show the ultimate model derived, and then backtrack rapidly to trace how it was arrived at. This also allows us to include some excerpts from the self-documentation form [7] and, in so doing, to indicate how a secondary analyst can verify or audit the data-analytic trail of such a qualitative study.

Figure 5 lays out the model. It may be divided into three parts: (a) the latitude issue; (b) the remaining seven predictors; and (c) the scenarios.

(a) The latitude stream runs along the bottom of the preimplementation phase (left-hand side of Fig. 5) and shows the cascade of negotiations from developer to end-users, together with the interactions between this variable and the other predictors. Drawing on the self-documentation form, here is how the stream was obtained:

Procedure	*Results, conclusions*
1. Read section in case reports to obtain bargaining points in time, and actors involved	Strong, decent data base in all twelve cases
2. Sketch out latitude stream	Straightforward: virtually identical for all NDN sites
3. Read case sections to see whether consented latitude actually led to modification	Explicit for all NDN sites; implicit but at higher inference level for IV-C sites
4. Connect latitude stream to other predictors—if more than three cases of *specific, explicit* influence and no reverse/inconsistent influence, draw arrow	Easy criterion to meet: "causal" flow makes logical sense

Note that the analytic procedures combine counting, transforming (text into "bargaining points") and inferring. The inferences appear plausible, but they also appear subjective (explicit–implicit inferences, logical sense), and make causal claims resting largely on the sequences reported in the case reports.

Next we turn to task (b), that of examining the relationships among the seven other predictors, above the latitude stream:

Procedure	*Results, conclusions*
1. Read sections in case reports to obtain sequence of predictors, starting with underlined predictors that assume causal influence	At least four sequence streams in twelve cases: needs further reduction
2. If at least three cases have a sequence linking two variables in the same order, draw arrow; if two cases, draw broken arrow	Easy criterion to meet: four cases for most arrows
3. Reread high–low cases for each box connected to another box, to see whether causal links work for both	Covariation works with underlined cases
4. Does model make sense?	Yes, but other models could make as good sense

Two observations should be made here. First, although the procedures are hardly elegant, they represent the rudiments of a data-analytic strategy that combines inductive and deductive methods. Second, the cross-site analysis is ultimately only as strong as the sum of the site-specific analyses. The relationships, influences and causal inferences are made already in the site reports [8]; the cross-site analyst extracts them, directly (from explicit text or magnitudes) or indirectly (by transforming text into magnitudes, and by inferring explicit relationships from what s/he determined are implicit ones). It should be remembered that the findings and inferences in the site reports themselves are not probabilistic in a statistical sense, nor even Bayesian, but inductive. Thus, we are really not talking at all about a numerical causal path or chain of multiple indicators as in a LISREL prediction model, but rather about a causal model built by someone like a forensic pathologist, a detective or an historian, using a progression of inferential analyses to run an evidential trace out to its end point.

Finally, we move to task (c), that of discovering whether there are different scenarios or families of sites with similar courses of events during the process of change in innovation. The scenarios run from column 3 to column 6 in Fig. 5. Here is how they were obtained:

Procedure	*Results, conclusions*
1. Recode chart (see Table X) to obtain types of change and times of change; cluster sites	Three or four overlapping patterns
2. Determine amounts of change (by counting component shifts) at times of first implementation	

3. Connect predictors to T1 (see column 3) by following decision rule: connect if three cases of explicit influence and no reverse/inconsistent influence	Four types for twelve cases; two types have a readiness problem
4. Do same exercise for later changes, types, extents, bargaining points; use decision rules: at least two key predictors the same; same overall amounts of change (numbers of components changed); same trends of change; e.g., low–moderate, high throughout (see Table X); similar bargaining and latitude-giving processes: same actors, same arguments, same results for outcome	Reduces twelve sites to six, but three profiles have only one case each
5. Collapse ultimate criterion measure (outcome) (distinction conceptually muddled)	Reduces to four families, but still one profile with only one case
6. For remaining eleven sites, derive overall name for each of four scripts: "overreaching", "locally refitting", "salvaging" and "enforcing"	
7. Derive names for later process factors and later changes made in four scripts	Some links between scenarios C and D
8. Reiterate each step, especially step 4	One site perhaps in two scenarios, but computation of component changes locate it in B, so kept as is

This somewhat circuitous trail took a day and a half. A second analyst then went over and verified it. Obviously, a second analyst might have done it differently, although s/he too would probably have looked for clusters (the scenarios) and factors (the predictors) in one shape or another. This is not unreasonable—two quantitative data analysts also do the same thing as a result of their leanings and experience. Some favor R, some Q analysis; some would adopt either. There are linearists and nonlinearists, etc. The important point is that the second analyst can follow and verify the procedures of the first.

A useful aid in this audit, we would argue, is provided by the successive data displays, which show another analyst the stages which have been traversed and, thereby, how to retrace the trail. The displays also render a similar service to the reader, by reducing the typically voluminous case studies of qualitative researchers into a more amenable form.

Concluding comments

Four major points have been made in this paper, behind the welter of detail that is probably the natural consequence of ethnographic concern for completeness. First, data reduction, also known as data transformation, occurs at all points in a study from design through to data collection and write-up. It is not something separate from analysis, it is analysis: analysis of a form which sharpens, sorts, focuses, throws away, organizes and clarifies data in such a way that final analysis can occur coherently.

The second main point is that data reduction and data analysis have as an indispensible accompaniment some form of data display, and that the display modes chosen will inevitably condition the processes and conclusions of analysis. The most frequently adopted and typical display mode for qualitative data to date—narrative text—is also the most cumbersome and limiting one imaginable. We here advocate much more elegance, simplicity and variety in display modes. We have emphasized matrices and figures of several sorts, and many other types can be generated that meet the need to display data coherently and compactly.

The third main point is that single-site and multisite analysis processes, at the present state of the art, are of course complex, perhaps even obsessively baroque, but not arcane, obscure or ineffable. It *is* possible to understand, we have found, how an analyst X got from point A to point B [9].

The fourth main point follows immediately. It is possible to understand processes such as those considered, if they have been documented accurately, using some reasonably standardized scheme. Such documentation permits an external audit, in Guba's terms; it allows reproducibility of findings—the core of science—and replicability of studies; it can support dialogue among researchers struggling with qualitative analysis that can lead to something resembling shared methodological canons. Such canons save energy, reduce doubt and anxiety over the status of conclusions, and enable accumulation of knowledge. However, the presence of such canons is not an unmixed blessing. Judging from the history of quantitative analysis methods, methodological canons also result in endless disputation and refinement of effort, intense socialization of novitiates into a received orthodoxy, and preoccupation with methods rather than with the substance of inquiry. It would be a great pity if the next cohort of graduate students had always to be mercilessly grilled on whether their predictor–outcome matrix followed Huberman–Miles Rule II, and if the proportion of qualitative methodological papers at researchers' conventions overwhelmed that of papers reporting substantive findings.

Perhaps it is possible to ward off at least some of the probable bad effects of becoming more systematic about qualitative data analysis. However, self-documentation is labor-intensive and not a total substitute for verbal elaboration. Filling out our forms usually took at least 15% of the total analysis time. Even taking that much care, we found that an uninitiated reader could not quite follow what had been done without added oral explanation (of the sort offered in the preceding sections, but even more detailed). Thus, although the present documentation method is actively recommended, with appropriate revisions and further iterations to make it easier and more useful, its limitations should be borne in mind. Above all, it is hoped that the work reported here will help supply some of the common language and syntax that qualitative researchers will need in elaborating a well documented and credible methodology.

Notes

[1] Hanson (1958) formulated an especially striking aphorism along the same lines. Causes, he wrote, "are certainly connected with events, but this is because our theories connect them, not because the world is held together by cosmic glue" (p. 64).

[2] Note that these methods have also been used successfully in another study (Huberman, 1981b).

[3] Beverly Loy Taylor and Jo Ann Goldberg were our colleagues in the design, data collection and within-site analysis portion of the study, with responsibility for eight of the twelve sites. Their energy, care and determination were crucial for the quality of the study.

[4] Note, however, the scaling confusion in the Table III matrix: the estimates range from low–high, absent–present, poor–good, whereas the survey scores (Crandall et al., 1982) are on a "not in place"–"fully in place" scale. This is what happens typically when formats are not fully standardized at the outset. Note also that individual responses have been aggregated.

[5] This technique combines and complicates such distinctions as the "constructive" versus "enumerative" strategy of data reduction made by Goetz and Lecompte (1981). Causal network variables were indeed generated from the local stream of behavior (constructive approach), but they could derive, directly or indirectly, only from the pre-established conceptual framework and research questions (enumerative approach).

[6] The survey data for the entire sample of 146 sites did not show that as much unacceptable change had occurred. That is understandable, since the users in the larger sample were only current users (2–6 per site), interviewed at one point in time about their present use. We assembled data from all users we could find, past and present, and tracked use over the course of a school year, supplementing interview data with observations. Under such circumstances it would be expected naturally to note that more adaptation in innovations had occurred. It should also be noted that in three of the five IV-C and in two of the seven NDN sites in the sample, adaptation and development were being encouraged explicitly or implicitly. We do not know whether such figures are typical.

[7] The quotes have been compressed and edited here and there to make them more readable to people not familiar with the study.

[8] We wish to stress this point heavily. Causality, where it happens, happens *locally*. Too many analysts seem to assume, like survey researchers, that some sort of averaged-out causality can be inferred from summed variables, each considered atomistically across all sites.

[9] The documentation also shows the reader the labor-intensiveness of the approach used. Although the episode we have described took a day and a half, there were many, many such episodes. Our deliberately thorough cross-site analysis occupies about 350 pages, while the original twelve site reports total about 1000. Less intensive approaches could of course be adopted using many of the techniques described here, but we do not wish to encourage the view that qualitative data analysis is cheap. It is probably as expensive as the data-processing costs typically incurred in mounting and analyzing a large and complicated survey data base.

Appendix I: Qualitative Analysis Documentation Form

Page 1

1. *Research issue being explored*: *Analyst* *Date*:

2. *In this analysis task, what specifically were you aiming to do?*
Say whether exploratory or confirmatory intent. Give context, and a short rationale, so overall picture is clear. Refer to past documents or activities if relevant.

3. *Describe the procedure that you followed, step by step, in sequence.*
Do this as a log, as you go. *Give details*. If form fills up, *stop*. Do page 2, even if task is incomplete. If task changes part way through, *stop*, do page 2.

Specific data sets in use [a]	Procedure (number each one; show *what* was done, *how* it was done; down *feelings* you had en route) [b]	Why you did this (include *assumptions* made)	Time taken	Preliminary conclusions drawn (number each one; distinguish *procedural* (P) from *substantive* (S)	Basis for conclusion (roughly stated)
—	—	—	—	—	—

[a] Show single site or multisites; examples are coded questionnaires, event listings, causal nets, write-ups, case report, site report (specify site), overviews, matrices, charts, etc. Give letters (A, B, C, etc.) to items used that are attached as exhibits.

[b] Be explicit. You are recording procedure for your own later write-up. Give *examples*, or allude to them, in matrix or report. Give *details*: which site, data source, decision rules used, etc. Include dead-ends, stuck places, etc.

Page 2

4. *Review the procedures you used.*
Circle those you were using.

1. Formula question — Select Refine Invent
2. Retrieve data — Gather Read

3.	Data classification/ reduction	Sort/clump Locate themes Create codes Write summaries Identify metaphors
4.	Code data	Assign Revise codes
5.	Connections/comparisons between qualitative/ quantitative data	
6.	Display data	Fill in Invent Revise (tables, matrices, text, nets, charts)
7.	Conclusions	Draw Test Support Propose hypotheses
8.	Verification of conclusions	Corroborate Establish confidence level Assess repeatability
9.	Write-up	Outline Organize information Write text Revise text
10.	Avoidance	Sharpen pencils Eat Scan papers
11.	Other (specify) _________	

5. *Review the conclusions you reached.*
Write in their numbers below to show the basis you used in reaching them. Write in other basis if used.

_____ Counting, adding up	_____ Convergence (repeated evidence)
_____ Triangulation (use of different data source to confirm)	_____ Divergence, meaning of outliers, exceptions
_____ Weighting of evidence	_____ Testing specific hypotheses in another part of data
_____ Rival explanations	_____ Extreme-bias cases
_____ Plausibility	_____ Reference to other studies, literature
_____ Other (specify)_____	_____ Contrasts, comparisons (specify) _____
_____ Other (specify)_____	

6. *Confidence in conclusions.*
Write numbers, assign confidence percentages, give reasons.

Conclusion	Confidence (%)	Reason
—	—	—

7. *How useful were the procedures and analysis methods you used?*

Strengths ____________
Weaknesses ____________
Suggested changes ____________

8. *Very important: Attach or refer to all exhibits, with letter code for each.*

Work notes/work sheets____________
Data tabulations____________
Displays ____________
Other exhibits____________
Resulting text (essential)____________

9. *Review page one.*
Procedures complete, including dead ends?
Are procedures clear enough for me to write up later (edit if needed)?
All exhibits lettered and attached? Text included?

10. *Approximate time to fill out this form* (during logging/afterwards).

Appendix II: Field-Study Conceptual Flow Chart (Original Version)

Impinging factors	*Internal context as "host"* (adopting school (time 1))	*Adoption decision*	*Cycle of transformations* (time 2, …, *n*)	*Outcomes* (reconfigured school (time *n*))
External context: community district office Assistance (external, internal) orientation interventions Innovative program: assumptions characteristics	Demographics ↓ Prior history with innovations ↓ Organizational rules, norms, work arrangements, practices within class and school at large, relationships ↑ ↓ User purposes assumptions, i.e., beliefs, motives	Decision to adopt Plan for implemen- tation Support for imple- mentation	Changes in innovation as presented to user ↑ ↓ Changes in user perceptions and practices ↑ ↓ Changes in orga- nizational rules, norms, practices relationships	Degree of insti- tutionalization ↑ ↓ Perceived gains and losses: individual and institutional, innovation-specific and metalevel, anti- cipated and unanticipated ↑ ↓ Side effects: positive and negative, antici- pated and unanticipated

References

Becker, H. (1958). "Problems of inference and proof in participant observation", *American Sociological Review* 23: 652–660.

Barton, A. and Lazarsfeld, P. (1955). "Some functions of qualitative analysis in sociological research", *Sociologica* 1: 324–361.

Berman, P. and McLaughlin, M.W. (1975). *Federal Programs Supporting Educational Change. Vol. IV. The Findings in Review*. Santa Monica, CA: Rand Corporation.

Bruyn, S. (1966). *Human Perspectives in Sociology*. Englewood Cliffs, NJ: Prentice–Hall.

Campbell, D. (1975). Degrees of freedom and the case study", *Comparative Political Studies* 8: 178–193.

Campbell, D.T. and Stanley, J. (1963). "Experimental and quasi-experimental designs for research", pp. 171–246 in N. Gage, ed., *Handbook of Research on Teaching*. Chicago: Rand McNally.

Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Crandall, D.P., Bauchner, J., Loucks, S. and Schmidt, W. (1982). *Study of Dissemination Efforts Supporting School Improvement. Vol. 1. Moulds of the School Improvement Process.* Andover, MA: Network Inc.

Cronbach, L. (1975). "Beyond the two disciplines of scientific psychology", *American Psychologist* 30: 116–127.

Douglas, J. (1976). *Investigative Social Research*. Beverley Hills: Sage.

Dreitzel, H., ed. (1970). *Recent Sociology*. Vol. 2. London: Macmillan.

Elmore, R. (1978). "Organizational models of social program implementation", *Public Policy* 26: 185–225.

Firestone, W.A. and Dawson, J.A. (1981). *To Ethnograph or Not to Ethnograph? Varieties of Qualitative Research in Education*. Philadelphia: Research for Better Schools.

Glaser, B. (1978). *Theoretical Sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory*. Chicago: Aldine.

Goetz, J.P. and Lecompte, M.D. (1981). "Ethnographic research and the problem of data reduction", *Anthropology and Education Quarterly* 12: 51–70.

Guba, E.G. (1981). "Criteria for assessing the trustworthiness of naturalistic inquires", *Educational Communication and Technology Journal* 29: 75–91.

Guba, E.G. and Lincoln, Y.S. (1981). *Effective Evaluation*. San Francisco, CA: Jossey–Bass.

Hanson, N. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hesse, M. *The Structure of Scientific Inference*. London: Macmillan.

Huberman, A.M. (1981a). Splendeurs, misères et promesses de la recherche qualitative", *Education et Recherche* 3: 233–249.

Huberman, A.M. (1981b). *School–University Collaboration Supporting School Improvement. Vol. 1. The Midwestern State Case.* Washington, DC: American University, Knowledge Transfer Institute.

Huberman, A.M. and Crandall, D.P. (1982). "Fitting words to numbers; some approaches to multisite, multimethod research in educational dissemination", *American Behavioral Scientist* 26: 62–83.

Huberman, A.M. and Miles, M.B. (1982). *Innovation Up Close: A Field Study in Twelve School Settings*. Andover, MA: Network Inc.

Kelley, H. (1973). "The process of causal attribution", *American Psychologist* 28: 107–128.

Kim, J. (1981). "Causes as explanations: a critique", *Theory and Decision* 28: 293–309.

Lindesmith, A. (1947). *Opiate Addiction*. Bloomington, IN: Principia Press.

Lofland, J. (1971). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* Belmont, CA: Wadsworth.

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

Miles, M.B. and Huberman, A.M. (1983). *The Realities of School Improvement: Analysis of Qualitative Data.* New York: Center for Policy Research.

Nisbett, R. and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgement.* Englewood Cliffs, NJ: Prentice Hall.

Schelling, T. (1963). *The Strategy of Conflict.* London: Oxford University Press.

Sieber, S.D. (1976). *A Synopsis and Critique of Guidelines for Qualitative Analysis Contained in Selected Textbooks.* Project on Social Architecture in Education. New York: Center for Policy Research.

Smith, L.M. (1978). "An evolving logic of participant observation in educational ethnology and other case studies", pp. 316–377 in L. Shulman, ed. *Review of Research in Education.* Vol. 6. Itasca, IL: Peacock.

Snow, R. (1974). "Representative and quasi-representative designs for research in teaching", *Review of Educational Research* 44: 265–292.

Stake, R. and Easley, J., eds. (1978). *Case Studies in Science Education.* Urbana, IL: Center for Instructional Research and Curriculum Evaluation.

Stearns, M., Green, D. and David, J. (1980). *Local Implementation of PL 94-142: First Year Report of a Longitudinal Study.* Menlo Park, CA: Stanford Research International.

Turner, R. (1953). "The quest for universals in sociological research", *American Sociological Review* 18: 604–611.

Wax, R. (1971). *Doing Fieldwork: Warnings and Advice.* Chicago: University of Chicago Press.

Wilson, S. (1977). "The use of ethnographic techniques in educational research", *Review of Educational Research* 47: 245–266.