DOCUMENT RESUME

ED 377 235 TM 022 496

AUTHOR Rossman, Gretchen B.; Wilson, Bruce L.
TITLE Numbers and Words Revisited: Being "Shamelessly Eclectic."
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Mar 91
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; *Educational Research; Evaluation Methods; Literature Reviews; Mathematics Tests; Policy Analysis; *Policy Formation; Public Policy; *Qualitative Research; Research Design; *Research Methodology; Social Science Research; *State Programs
IDENTIFIERS *Mixed Method Triangulation

ABSTRACT

This paper offers concrete examples from recent policy research about how qualitative and quantitative methods can be combined to better address research questions. Using a conceptual framework developed by Rossman and Wilson (1985), later expanded by Greene, Caracelli, and Graham (1989), the paper discusses how the design and analysis phases of research can apply mixed methods to enhance understandings of social phenomena. A purposive sample of 57 mixed-methods evaluation studies was used to test components of the framework. A new typology of mixed-methods purposes is developed that identifies: (1) corroboration; (2) elaboration; (3) development; and (4) initiation. Crossing the four purposes with research design or analysis stages yields a 4 x 2 matrix. Examples of how methods can be effectively combined are drawn from two policy studies that focused on the effects of state policies on local school districts. In both studies a combination of qualitative and quantitative methods was consciously included. Mixed-methods approaches work best when more than one researcher is engaged, when norms of respect and collegiality prevail, and when an attitude of healthy skepticism about both theory and method exists. (Contains 25 references.) (SLD)

TM

NUMBERS AND WORDS REVISITED:

BEING "SHAMELESSLY ECLECTIC"

Gretchen B. Rossman
University of Massachusetts

Bruce L. Wilson
Research for Better Schools

March 1991

Paper presented at the annual meeting of the American Educational Research Association, April 5, 1991.

The preparation of this paper was partly supported by funds from the US Department of Education, Office of Educational Research and Improvement (OERI). The opinions expressed do not necessarily reflect the position of OERI, and no official endorsement should be inferred. This draft is being distributed on a restricted basis to obtain suggestions for revision and should not be circulated further without the permission of Research for Better Schools, Inc.

NUMBERS AND WORDS REVISITED: BEING "SHAMELESSLY ECLECTIC"[1]

The theoretical literature on methods has increasingly called for research that combines methods. Except for the purists who argue that different methods derive from mutually exclusive views about society and social science research (Smith, 1983; Burrell & Morgan, 1979), there is growing acknowledgment that complex social phenomena can usefully be understood by looking at them both quantitatively and qualitatively. There are essays on why mixed-methods research should be done (e.g. Mathison, 1988) and conceptual frameworks to assist those who contemplate doing it (e.g. Greene, Caracelli, & Graham, 1989). Yet, there continues to be a dearth of examples where what we preach is actually being practiced. For example, in reviewing the two most prominent journals on policy research, Educational Evaluation and Policy Analysis and Educational Policy, we found only a small handful of examples of published studies over the past two years where multiple methods were employed.

This paper offers concrete examples from recent policy research about how qualitative and quantitative methods can be combined to better address research questions. Using a conceptual framework developed by Rossman and Wilson (1985) and then expanded by Greene, Caracelli, & Graham (1989), we describe how both the design and analysis phases of research can apply mixed-methods to enhance the research purposes of corroborating, elaborating, developing, or initiating understandings of social phenomena.

A Typology of Mixed Methods Research

An initial attempt at conceptualizing different purposes for mixed methods research was outlined by the authors (Rossman & Wilson, 1985). In that article we argued that three purposes are served at the analysis stage for combining multiple methods in a single study: corroboration, elaboration, and initiation. Examples were offered from a three-year evaluation study of the services provided by regional education service agencies. Those examples showed how both quantitative and qualitative methods could be used reciprocally to inform each other for purposes of corroborating, elaborating, or initiating research findings.

The mixed methods conversation was greatly enhanced by the contribution of Greene, Caracelli, and Graham in 1989. They extended the theoretical framework for mixed methods research by elaborating our initial three purposes and suggesting an additional purpose for combining methods: development[2]. A purposive sample of 57 mixed-methods evaluation studies were used to test the components of their framework. By combining the two works, we generate a new typology of mixed methods purposes: corroboration, elaboration, development, and initiation. We argue that each purpose may be applied at either the research design or analysis stage, with one exception -- initiation. By crossing the four purposes with two stages, we generated a 4 x 2 matrix. Before we present the application of this typology with examples from two recent studies, we define each mixed-method purpose and discuss the distinction between design and analysis stages of the research process.

Corroboration

The first purpose is classical triangulation where different methods are employed to test the consistency of findings from one method to another. Jick (1979: 603) refers to this design as "the archetype of triangulation strategies". Designed to be implemented essentially simultaneously but independently of one another, qualitative and quantitative methods are expected to generate data that will "pinpoint the values of a phenomenon more accurately by sighting in on it from different methodological viewpoints" (Brewer & Hunter, 1989: 17). This corroborative purpose maps onto what Greene, Caracelli, and Graham (1989) refer to as triangulation or the correspondence of results from different methods. Rossman and Wilson (1985) refer to corroboration as convergence, with one method used to confirm the results of the other.

A useful example of corroboration is offered by Brewer and Hunter (1989). It derives from crime incidence studies. Historic assessments of crime rates relied on unobtrusive measures: police records. Critiqued because of a bureaucratic bias, crime studies now often incorporate victimization studies: surveys of relevant segments of the population for their reports of crime. Thus, the use of both methods -- police records and surveys -- converging on a single phenomenon -- crime rates -- is an example of classical triangulation.

Elaboration

The second purpose in our typology (Rossman & Wilson, 1985) is elaboration. This is similar to the Greene, Caracelli, and Graham (1989) concept of complementarity. Elaboration provides a richness and detail that is often lacking if just one method is employed. As Greene, Caracelli and Graham (1989: 259) note, this purpose "enhance(s), illustrate(s), clarif(ies) the results from one method with the results from the other". If we think of social phenomena as gems, elaboration designs are intended to illuminate different facets of the phenomenon of interest.

In a study of educational attainment of Thai children, Fry, Chantavanich, and Chantavanich (1981) used econometric analyses to isolate school effectiveness. The researchers then conducted indepth ethnographies at the schools identified at either extreme (predicted effectiveness compared with actual effectiveness) to better understand "their actual educational processes" (p. 154). Ethnographic information was used to elaborate and extend understanding of the influence of particular schools on educational attainment as documented by quantitative results.

Development

The third purpose, one not captured in our original typology, is described by Greene, Caracelli, and Graham (1989) as development. This new purpose offers an important contribution to our original conceptualization. Development uses the efforts from one method to inform the other method.

That is, the results generated by one method shape subsequent instrumentation, sampling or analysis strategies of the other method.

An example comes from the study of organizational cultures. Siehl and Martin (1988) describe a two-stage process for studying organizational culture. Noting that qualitiative methods are uniquely suited for capturing complexity (Siehl & Martin, 1988), they lament that such description does not permit answering important theoretical questions. Comparisons across time and groups demand systematic assessment through "repeatable, easily administered instruments" (Siehl & Martin, 1988: 80). In order to capture the ambiguity and contradiction of cultural beliefs, however, the construction of such instruments must be well grounded in the tacit, everyday life of the organization. They propose a two-stage inquiry process where indepth, long-term ethnographic fieldwork results in rich description of the content of the culture. These descriptive data are then used to construct a questionnaire which is subsequently administered to new organizational members or potential members to assess their "levels of cultural awareness and commitment" (Siehl & Martin, 1988: 82). In this example, construction of the questionnaire cannot proceed until the qualitative data-gathering has been completed; the questionnaire is developed from the fieldwork.

In the above example, the development purpose is achieved at the analysis stage but, we infer, the original design carefully stipulated this developmental sequence. Research and evaluation easily could be designed with an initial period of participant-observation followed by development and administration of a questionnaire. Applying the concept of "emergent design" (Lincoln & Guba, 1985; Jorgensen, 1989) from qualitative inquiry, however, extends this purpose to include the circumstances when initial design decisions are altered during the conduct of the study in a developmental fashion; these ways may not originally have been imagined or designed by the evaluators. Emergent development design occurs when initial design decisions did not include the developmental use of different methods; decisions to use additional, different methods emerged from early findings. A variation on this emergent development design would be the case where the initial design included development features but ones in addition to those originally planned for become included. Thus, the emergent development purpose suggests that evaluators (designers) respond and adapt their use of different methods to address increasing complexity as they learn more about a phenomenon (or learn what they do not know) and seek to more fully describe it.

Initiation

The fourth purpose is initiation, a term used by both Rossman and Wilson (1985) and Greene, Caracelli, and Graham (1989). Initiation designs are intended to uncover "paradox and contradiction" (Rossman & Wilson, 1985: 633), suggest ways to reframe the research question, and generally challenge the original conceptual framework of the evaluation. Moving well beyond the convergence orthodoxy of classical concepts of triangulation, initiation invites the unexpected.

As we describe below, previous mixed-methods work suggests that pure (or even somewhat tainted) examples of initiation at the design phase of research do not exist. Greene, Caracelli, and Graham (1989: 268) concur, noting that "purposeful initiation may well be rare in practice". It seems implausible that evaluators could design a study with the specific intent of uncovering the unexpected, the paradoxical, or the contradictory. These may well be the serendipitous results of analyses of data generated from different methods. Brewer and Hunter (1989: 19) conclude that "research findings that sharply and persistently diverge lead social scientists to rethink research problems".

An example of initiation at the analysis stage is offered by Rossman and Wilson (1985). As part of the evaluation of regional educational service agencies, it was reported that there are two strategies for promoting reform in education by the staff of these agencies: assistance and enforcement (see Firestone & Wilson, 1983). Interviews with staff in enforcement-oriented agencies revealed an ambivalence about the enforcement role. Indeed, many were trying to redefine their role as assistance. An unexpected explanation for that ambivalence came from the survey data where it was discovered that the vast majority of staff had previously been employed as teachers or administrators in the schools over which they now had enforcement powers. Thus, the quantitative data initiated a new interpretation that helped us understand why there was so much ambivalence about the enforcement role described in the interview data. This combination of methods gave the findings important new meaning.

These four purposes can be fruitfully applied at both the design and analysis stages of inquiry (with the exception of initiation designs). Before discussing the application of this typology to two recent policy studies, we clarify the distinction between the design and analysis stages of research.

Design and Analysis

Discussions and examples of combining methods at all stages of inquiry may well be useful for explicating the subtle inductive and deductive processes that result in powerful integration of multiple methods. During early conceptualization of the research when ideas are fluid and possibilities roam free, in more formal design decisions and justifications for those decisions, while gathering data and engaged in concomitant initial and tentative analyses of them, during more formal analytic procedures and drafts of written reports, as well as in the final text production where the results of combining methods are presented. detailed descriptions of false starts, stalls, great leaps of faith, wonderful insights, and steady sometimes tedious work could contribute substantially to the mixed-methods literature. We focus here on the design and analysis stages and hope that others will go further in developing examples for all stages of inquiry.

Much previous work has focused on the design stage, that phase of inquiry (often with an indeterminate beginning), when the kernel of an idea jumps out or, in the case of evaluations or policy studies, is presented.

Design incorporates this early work of idea generation and design possibilities, moves through the "funneling down" process as ideas become concrete decisions, and embraces the more formal design presented in the text of a proposal or research report. However, as noted above, in complex mixed-methods work, we find examples of what could be called "emergent design" where the researcher alters the original design based on new understandings or might well have conceived the study to have emergent elements. For these cases, the design phase incorporates emergent design processes too.

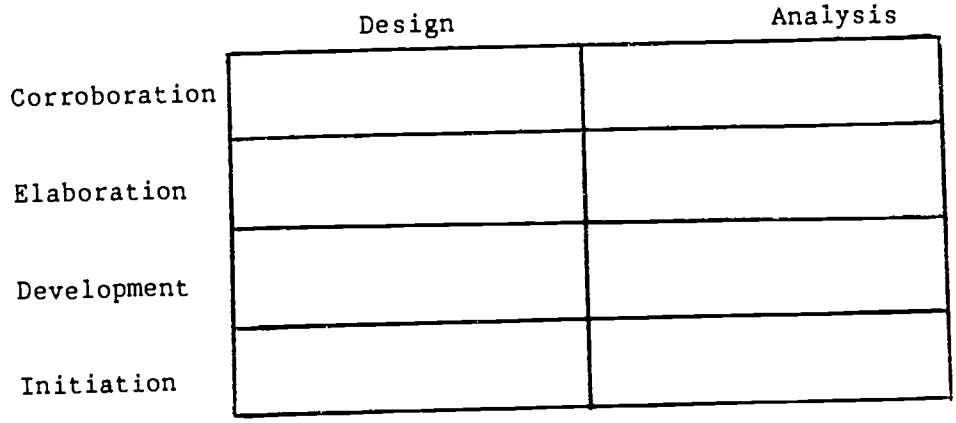
Working much as architects do, the researcher begins a process of iterative idea generation and refinement, as a blueprint for the investigation takes shape. Working back and forth between design possibilities and research questions, the researcher experiments conceptually with different elements, assesses their usefulness, soundness and fit with the emerging design, and decides what elements to include and which to discard. At this stage of the inquiry, the possibilities for combining methods are constrained by logistics, resources, researcher skills, and stakeholder preference. The design processes, then, result in a set of decisions -- a plan or blueprint -- for the conduct of the study.

Analysis is variously defined as the process of bringing order to the mass of gathered data (Schatzman & Strauss, 1973; Marshall & Rossman, 1989). It entails thinking that is "self-conscious, systematic, organized, and instrumental" (Schatzman & Strauss, 1973: 109) in which the researcher discovers the significant variables and their relationships. The data are organized into categories which can then be manipulated conceptually (qualitative analysis) or statistically (quantitative analysis). This phase of inquiry also embraces a range of processes. Early analysis begins as the researcher plays with ideas, poses research questions, and constructs conceptual frameworks. As the research questions are refined, methods chosen to help respond to them, and data gathered, the researcher engages with the phenomenon and begins to learn. These processes and decisions preclude some ways of thinking about the phenomenon and highlight others. The researcher implements more formal analytic strategies when data have been generated. It is during these more formal processes when discussions about combining multiple methods during data analysis are most salient.

The combination of purposes with stages of the research process produces a 4 x 2 matrix (see Figure 1). The remainder of this article is devoted to a discussion of the application of the typology to two mixed-methods policy studies conducted by researchers at Research for Better Schools, the mid-Atlantic regional education laboratory. Specific examples, drawn from both studies, are used to illustrate how each purpose can combine methods at both the design and analysis stages. Before a discussion of the studies, however, we would like to draw attention to the difficulties encountered in our analyses. Recall from the above discussion that the typology was derived by integrating the four purposes for combining methods with the two research stages of interest. In applying the typology, we found that we had to reconstruct the intent driving particular design decisions prior to their implementation, and struggled with distinguishing

between the design and analysis phases of the research. The difficulty of "retrospective rationality" (Staw, 1980) is a thorny one in making judgments about intentionality at the design state of any project. Unless original formal designs are available, we must trust the narrative constructed in a final report, when, with the perfect clarity of hindsight, we understand the rationality of previous decisions.

Figure 1: Typology of Purposes and Stage for Mixed-methods Research



With these caveats in mind, we turn to an overview of the two studies that form the empirical base for the examples of different mixed-methods purposes at work. The first is a study of state mandated testing; the second, of high school graduation requirements reform.

Description of the Studies

Examples of how methods can be effectively combined to inform important questions are drawn from two major policy studies recently completed by Research for Better Schools (Corbett & Wilson, 1991; Wilson, Rossman, & Adduci, 1991)[3]. Both studies focus on the question of the effects of state policies on local school districts. Over the course of the 1980s state-level policymakers have taken on increased responsibility for education reform. Faced with growing criticism for the failure of our schools to produce successful learners, states have actively sought ways to influence the quality of instruction. Two common state-level reform strategies included the the introduction of mandatory minimum competency tests as an accountability tool to assess the quality of instruction and the regulation of courses and course-taking patterns through stricter graduation requirements. Both broad policy initiatives were embraced widely during the 1980s when nearly two-thirds of the states mandated minimum competency tests and when 39 of the 43 states responsible for defining minimum graduation requirements increasing graduation credits. The two studies addressed the degree of impact that these policies had on local school districts.

Integral to both these investigations was a mixed-method approach to learn as much as possible about both intended and unintended effects. A combination of qualitative and quantitative methods was consciously included in the design and emerged as part of the analysis to help respond to the research questions. A brief overview of each study is offered to set a context for understanding the examples provided later.

Testing Study

This policy study (Corbett & Wilson, 1991) began as a comparison of local response to low-stakes and high-stakes testing conditions defined by two states. The "level of stakes" associated with a test is the extent to which students, teacher, administrators and/or parents perceive test performance to be "used to make important decisions that immediately and directly affect them" (Madaus, 1988: 86). Relatively minor consequences attended students' performance on Pennsylvania's minimum competency tests in language and mathematics. Maryland's high-stake strategy, on the other hand, required students to pass reading, writing, math and citizenship tests in order to receive a high school diploma. This distinction blurred during the study as Pennsylvania's testing program became more high-stakes when the Chief State School Officer released district rankings on the test and touted those results as an appropriate indicator of school effectiveness.

The research hypothesized that implementing a mandatory statewide testing program would stimulate adjustments in school district organization, technology, and culture at the local level -- particularly in high stakes situations. The greater the perception that mandatory testing programs would be used as inputs into important educational decisions, the more far-reaching the local ramifications would be for (a) the school district's goals and operating procedures, (b) what the district felt students should learn and the ways in which staff members accomplished this work, and (c) the values -- or definitions of desirable ways of operating -- already in place. An initial model was developed that suggested these system adjustments would vary as a function of the district context and the character of the state testing program. This model was altered as new data added additional insights.

The research was guided by two assumptions about implementation of state policies. The first assumption was that districts respond variably to policy initiatives; the second that the variability is largely a product of differences in the districts' organizational, political, economic, and cultural contexts. These assumptions had important implications for the methods employed. To document variability in response meant that the study had to have a large enough sample of sites to gauge the expected variability and to convince the reader that the variability was real rather than an artifact of selection. To understand how contextual conditions intertwined with school district personnel's behavior meant that the study had to give researchers sufficient time to talk to and observe people in their sites.

The study was conducted in three phases over a period of three years. First, a preliminary round of qualitative fieldwork was performed wherein researchers visited each of 12 school districts (six in each state) to interview a wide variety of staff members. Second, the results from the interviews were used to design a questionnaire administered throughout the districts in the states studied. Third, the survey results were used to structure a final round of feedback and interviews in the 12 sites originally visited.

Pathways Study

The policy study of changes in graduation requirements (Wilson, Rossman, & Adduci, 1991) was a four-year effort to investigate the effects of a new state policy in five high schools across the state of Maryland. The five schools were chosen because they are diverse and represent the full range of high schools Maryland's students might attend during their school careers. The overall research approach incorporated multiple methods in multiple sites to assess these effects. The policy was structured such that a phase-in was required with the policy being passed in 1986 but the effects not being fully in place until 1989. The researchers had access to data before, during and after the implementation of the policy so the full range of impact could be assessed.

The research was guided by two concurrent emphases, one a general interest in policy implementation and the other a focused inquiry into the effects of that policy on students' opportunity structure. The emphasis on policy implementation relied on the concept of backward mapping (Elmore, 1980) to track the effects of a policy. This approach assumes that the most profound knowledge can be gained through understanding the policy at the local level, the level at which it must be put into practice. By focusing on those who must implement the policy and their interpretation of the policy, a more complete picture of the policy's effects as well as important shifts in meaning can be obtained than if one tracked the policy from the top down. The design invested heavily in looking at individual school responses to the new policy with particular emphasis on the perspectives of classroom teachers. This bottom up focus did not ignore flow in the other direction. Indeed, an important part of the research involved unpacking the historic intent of the policy by interviewing the original framers of the policy. The flow was not only observed at both the top and the bottom; district and state department perspectives were also sought.

Whether there were important differences at the local level in responding to a uniform policy change offers important but incomplete insights about the policy. It is also important to explore the bottom line question of what impact these changes had on students' high school careers. The concept of opportunity structure was introduced to help focus that emphasis (Rosenbaum, 1978). Of interest were the immediate and more distant forces that influence students as they make decisions about their high school careers. These forces run the gamut from friends, family, printed

materials, teachers, counselors, school, district and even state policies. Additional internal pressures -- individual hopes and aspirations, fears and doubts, and expectations and beliefs about what he or she ought to be doing rounded out the concepts that guided the research questions.

Data collection was organized around six components outlined below:

1. A set of qualitative, indepth interviews were conducted with key state actors who provided recommendations for the state bylaws that framed the new requirements. This historical perspective is an effort to better understand the assumptions, values, and purposes of the new requirements.

2. A quantitative review of master schedules and course catalogues at the five high schools was conducted to document on a school level the changes in the quantity and character of the courses being offered. This was implemented over the course of a six year period prior to and during the implementation of the policy.

3. Interviews with key district personnel and document reviews of district materials were performed to assess the degree to which the new requirements produced significant change in local practice. This was necessary since several districts already had in place stricter requirements than those required by the state.

4. A set of qualitative, indepth interviews with students, teachers, counselors, and building administrators at each of the five high schools was conducted to document the effects of the new requirements. Three sets of interviews were conducted at each site: initial interviews shortly after the new policy was passed, interviews while the schools were in the middle of implementing new curricula and programs to respond to the policy, and a final round of interviews after the policy had been fully in place.

5. A quantitative, records-based analysis of student transcripts was performed to document course-taking patterns on an individual student level. Data were collected from a large sample of students at each school both prior to and after the new requirements had taken effect.

6. A set of semi-structured interviews were conducted with state colleges (2 and 4 year) and universities as well as local employers near the five schools to assess their views about the effects of the new requirements on the quality of recruits.

The Typology Applied

This section reviews each of the four mixed-methods research purposes and within each offers an example of how it was carried out in either the Pathways or Testing study. The exception is the initiation/design cell of the 4 x 2 matrix.

Corroboration

The term corroboration draws attention to ways the data converge, trying to address the same question, or lending support for one another. An example of corroboration design comes from the Pathways study where we investigated the question of whether the graduation requirements reform, as it shaped the high school curricular offerings, was affecting the size and composition of various departments. Quantitative findings in the Testing study led us at the analysis stage to turn to the qualitative data to seek confirmation of what the numbers were communicating.

Design. Previous research on changes in graduation standards found that some departments had shrunk dramatically over the decade of the 1980s, while others had grown. These findings came from the early California studies (Grossman, Kirst, Negash, Schmidt-Posner, & Garet, 1985) as well as from more recent multi-state policy studies (McDonnell, 1988; Clune, 1989). Growth trends in mathematics and science departments were clear and consistent; substantial shrinkage in physical education and vocationally-oriented departments was large; but patterns in other departments such as foreign language, art, music, and social studies were mixed.

We wanted to investigate whether teachers in the five representative Maryland high schools felt their jobs were at risk as a result of the new graduation requirements. We thought we might find evidence of substantial shifting between departments with teachers in some departments being laid off or in fear of being laid off. To address these research questions, we included a design component where teacher interview data could be corroborated with master schedule data. The intent was to use unobtrusive, archival data gathered annually by most schools to assess actual class size within departments and numbers of classes given by each department. To the extent that these data were available historically, we could assess growth and shrinkage across departments in each school over time.

We planned to use these data in conjunction with data generated from teacher interviews, also available over time. The interviews would provide detailed, qualitative descriptions of teachers' perceptions about changes in their departments and across the whole school, allowing us to glimpse interdepartmental conflict and competition if present. Thus, we intended that the corroborative design would allow interesting comparisons of data generated through distinctive methods assessing the same phenomenon.

Analysis. In the Testing study survey results addressed the issue of local variation in response to state mandated tests. Wide variations were documented in curricular and instructional adjustments across districts. Regression analyses resulted in factors explaining this variation. One of the most important explanatory variables was the "political climate" or how positive the relationship was between the district and the state. Specifically, positive relationships were associated with more curricular adjustments. At the analysis stage we turned to the third phase interviews to seek convergence for these findings. Corroborative examples from both

ends of the response continuum were found. For example, in one district where no changes occured in the curriculum, a district administrator characterized the district/state relationship in this way: "The state has become someone we have to beat rather than a partner to work with". On the other extreme was a district that accepted the state's increased role in monitoring educational outcomes and worked hard to find creative instructional techniques to improve student performance. The qualitative descriptions of how these two districts responded to the state mandate corroborated and offered convergence to the quantitative findings.

Elaboration

The mixed-methods purpose of elaboration maps the Green, Caracelli, and Graham (1989) category of complementarity onto our original term. This is a frequently-found purpose for combining methods, where data from one source elaborate, extend, clarify, illuminate or help interpret data from another method. For this purpose, qualitative methods are often incorporated into a study to "put meat on the bones" of quantitative findings, most often a questionnaire; qualitative data are used to help interpret survey responses. It is also possible (although less frequently found) that data generated through quantitative measures would be used to illuminate or extend understanding of indepth, perhaps ethnographic themes or patterns. This could occur through the assessment of the frequency of patterns of beliefs or values, as in the Siehl and Martin work (1988) mentioned above. Elaboration at the design stage is illustrated by the Testing study's use of interview data to elaborate survey findings. The example at the analysis stage comes from the Pathways study where demographic data enriched teacher reports and formal documents enhanced understanding of transcript analyses.

Design. The Testing study offered a useful example where interview data collected during the third phase of the study enriched quantitative findings from the second phase survey. During phase two of the research design a questionnaire was administered to 300 districts across the two states. A wealth of information was assembled as a result of that data collection effort. However, the researchers acknowledged in the design that these quantitative results would not tell the full story. A problem associated with quantitative research is that researchers cannot be sure that the meanings they attach to words on a survey and to the resulting statistical summaries are similar to those held by respondents; the data have become de-contextualized.

To compensate for this potential problem the researchers built into the design a third phase where we returned to the original twelve sites and sought educators' reactions to the survey results. Essentially, we summarized key survey findings and then asked for their interpretations. The qualitative responses provided detail and enhancement to the numbers from the survey. For example, we found that the majority of the survey respondents indicated that the curriculum had improved as a result of the state testing program. Yet, when we probed this in the followup interviews, we found that most of the interpretations of "improvement" focused on tightening of the curriculum with adjectives like "structured," "more

focused," "coordinated," "more systematic". Thus, the qualitative results offered an elaboration not available if just the quantitative results were analyzed in isolation.

Analysis. Several examples of analysis strategies for the purpose of elaborating data from one method with data from another come from the Pathways study. We routinely incorporated data obtained during site visits into the analysis process; these data provided insights into and explanations of events and teachers' perspectives. For example, in discussing the differences among teachers' reports about curriculum changes at the five high schools, we incorporated information about schedules, enrollment trends, and student population served. These data all were obtained during fieldwork site visits and were used to complement interview data.

Another example from Pathways is found in the analysis of transcript data. We analyzed whether students were earning more credits overall after implementation of the new graduation requirements and whether these differences varied by academic track. The quantitative measure of credits yielded the finding that, for the class of 1989 (those graduating under the new requirements), students in the general track took fewer credits than those in other tracks. The differences among tracks were greatest at the one high school with a magnet program. Using data gleaned from our site visits, through casual conversations with administrators, and through formal documents describing the magnet program, we were able to explain the large differences at this high school. The qualitative data were brought into the analysis to help extend understanding of the quantitative patterns.

Development

Greene, Caracelli & Graham (1989) define development as the process by which the results from one method help to inform the development of the other method (p. 259). A useful design example from the Testing study involved data from the qualitative interviews in phase one of the research being used to develop a questionnaire which was the basis for the quantitative data collected in phase two. The Pathways study offers an analysis stage example where interview data helped shape the development of the questions to be asked of student transcript data.

Design. A review of existing research on the local effects of state testing programs revealed that there was very little empirical information to assist us in the development of a questionnaire. Armed with this knowledge, we thought it appropriate to build into the design some investigative work that would help us better understand the issues before attempting to construct a survey. Thus, we conducted interviews at six districts in each of the two states under study.

A researcher spent two-to-three days in each district with the first day in the central office talking with the superintendent, the person responsible for handling the testing program, and other district staff who dealt with the test. School interviews were conducted with administrators,

guidance counselors, teachers and students. The guiding principle in the interviews was to get informants to talk about what was important to them about implementing statewide testing programs. Data summary charts were produced for each of the twelve districts. These findings provided the outline for the development of the survey instrument. The interview data revealed five themes around which a series of questions were developed. The themes included: the context in which the districts operated, the responses the districts made to tests, the strategies they employed to carry out those responses, the uses to which the tests were put, and the impacts of the tests. In addition to the major themes around which the questions were organized, the interview data also offered specific information for the wording of survey questions. Thus, the qualitative data in phase one of the study were designed to inform the development of the survey in phase two.

Analysis. The original Pathways design called for substantial interviewing and collection of student transcript data; these were the primary elements of the design. While we had framed general questions for the transcript analyses, specific questions did not emerge until we had considerable understanding of the issues at each site and had begun to generate the conceptual framework for the final report. Analysis of early interviews directly shaped the analytic questions for the transcript data. In particular, the interview data helped develop a focus on opportunity structure and how graduation requirements reform was affecting the opportunities of at-risk youth.

Additionally, early interviews with students led us to question the rigidity of tracking structures. Because some students talked with us about the variety of classes they took, we became less sure that a strict three-tiered classification system accurately reflected the reality of course-taking patterns in the five high schools. As a result, we began to play with alternative empirical measures of track and devised an algorithm that more accurately and discretely described students' course-taking patterns. Had we not allowed the interviews to shape subsequent analyses of transcript data, we would not have developed this new algorithm.

Initiation

Initiation is defined as the discovery of paradox or contradiction that leads to a reinterpretation of results. Such a purpose is difficult to conceive at the design stage, and we remain unconvinced that design decisions can intentionally build attempts to capture the paradoxical and the contradictory. These are the results of analysis and therefore we have some difficulty with this category.

Initiation at the analysis stage is illustrated by an example from the Testing study where the intent of the state policy (mandated testing program) was to pressure local districts in a positive manner to use the test results in a systematic examination of educational purpose, process or structure. Such an examination should inform which actions could be taken to improve the quality of instruction.

Analysis. What we discovered in the analysis phase of the Testing study, unanticipated in the original design, was that rather than conformity being the primary local response (as was the policy intent), in many cases rebellion occurred. Instead of accepting the state test as a valid indicator of student mastery of basic skills and using the results as a basis for making improvements, several districts rejected the goals of the state policy and replaced them with their own. These "rebellious" districts replaced the intended goals of accepting the tests as valid indicators of student learning and using the scores as the basis for long-term improvement with a short-term goal of raising test scores to diffuse community pressure. Thus, an important contradiction emerged during the analysis between intentions and local actions, and as a result new meaning emerged about the local effects of state policy.

Conclusions

Mixing methods in evaluation research and policy studies holds promise for strengthening our understanding of complex social phenomena and hence, the possibilities for improving social conditions of nagging intractability. When we illuminate that complexity through multiple lenses, we see more facets than when we only use one.

Although we have not tested this idea empirically, it appears that mixed-methods work is more prevalent in evaluations than in policy studies or research. The pragmatic demand of evaluations to address many aspects of a program or an intervention in order to respond to questions from stakeholders may have led to more eclectic methodological approaches. Our sense is that many policy studies reported in monograph form also rely on multiple methods. Similar demands to assess many facets of policy implementation and effects may operate here. These examples should be highlighted in the literature and discussed in more detail.

The conduct of mixed-methods research works more smoothly when certain conditions obtain. The first is that more than one researcher be engaged in the research. Each researcher brings particular methodologic expertise to the endeavor and the pooling of those skills offers a more comprehensive perspective. However, in order for that to work each expertise must be acknowledged by the other members of the team so that disputes over authority to resolve issues are minimized.

The second condition is that norms of respect and collegiality should prevail. When we generate ideas and let them float to see if they make sense, we as scholars are vulnerable to criticism or ridicule. We expose our thinking -- our most valuable asset. If norms of respect and collegiality do not develop, the ideas framing the study, shaping the methodologic mix, and building conclusions will be impoverished.

The final condition is an attitude of healthy skepticism about both theory and method (see Brewer & Hunter, 1989, chapter 2 for a discussion). We should be skeptical about quick fixes, the one best model, and panaceas, whether for practice or research. A curious change seems to be afoot in the

discourse about research, educational policy, and educational practice. In research we hear of the value of multiple methods and the need to describe multiple perspectives. State policies promote local experimentation, autonomy, and control. And classroom practices foster constructivist approaches to knowledge and stress students as active learners. All these trends express a weariness with ideological harangues whether these have exhorted that one particular paradigm is clearly superior to another (whether positivists, interpretive, or critical -- advocates of each make these claims); or have demanded that a particular set of authority relations should have dominance (e.g. the state defining particular outcomes for school districts); or have claimed miracles in classroom practice after a half-day workshop (e.g. the conversion-like experiences of "Hunterized" teachers). The authoritarian view of the one best way is being challanged. Educators are reclaiming responsibility to conduct business in ways that meet local needs.

From our perspective, we see a similar movement in the research and evaluation community. Several years ago Lortie (1982) called upon a group of budding scholars to be "shamelessly eclectic" in using concepts from various disciplines to frame inquiry and insight. We echo that call here: we should be shamelessly eclectic in our use of methods to understand the intractable and persistent problems of education today.

Endnotes

1. This phrase comes to us from Lortie (1982).

2. The Greene, Caravelli, and Graham typology includes a fifth purpose for mixed-methods evaluations: expansion. Closely linked to their purpose of complementarity (what we call elaboration), expansion designs are characterized by the use of distinct methods to elaborate the "scope, breadth, and range" (Greene, Caravelli, & Graham, 1989: 269) of the study. This is seen most clearly in the use of qualitative methods to evaluate program prcesses while quantitative methods assess program outcomes. Thus, different methods in fact assess distinct phenomena. In applying this purpose and design concept to our policy research, we struggled with the notion of distinct phenomena. In personal communication with one of the authors (Greene, 1991), we decided that this fifth purpose may not fit policy research as easily as evaluation studies. However, space does not permit a full explication of the differences between evaluations and policy studies here.

3. Rather than reference the two studies each time they are mentioned in the text, a single citation is offered only at the outset.

References

Brewer, J. & Hunter, A. (1989). Multimethod research: A synthesis of styles. Newbury Park, CA: Sage.

Burrell, G., & Morgan, G. (1979). Sociological paradigms and organizational analysis. London: Heinemann.

Clune, W. H., with White, P., & Patterson, J. (1989). The implementation and effects of high school graduation requirements: First steps toward curricular reform. CPRE Research Report Series RR-011. New Brunswick, NJ: Center for Policy Research in Education.

Corbett, H. D., & Wilson, B. L. (1991). Testing, reform, and rebellion. Norwood, NJ: Ablex.

Elmore, R. (1980). Backward mapping: Implementation research and policy decisions. Political Science Quarterly, 94, 601-616.

Firestone, W. A., & Wilson, B. L. (1983). Assistance and enforcement: Strategies for knowledge transfer and program reform. Knowledge: Creation, Diffusion, Utilization, 4, 429-452.

Fry, G., Chantavanich, S., & Chantavanich, A. (1981). Merging qualitative and qualitative research techniques: Toward a new research paradigm. Anthropology and Education Quarterly, 12(5), 145-158.

Greene, G. C. (1991). Personal communication with authors.

Greene, G. C., Caracelli, V. J., & Graham, W. F. (1989). Towards a conceptual framework for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 11(3), (255-274).

Grossman, P., Kirst, M., Negash, W., Schmidt-Posner, J., & Garet, M. (1985). Curricular change in California comprehensive high schools. 1982-83 to 1986-87. PACE Policy Paper 85-7-4. Berkeley, CA: Policy Analysis for California Education.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. Administrative Science Quarterly, 24, 602-611.

Jorgensen, D. L. (1989). Participant observation: A methodology for human studies. Newbury Park, CA: Sage.

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

Lortie, D. C. (1982). Remarks made at a fireside chat at the annual meetings of the American Educational Research Association, Montreal.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), Critical issues in curriculum: 87th yearbook of the NSSE, Part 1. Chicago: University of Chicago Press.

Marshall, C., & Rossman, G. B. (1989). Designing qualitative research. Newbury Park, CA: Sage.

Mathison, S. (1988). Why triangluate? Educational researcher, 17(2), 13-17.

McDonnell, L. M. (1988). Coursework policy in five states and its implications for indicator development. Working paper. New Brunswick, NJ: Center for Policy Research in Education.

Rosenbaum, J. E. (1978). The structure of opportunity in school. Social Forces, 57(1), 236-256.

Rossman, G.B. & Wilson, B.L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. Evaluation Review, 9(5), 627-643.

Schatzman, L., & Strauss, A. L. (1973). Field research: Strategies for a natural sociology. Englewood Cliffs, NJ: Prentice-Hall.

Siehl, C. & Martin, J. (1988). Measuring organizational culture: Mixing qualitative and quantitative methods. In M. O. Jones, J. D. Moore, & R. C. Snyders (Eds.), Inside organizations: Understanding the human dimension, Newbury Park, CA: Sage.

Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. Educational Researcher, 12(3), 6-13.

Staw, B. M. (1980). Rationality and justification in organizational life. In B. M. Staw & L. L. Cummings (Eds.), Research in organizational behavior, Volume 2. Greenwich, CT: JAI Press.

Wilson, B. L., Rossman, G. B., & Adduci, L. (1991). Maryland's graduation requirements: Local effects of policy reform. Philadelphia: Research for Better Schools.