

Toward a Conceptual Framework for Mixed-Method Evaluation Designs

Jennifer C. Greene, Valerie J. Caracelli, and Wendy F. Graham
Cornell University

In recent years evaluators of educational and social programs have expanded their methodological repertoire with designs that include the use of both qualitative and quantitative methods. Such practice, however, needs to be grounded in a theory that can meaningfully guide the design and implementation of mixed-method evaluations. In this study, a mixed-method conceptual framework was developed from the theoretical literature and then refined through an analysis of 57 empirical mixed-method evaluations. Five purposes for mixed-method evaluations are identified in this conceptual framework: triangulation, complementarity, development, initiation, and expansion. For each of the five purposes, a recommended design is also presented in terms of seven relevant design characteristics. These design elements encompass issues about methods, the phenomena under investigation, paradigmatic framework, and criteria for implementation. In the empirical review, common misuse of the term triangulation was apparent in evaluations that stated such a purpose but did not employ an appropriate design. In addition, relatively few evaluations in this review integrated the different method types at the level of data analysis. Strategies for integrated data analysis are among the issues identified as priorities for further mixed-method work.

The inevitable organizational, political, and interpersonal challenges of program evaluation mandate the use of multiple tools from evaluators' full methodological repertoire (Cook, 1985; Mathison, 1988). In recent years, this repertoire has been considerably expanded with the acceptance of qualitative methods as appropriate, legitimate, and even preferred for a wide range of evaluation settings and problems. Concomitantly, evaluators have expressed renewed interest in mixed-method evaluation designs

employing both quantitative and qualitative methods (e.g., Cook & Reichardt, 1979; Madey, 1982; Rossman & Wilson, 1985; Smith & Louis, 1982). However, the territory of mixed-method designs remains largely uncharted. Of particular need is a clear differentiation of alternative purposes for mixing qualitative and quantitative methods and of alternative designs, analysis strategies, and contexts appropriate for each purpose (Greene & McClintock, 1985). For example, in current practice, quite different mixed-method designs are advocated and used in varied evaluation contexts for the common proclaimed purpose of *triangulation*. Such practice muddles the concept of triangulation as originally construed and remains insensitive to other possible benefits of mixed-method designs (Mathison, 1988). Further, just as careful planning and defensible rationales accompany the design and

An earlier version of this paper was presented as a panel at the 1988 Annual Meeting of the American Evaluation Association in New Orleans. The authors are indebted to Melvin Mark for his insightful and constructive comments on the work presented herein, both at the conference and in subsequent personal communications (Mark, 1988).

implementation of evaluation case studies, ethnographies, surveys, and quasi-experiments, so must similar thoughtfulness be given to the design and implementation of mixed-method studies.

Toward these ends, the present study contributes to the development of a conceptual framework, thus enabling more thoughtful and defensible mixed-method evaluative inquiry. In this study, we defined mixed-method designs as those that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to any particular inquiry paradigm. Through an analytic review of first theoretical and then empirical literature on mixed-method inquiry, this study generated valuable information on mixed-method *purposes* and *design characteristics*. Review procedures and findings for these two components of our mixed-method conceptual framework thus constitute the focus of the present discussion. Relatively little information was garnered relevant to other components of this framework, including the differential *utilization* of quantitative and qualitative information, *data analysis strategies* and *contexts* appropriate for mixed-method inquiries, as well as mixed-method *project management* and *resource* issues. These concerns are briefly discussed at the conclusion of the present article as issues warranting further work.

Theoretical Base

This study on mixed-method evaluation inquiry was grounded in an initial review of four theoretical starting points, selected for their conceptual attention to one or more of the key issues represented in our mixed-method conceptual framework.

Triangulation. (See Campbell & Fiske, 1959; Denzin, 1978; Webb, Campbell, Schwartz, & Sechrest, 1966; see also Mathison, 1988, for an excellent discussion of triangulation from these same sources.) From its classic sources, triangulation refers to the designed use of multiple methods, with offsetting or counteracting biases, in investigations of the same phenomenon in

order to strengthen the validity of inquiry results. The core premise of triangulation as a design strategy is that all methods have inherent biases and limitations, so use of only one method to assess a given phenomenon will inevitably yield biased and limited results. However, when two or more methods that have offsetting biases are used to assess a given phenomenon, and the results of these methods converge or corroborate one another, then the validity of inquiry findings is enhanced. As noted by Greene and McClintock (1985), this triangulation argument requires that the two or more methods be intentionally used to assess the same conceptual phenomenon, be therefore implemented simultaneously, and, to preserve their counteracting biases, also be implemented independently.

Multiplism. (See Cook, 1985; Mark & Shotland, 1987; Shotland & Mark, 1987.) Thomas Cook's critical multiplism acknowledges the decreased authority of social science theory and data in a postpositivist world and then seeks to reaffirm and strengthen the validity of, and thereby users' confidence in, empirical work by extending the basic logic of triangulation to all aspects of the inquiry process.

The fundamental postulate of multiplism is that when it is not clear which of several options for question generation or method choice is "correct," all of them should be selected so as to "triangulate" on the most useful or the most likely to be true Multiplism aims to foster truth by establishing correspondences across many different, but conceptually related, ways of posing a question and by ruling out whether any obtained correspondences are artifacts of any epiphenomena of value, substantive theory, or method choice that may have been inadvertently incorporated into individual tests. (Cook, 1985, pp. 38 and 46)

Congruent with the basic logic of triangulation, Cook's multiplism emphasizes enhanced validity via convergence of results from multiple methods, theoretical orientations, and political or value perspectives. Cook also acknowledges that the results of multiple methods may serve more comple-

mentary than convergent purposes, as when different methods are used for different components of a multitask study. Elaborating on this point, Mark and Shotland (1987) offer three different purposes for multiple-method designs: (a) triangulation, which seeks convergence of findings; (b) bracketing, which seeks a range of estimates on the correct answer (or triangulation with a confidence interval); and (c) complementarity, in which different methods are used to assess different study components or phenomena, to assess the plausibility of identified threats to validity, or to enhance the interpretability of assessments of a single phenomenon—for example, via broader content coverage or alternate levels of analysis.

Mixing methods and paradigms. (See Guba & Lincoln, 1984; Kidder & Fine, 1987; Reichardt & Cook, 1979; Rossman & Wilson, 1985; Smith, 1983; Smith & Heshusius, 1986.) This set of references was selected primarily for their common discussion of the following design issue: Are mixed-method evaluation designs, in which the qualitative and quantitative methods *are* linked to contrasting inquiry paradigms, meaningful, sensible, and useful? Rossman and Wilson (1985) outline a continuum of three stances on this issue: the purists, the situationalists, and the pragmatists.

The purists (including Guba & Lincoln, 1984; Smith, 1983; and Smith & Heshusius, 1986) answer with an unequivocal “no” to the issue posed. They argue that the attributes of a paradigm form a “synergistic set” that cannot be meaningfully segmented or divided up. Moreover, different paradigms typically embody incompatible assumptions about the nature of the world and what is important to know, for example, realist versus relativist ontologies. So, mixed-method evaluation designs, in which the qualitative and quantitative methods are conceptualized and implemented within different paradigms (characteristically, interpretive and postpositivist paradigms, respectively), are neither possible nor sensible.

In contrast, Reichardt and Cook (1979) argue pragmatically that paradigm attributes are logically independent and therefore can

be mixed and matched, in conjunction with methods choices, to achieve the combination most appropriate for a given inquiry problem. The practical demands of the problem are primary; inquirer flexibility and adaptiveness are needed to determine what will work best for a given problem. Or, in the pragmatic view of Miles and Huberman (1984), epistemological purity does not get research done.

The middle-ground situationalist position, articulated by Kidder and Fine (1987), retains the paradigmatic integrity stance of the purists but also argues, like the pragmatists, that our understanding of a given inquiry problem can be significantly enhanced by exploring convergences in stories generated from alternate paradigms. Congruent with Cook’s proposal for aggressive meta-analyses, Kidder and Fine suggest that such explorations occur across studies, in particular, across quantitative (postpositivist) and qualitative (interpretivist) studies. This strategy may yield “stories that converge” or discrepancies that invoke fresh perspectives and new, more illuminating explanations.

In a similar vein, Rossman and Wilson (1985) sought their own middle ground on this issue of mixing paradigms by outlining three functions for mixed methodology: (a) corroboration, as in establishing convergence; (b) elaboration, as in providing richness and detail; and (c) initiation, which “prompts new interpretations, suggests areas for further exploration, or recasts the entire research question. Initiation brings with it fresh insight and a feeling of the creative leap Rather than seeking confirmatory evidence, this [initiation] design searches for the provocative” (Rossman & Wilson, 1985, pp. 637 and 633).

Mixed-method design strategies. (See Greene, 1987; Greene & McClintock, 1985; Knapp, 1979; Madey, 1982; Mark & Shotland, 1987; Maxwell, Bashook, & Sandlow, 1986; Sieber, 1973; Trend, 1979.) This more diverse set of references was reviewed primarily for additional ideas on alternative mixed-method purposes and on design characteristics that may differentiate among these purposes. Building on the work of

Greene and McClintock (1985), Greene's (1987) synthesis of these ideas with those represented in the other three theoretical starting points served as a key foundation for the present conceptual work (and is thus incorporated within our later presentation of findings).

Empirical Review

Believing that sound conceptual work requires an interplay of theory and practice, we next conducted a comprehensive review of a purposive sample of 57 mixed-method evaluation studies. Our review guide included all of the components of our mixed-method conceptual framework (purpose, design characteristics, utilization, data analysis, contexts, management, and resources), with directed emphasis from the theoretical starting points on the first two.¹ The sample was purposive in that we aimed to identify studies in which the mixed-method aspect of the design was prominent and thus in concert with our research objectives. The search was limited to studies reported from 1980 to 1988. We also sought a broad representation of different approaches to evaluation, different kinds of evaluands, and different types of evaluation documents. Our final sample, which included 18 published evaluation studies, 17 evaluation reports, and 22 evaluation papers, met all of our sampling criteria except representation across evaluands. Compared with other data bases employed during sampling, ERIC yielded many more appropriate studies; hence, our sample tilted toward mixed-method evaluations conducted on educational programs.

Our reviews of these selected literatures on the theory and practice of mixed-method evaluation yielded most importantly a set of five different mixed-method purposes and seven relevant design characteristics. These results are presented in the following two sections.

Results for Mixed-Method Purposes

Theory. The five mixed-method purposes generated from our theoretical review are

presented in Table 1 and briefly elaborated below.

A mixed-method design with a *triangulation* intent seeks convergence in the classic sense of triangulation. The use of both a qualitative interview and a quantitative questionnaire to assess program participants' educational aspirations illustrates this triangulation intent. In conjunction with this intent, Shotland and Mark (1987) caution that different methods may be biased in the same direction or, in fact, may be asking different questions. Variations within this triangulation purpose include Campbell and Fiske's (1959) advocacy of multiple methods to evaluate discriminant as well as convergent validity, and Mark and Shotland's (1987) idea of using multiple methods to bracket rather than converge on the correct answer. This idea of triangulation with a confidence interval is drawn from Reichardt and Gollob (1987).

In a *complementarity* mixed-method study, qualitative and quantitative methods are used to measure overlapping but also different facets of a phenomenon, yielding an enriched, elaborated understanding of that phenomenon. This differs from the triangulation intent in that the logic of convergence requires that the different methods assess the same conceptual phenomenon. The complementarity intent can be illustrated by the use of a qualitative interview to measure the nature and level of program participants' educational aspirations, as well as *influences* on these aspirations, combined with a quantitative questionnaire to measure the nature, level, and *perceived ranking within peer group* of participants' educational aspirations. The two measures in this example are assessing similar, as well as different, aspects of the aspirations phenomenon. One variation within this complementarity intent is the use of different methods to assess different levels of a phenomenon (Mark & Shotland, 1987), which we characterized with the analogy of peeling the layers of an onion.

Sieber (1973) and Madey (1982), for sociological and evaluation contexts, respectively, provide many creative examples of

TABLE 1
Purposes for mixed-method evaluation designs

Purpose	Rationale	Key theoretical sources
TRIANGULATION seeks convergence, corroboration, correspondence of results from the different methods.	To increase the validity of constructs and inquiry results by counteracting or maximizing the heterogeneity of irrelevant sources of variance attributable especially to inherent method bias but also to inquirer bias, bias of substantive theory, biases of inquiry context.	Campbell & Fiske, 1959 Cook, 1985 Denzin, 1978 Shotland & Mark, 1987 Webb et al., 1966
COMPLEMENTARITY seeks elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method.	To increase the interpretability, meaningfulness, and validity of constructs and inquiry results by both capitalizing on inherent method strengths and counteracting inherent biases in methods and other sources.	Greene, 1987 Greene & McClintock, 1985 Mark & Shotland, 1987 Rossman & Wilson, 1985
DEVELOPMENT seeks to use the results from one method to help develop or inform the other method, where development is broadly construed to include sampling and implementation, as well as measurement decisions.	To increase the validity of constructs and inquiry results by capitalizing on inherent method strengths.	Madey, 1982 Sieber, 1973
INITIATION seeks the discovery of paradox and contradiction, new perspectives of frameworks, the recasting of questions or results from one method with questions or results from the other method.	To increase the breadth and depth of inquiry results and interpretations by analyzing them from the different perspectives of different methods and paradigms.	Kidder & Fine, 1987 Rossman & Wilson, 1985
EXPANSION seeks to extend the breadth and range of inquiry by using different methods for different inquiry components.	To increase the scope of inquiry by selecting the methods most appropriate for multiple inquiry components.	Madey, 1982 Mark & Shotland, 1987 Sieber, 1973

mixing methods for *development* purposes. All involve the sequential use of qualitative and quantitative methods, where the first method is used to help inform the development of the second. For example, a quantitative survey of program participants' educational aspirations could be used to identify a purposive sample for more in-depth interviews about these aspirations.

For a given mixed-method study, *initiation* as the discovery of paradox and fresh perspectives may well emerge rather than constitute a planned intent. However, in complex studies, as well as across studies, both consistencies and discrepancies in qualitative compared with quantitative findings can be intentionally analyzed for fresh insights invoked by means of contradiction and paradox.

A mixed-method study with an *expansion* intent is a "multitask" study in Cook's (1985) multiplism framework or a study that aims for scope and breadth by including multiple components. In evaluation contexts, this mixed-method expansion purpose is commonly illustrated by the use of qualitative methods to assess program processes and by quantitative methods to assess program outcomes.

Practice. Our empirical review results substantially confirmed this conceptualization of mixed-method purposes. For all studies with a discernible rationale for mixing methods, this rationale matched one or more of these five purposes. Hence, we offer this set of purposes as representing both the practice and potential of mixed-method evaluation strategies (see also Smith, 1986) and as progress toward a common parlance for conceptualizing and describing mixed-method rationales in program evaluation.²

In the empirical review, mixed-method purposes were tabulated both according to the study authors' statement of purpose and by our definitions. As shown in Table 2, the authors' stated primary or secondary purpose for using a mixed-method design was often triangulation (23%) or expansion (26%). However, in a similar proportion of evaluations, no purpose for the mixed-method design was stated or could be readily

inferred. By our definitions, four fifths of the primary purposes and one half of the 70 total purposes were either complementarity (not triangulation) or expansion.

The more interesting finding in Table 2 is the backward Z pattern formed by this cross-tabulation. The diagonal represents agreement between the authors' statements and our definitions of mixed-method purposes. For example, five of the empirical studies reviewed had a primary or secondary purpose of triangulation (upper left cell) according to both determinations. When there was such agreement about purpose, the authors were usually very explicit in their stated rationale for the particular mixed-method design chosen. For example, in an evaluation of a physical education project, "the data were examined . . . and presented employing the processes of triangulation and corroboration in order to arrive at valid and reliable statements" (Moody, 1982, Abstract). Additional illustrations of this diagonal, or instances of our mixed-method purposes in evaluation practice, follow.

The evaluation instruments were designed to give overlapping [complementarity] and cross checking [triangulation] assessments of the perceptions of those involved. (Peters, Marshall, & Shaw, 1986, p. 16)

Overall, the methodologies used confirmed that any paper-and-pencil instrument ought to be supplemented by qualitative methods. This would enrich and provide depth to the statistical data obtained. (Martin, 1987, pp. 14–15) [complementarity]

Quantitative methods can establish the degree to which perceptions are shared, but uncovering the perceptions themselves must be [first] done naturalistically. (Gray & Costello, 1987, p. 12) [development]

Qualitative in addition to quantitative methods were included so the evaluation could "tell the full story." (Hall, 1981, p. 127) [expansion]

The whole is greater than the sum of the parts when qualitative and quantitative approaches and methods are combined. (Smith, 1986, p. 37) [initiation]

The horizontal lines forming the top and bottom of the backward Z show disagree-

TABLE 2

Empirical review: Crosstabulation of mixed-method purposes as stated and by our definition

Authors' statement	By our definition										Totals					
	Triangulation		Complementarity		Development		Initiation		Expansion		P		No.	%	s	% all
	P	s	P	s	P	s	P	s	P	s						
Triangulation	3	2	5	2	2			1	1			11	19	5	23	
Complementarity			8	2								8	14	2	14	
Development					5	1						5	9	1	9	
Initiation							1	1				1	2	1	3	
Expansion									16	2		16	28	2	26	
Not stated			5	1			1	1	10			16	28	2	26	
Totals	3 (5%)	2	18 (32%)	5	7 (12%)	1	2 (4%)	3	27 (47%)	2		57		13		
% all	7%		33%		11%		7%		41%							

Note. P = primary purpose; s = secondary purpose.

ment between the authors' stated intentions for the mixed-method design and our determination of purposes. This discrepancy is of one of two types: Either the authors stated triangulation as the purpose for the mixed-method design when it was not, or the authors did not state a purpose when we were able to identify one. The latter discrepancy is difficult to illustrate because of the absence of a stated purpose by the authors. An excerpt from Moran (1987) illustrates the first discrepancy. This evaluator stated triangulation as the purpose for the mixed-method design, when we identified primary and secondary purposes of development and initiation, respectively.

Some researchers maintain that a two-tiered methodology is not really a *triangulation*. Greene and McClintock (1985) contend that . . . "a nonindependent, sequential mixed-method strategy loses the capacity for triangulation. In this strategy, the methods are deliberately interactive, not independent, and they are applied singly over time so that they may or may not be measuring the same phenomenon." Given the dynamic nature of the public service, it is difficult to discern how any evaluation routine could meet the[se] criticisms The idea behind an interactive sequential methodology is not to measure the same phenomenon at the same time, but to use the findings of one methodology to inform the issues to be addressed in the subsequent evaluation. Under this construct, qualitative data are employed to ensure that the quantitative study is current. Quantitative data in turn are used to reformulate the issues for the qualitative study. (Moran, 1987, pp. 623–624, emphases added)

Results for Mixed-Method Design Characteristics

The seven characteristics of mixed-method designs presented in Table 3 represent an integration of results from our theoretical and empirical reviews. Although the empirical review did not alter the initial set of theoretically-derived design characteristics,³ it did serve to refine and clarify our conceptualization of each. Nonetheless, we

do not consider this set of mixed-method design characteristics to be exhaustive, but rather we anticipate future refinements and additions. Brief descriptions of the seven mixed-method design characteristics generated in this study follow. Empirical results for these design characteristics are presented in the next section, differentiated by primary mixed-method purpose.

Methods. The methods characteristic represents the degree to which the qualitative and quantitative methods selected for a given study are similar to or different from one another in form, assumptions, strengths, and limitations or biases (as argued by Campbell & Fiske, 1959). A scaled questionnaire and structured interview would be considered similar, whereas an achievement test and open-ended interview would be considered different. Mid-range positions can occur when the methods share some characteristics, especially biases, but not others, as in the combined use of a quantitative written questionnaire and a qualitative critical incident (also written) diary.

Phenomena. The term *phenomena* refers to the degree to which the qualitative and quantitative methods are intended to assess totally different phenomena or exactly the same phenomenon. When different methods are implemented to assess different phenomena, the methods are usually responding to different questions. To illustrate, quantitative measures like standardized achievement tests are often used to assess the degree of success of an educational program, and qualitative measures such as interviews and observations are used to understand how and why a program is successful or unsuccessful.

Mid-range phenomena positions occur when qualitative and quantitative methods overlap in their intent, yet also capitalize on the strengths of one or both methods to secure additional information. For example, Smith and Robbins (1984) used quantitative surveys to provide a detailed picture of the nature, causes, and consequences of parental involvement in four different federal programs. A qualitative site review, which included interviews, observations, and docu-

TABLE 3
Mixed-method design characteristics

Characteristic	Refinement				Key		
Methods	Similar	_____				Different	
Phenomena	Different	A	B	C	D	Same	<p>A = totally different phenomena</p> <p>B = dissimilar but related phenomena</p> <p>C = overlapping phenomena or different facets, dimensions of a single phenomenon</p> <p>D = exactly the same phenomenon</p>
Paradigms	Different	A	B	C	D	Same	<p>A = all qualitative methods in one paradigm, all quantitative methods in another</p> <p>B = all qualitative <i>or</i> quantitative methods in one paradigm, most of the other methods in another</p> <p>C = most methods of both types in one paradigm, a few methods of one or both types in another</p> <p>D = all methods in the same paradigm</p>
Status	Unequal	_____				Equal	
Implementation: Independence	Interactive	_____				Independent	
Implementation: Timing	Sequential	A	B	C	D	Simultaneous	<p>A = sequential: different methods implemented sequentially</p> <p>B = bracketed: one method implemented before and after the other</p> <p>C = concurrent: one method implemented within the time frame spanned by implementation of the other</p> <p>D = simultaneous: different methods implemented simultaneously</p> <p>E = not applicable: one method represents the use of existing data</p>
Study	> One study	_____				One study	

ment analysis, was intended to secure similar information on parental involvement, as well as additional information about the effects of parental involvement in varied program settings (e.g., rural vs. urban).

Paradigms. The design characteristic labeled *paradigms* refers to the degree to which the different method types are implemented within the same or different paradigms. We recognize that any given pair of quantitative and qualitative methods either is or is not implemented within the same paradigm, rendering this design characteristic dichotomous. Evaluation practice, however, commonly includes multiple methods of both types. Thus, the ratings in Table 3 are intended to be holistic, representing the degree to which the whole set of methods is conceptualized, designed, and implemented within the same or different epistemological frameworks. Assessments of this design element should be made independently of the relative number and status of qualitative versus quantitative methods.

Status. This characteristic represents the degree to which a study's qualitative and quantitative methods have equally important or central roles vis-à-vis the study's overall objectives. In contrast to paradigms, the status design characteristic should directly reflect the relative weight and influence of the qualitative and quantitative methods with respect to their frequency and their centrality to study objectives.

Implementation: Independence. The degree to which the qualitative and quantitative methods are conceptualized, designed, and implemented interactively or independently can be viewed on a continuum. Sometimes a study includes both components, representing a mid-range position. For example, in part of Louis's (1981) study, mixed-method implementation was independent: Standardized data collection by central project staff occurred simultaneously with the development of 42 miniethnographies by field staff, who worked without knowledge of the central staff's emerging findings. Part of Louis's study was also interactive: During analysis and interpretation, every individual who contributed as a

major author or analyst to the study was familiar with all data available.

Implementation: Timing. Although we represent this characteristic as a continuum, we again recognize that a given pair of methods is typically implemented concurrently or sequentially, not in between. Yet, a short quantitative method could be paired with a longer qualitative method, or pre-post tests could be implemented before and after participant observation (illustrating, from Table 3, "concurrent" and "bracketed" timing, respectively). Variation on this design element also arises from the use of multiple methods within a mixed set. With reflection we refined this characteristic by dividing it into categories (see Table 3) that could be assessed for a whole set of mixed methods or, as appropriate, for each pair of methods.

Study. The final design characteristic labeled *study* is essentially categorical. The empirical research either encompassed one study or more than one study. Although our own review yielded little variation on this design characteristic (all but four evaluations represented a single study), it remains an important consideration for continued discussion of mixed-method designs (Cook, 1985; Kidder & Fine, 1987).

Mixed-Method Purposes × Design Characteristics: Recommended Designs

To review, the long-range goal of this study is the development of a conceptual framework that could inform and guide the practice of mixed-method inquiry. Such guidance would include a description of the kind of design (and analysis, context, etc.) most appropriate for a given mixed-method purpose. For this reason, we analyzed the empirical review results on mixed-method design characteristics separately for studies grouped by *our* definition of primary purpose. This analysis is presented in Figure 1.

Each row of Figure 1 represents one of our five mixed-method purposes; each column presents a single design characteristic and the scale by which it was rated. The five points in these scales correspond to the following ratings of these design elements which we viewed as continua during our

PURPOSE (N=57)

Methods 1—5
Similar Different

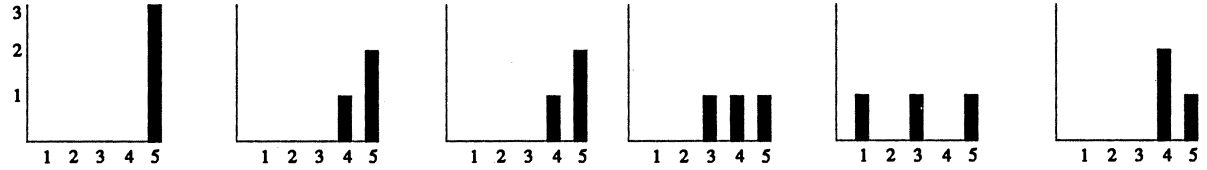
Phenomena 1—5
Different Same

Paradigms 1—5
Different Same

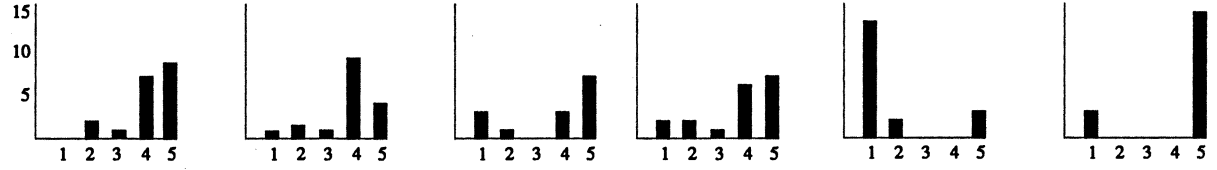
Status 1—5
Unequal Equal

Implementation 1—5
Interactive Independent Sequential Concurrent

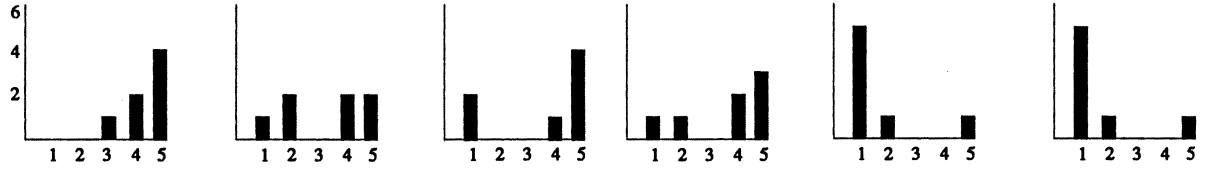
TRIANGULATION
(n=3)



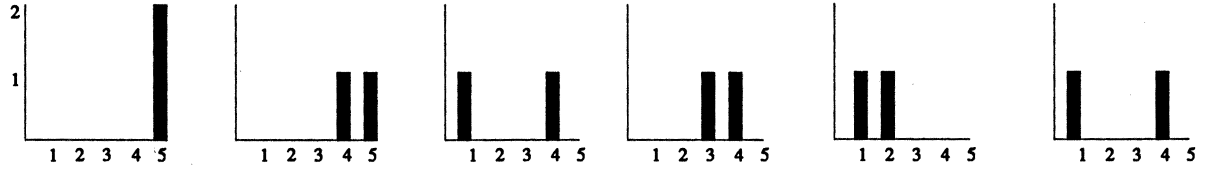
COMPLEMENTARITY
(n=18)



DEVELOPMENT
(n=7)



INITIATION
(n=2)



EXPANSION
(n=27)

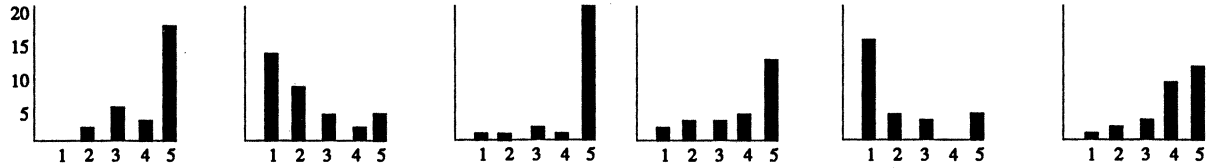


FIGURE 1. Mixed-method purposes by design characteristics

Note. Study not reported because only four evaluations included more than one study. In a few studies, some design elements could not be rated because of missing information.

empirical review: at either end (1 and 5), near either end (2 and 4), near the middle (3). Each cell entry in Figure 1 thus displays the distribution of our ratings on a single design characteristic for a given mixed-method purpose. For example, the graph in the upper left-hand cell shows that the qualitative and quantitative methods were rated *different* (a score of 5) in all three evaluations with a triangulation purpose.

Incorporating these empirical review results on design characteristics, this section heuristically presents a recommended design for each of the identified mixed-method purposes. There are three caveats to keep in mind as these recommendations are presented. First, the importance of these seven characteristics to mixed-method designs is generally supported by our empirical review. Nonetheless, we have greater confidence in our definitions of mixed-method purposes and consider elements of mixed-method design choice an open area of investigation. Second, mixed-method strategies are often guided by more than one purpose. Thus, designs will not appear as pristine on these characteristics in practice as we have set them forth here. Third, we acknowledge that some departures from these recommended designs can be readily defended. Mark (1988), for example, suggested that for a triangulation design, the different methods need not be implemented simultaneously if the phenomenon of interest is stable over time. In short, we present these recommended designs to underscore the importance of design element choice in mixed-method frameworks, but we present them as working ideas rather than prescriptive models.

Figure 2 profiles our five recommended mixed-method designs. In this figure, each letter represents a different mixed-method purpose. Individual letters denote a recommended position on a design characteristic for a particular purpose. Letters with bars indicate that the recommended position can range somewhat. The omission of a letter means that a *specific* position on a characteristic is not warranted.

Triangulation (T) design. The combined

use of quantitative and qualitative methods for the purpose of triangulation dominates current discussions about mixed-method rationales. Yet, as indicated by our empirical review (see Table 2), methodological triangulation in its classic sense is actually quite rare in mixed-method practice. Our recommended triangulation design is based on the logic of convergence embedded in the classic conceptualization of triangulation.

This logic requires that the quantitative and qualitative methods be different from one another with respect to their inherent strengths and limitations/biases and that both method types be used to assess the same phenomenon. Methods that are biased in the same direction or that ask/answer different questions can undermine the triangulation logic and result in spurious inferences (Shotland & Mark, 1987). Relatedly, the methods need to be conceptualized, designed, and implemented within the same paradigmatic framework (Greene & McClintock, 1985; Kidder & Fine, 1987). Strong between-methods triangulation is also enhanced when the status of the different methods—that is, their relative weight and influence—is equal and when the quantitative and qualitative study components are implemented independently and simultaneously. Across mixed-method purposes, the recommended independent implementation of the different methods is unique to triangulation.

Complementarity (C) design. One apparently common purpose for combining qualitative and quantitative methods is to use the results from one method to elaborate, enhance, or illustrate the results from the other. The recommended complementarity design depicted in Figure 2 is similar to the triangulation design, with the exception of the phenomena and implementation-independence characteristics. The phenomena characteristic has a slight range, indicating that the quantitative and qualitative methods should be used to examine overlapping phenomena or different facets of a single phenomenon. In complementarity designs, the paradigmatic framework for both types of methods should also be similar, and in-

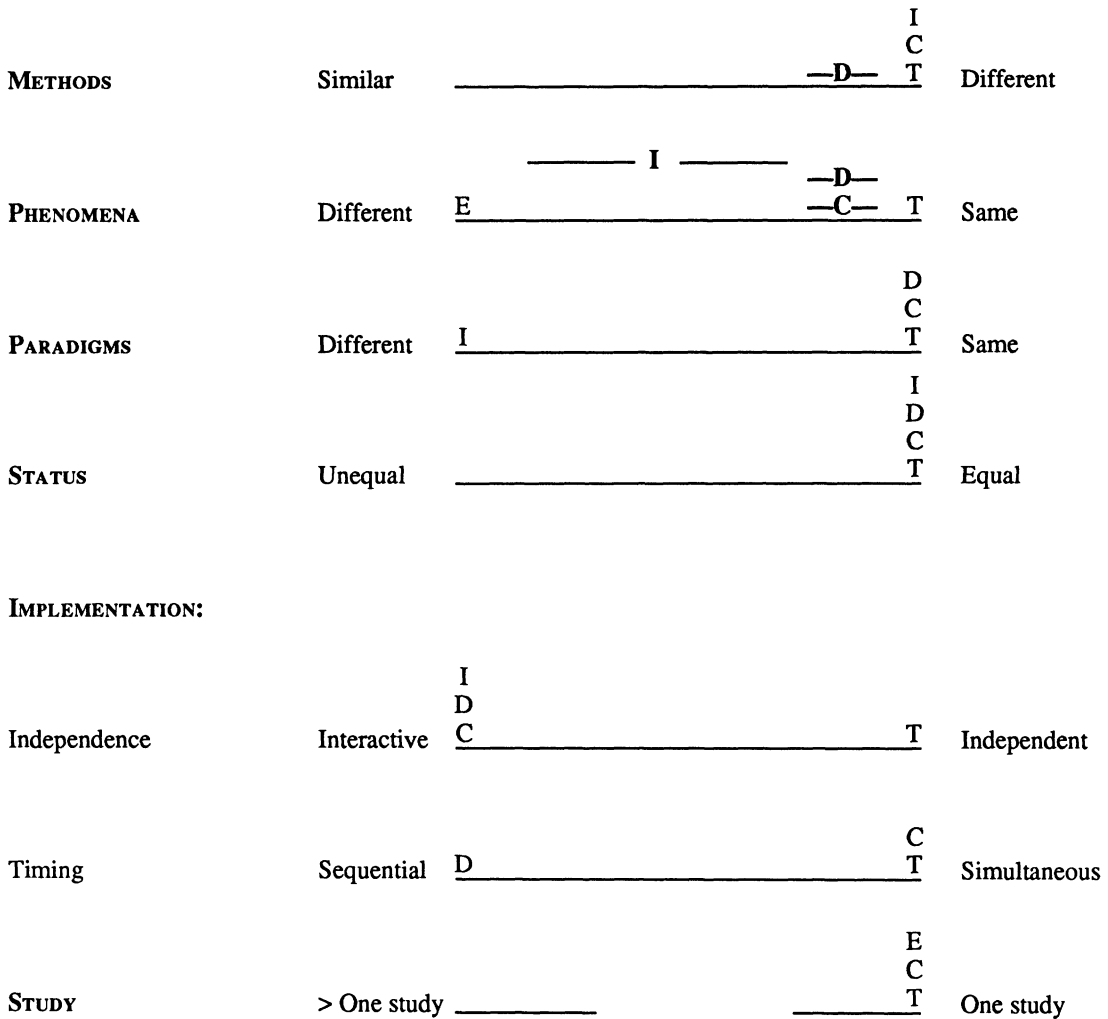


FIGURE 2. Recommended mixed-method designs

Note. T = triangulation; C = complementarity; D = development; I = initiation; E = expansion.

terpretability is best enhanced when the methods are implemented simultaneously and interactively within a single study.

In our empirical review there were 18 mixed-method evaluations implemented for the primary purpose of complementarity. A comparison of the design characteristics of these 18 studies (see Figure 1) with our recommended complementarity design yields considerable congruence. On each design characteristic with the exception of status, approximately three fourths of these mixed-method studies were judged to be at or close to our recommended position. Somewhat more variability was evident for status. This

congruence of theory with practice supports and encourages both.

Development (D) design. The salient feature of our recommended development design is the sequential timing of the implementation of the different methods. One method is implemented first, and the results are used to help select the sample, develop the instrument, or inform the analysis for the other method. By definition, then, implementation is also interactive, and the different methods are used to assess the same or similar phenomena, conceptualized within the same paradigm. We further maintain that strong development designs use

dissimilar methods of equal status. Mixed-method studies with a development intent can occur within a single study or across studies, conducted sequentially to capitalize on the benefits derived from each method type.

Like the theory-to-practice comparison for complementarity, the design characteristics of the seven empirical mixed-method studies conducted for purposes of development (see Figure 1) were quite congruent with this recommended design. The designs of five or six of these studies were at or close to the recommended position on all characteristics except phenomena. Surprisingly, for this design element, three studies used mixed methods to assess different rather than similar phenomena. An interesting variation was found in one study that implemented the different methods simultaneously rather than sequentially (Bower, Anderson, & Thompson, 1986). In this case, a small “prefatory naturalistic study” provided a descriptive base of information, which was then used for three successively larger “waves” of data collection, each of which included both quantitative and qualitative measures.

Gray and Costello (1987) stretch our conceptions about this design by advocating mixing methods, as well as paradigms for development purposes. Their main argument is that the use of naturalistic qualitative methods to assess *context* first does not preclude the use of positivist quantitative methods for other purposes later in the study. Gray and Costello’s work also supports our call for a more thorough understanding of the contexts appropriate for various mixed-method purposes and of the influence that contextual factors may have on mixed-method designs.

Initiation (I) design. In a mixed-method study with an initiation intent, the major aim of combining qualitative and quantitative methods is to uncover paradox and contradiction. Jick (1983) discussed similar purposes in outlining his “holistic triangulation” design. Rossman and Wilson (1985) demonstrated that iterative use of both method types can intentionally seek areas of

nonconvergence in order to “initiate interpretations and conclusions, suggest areas for further analysis, or recast the entire research question” (p. 633).

Nonetheless, purposeful initiation may well be rare in practice. One excellent example of a more emergent initiation design from our empirical review is Louis’s (1981) evaluation of the Research and Development Utilization program (RDU). This eight-million-dollar demonstration project was funded by NIE (National Institute for Education) between 1976 and 1979 to promote the adoption of new curricula and staff development materials in 300 local schools. Louis discusses key features and examples of the “cyclical interaction” model developed during the course of this evaluation, including the following:

1. Purposive sampling of particular cases was combined with random sampling for survey or other structured data collection in order to maximize both discovery and generalizability.

2. An iterative approach to instrumentation for both field data collection and more standardized instruments was achieved through ongoing interaction between qualitative and quantitative analyses.

3. Analysis began with the first data collection and occurred at periodic intervals throughout the project. The same staff engaged in simultaneous analysis of both qualitative and quantitative data. Testing and verification of both types of data sources increased reliability and validity.

A second example of a more emergent initiation design is Maxwell et al.’s (1986) evaluation of the use of “medical care evaluation committees” in physician education. In this unusual study, ethnographic methods were employed within an experimental framework. Initiation features were evident in the authors’ comments regarding the advantages of the ethnographic approach: “It allowed us to discover aspects of the committees’ educational functioning that we had not anticipated and would have missed had we relied on quantitative methods” (p. 138). Specifically, the qualitative data prompted a recasting of how medical care evaluation

committees influenced physicians' performance. In the original hypothesis, committee participation was expected to directly increase a physician's knowledge and thereby enhance his/her performance. The data indicated, however, that committee participation served to increase the physician's confidence to apply knowledge he/she already had, and this enhanced confidence underlay performance changes.

Drawing in part on these empirical examples, our recommended design for a mixed-method evaluation with an initiation intent incorporates two distinctive features. First, the phenomena investigated with initiation-oriented mixed methods could cover a broad range. Second, to maximize the possibility of unlikely findings, mixing paradigms in this design is acceptable and even encouraged. This advocacy of mixed epistemological frameworks is congruent with Cook's (1985) call for multiple theoretical and value frameworks in applied social inquiry.

Expansion (E) design. In our empirical review, the most frequently cited mixed-method purpose was expansion. This suggests that many evaluators are mixing methods primarily to extend the scope, breadth, and range of inquiry by using different methods for different inquiry components. Typically, in the empirical studies reviewed, quantitative methods were used to assess program outcomes, and qualitative measures to assess implementation. Figure 2 recommends only two elements for a mixed-method expansion design. The empirical work would be encompassed within a single study, and, unique to expansion designs, the phenomena investigated would be distinct. Our sample of mixed-method expansion designs is fairly congruent with these recommendations (see Figure 1).

The decision to "expand" an evaluation to include both process and product components is undoubtedly motivated by the desire to produce a more comprehensive evaluation. However, in many of the evaluations of this genre that we reviewed, there was a paramedic quality to the qualitative component. That is, qualitative data often

appeared in the emergency room of report writing as a life-saving device to resuscitate what was either a failed program or a failed evaluation. Problematic programs or evaluations with insufficient (quantitative) controls or statistical power were discussed in terms of (qualitative) participant experiences, implementation impediments, and recommendations for program improvement.

What is at issue here is how qualitative and quantitative methods in an expansion design can be mixed meaningfully and effectively. Even in the stronger expansion studies reviewed, the qualitative and quantitative methods were kept separate throughout most phases of the inquiry. The term *parallel design* (Louis, 1981) may appear more appropriate. Yet, we prefer the term *expansion* because we believe it more accurately reflects the "multitask" intent of such studies in Cook's multiplism framework.⁴ We also believe that mixed-method expansion studies have not yet tested the limits of their potential. For example, in a higher order expansion design, a more integrated use of methods could be achieved by employing combinations of qualitative and quantitative methods to assess both implementation and outcomes. Such a study may well incorporate elements of triangulation and complementarity into its design, becoming, in effect, a multipurpose study. Or a higher order expansion design could use a mix of different methods, each creatively designed to assess conceptual strands that span or link program implementation and outcomes. The major benefit of such higher order designs would be strengthened inferences. In contrast, our review suggested that the current normative expansion design keeps the different methods separated and thus does not realize such benefits.

In summary, Figure 3 presents a funnel array of recommended design options for the various mixed-method purposes. This array indicates that design options are relatively constrained and narrow for some mixed-method purposes but more flexible and wider for others. The order from most to least constrained design options for

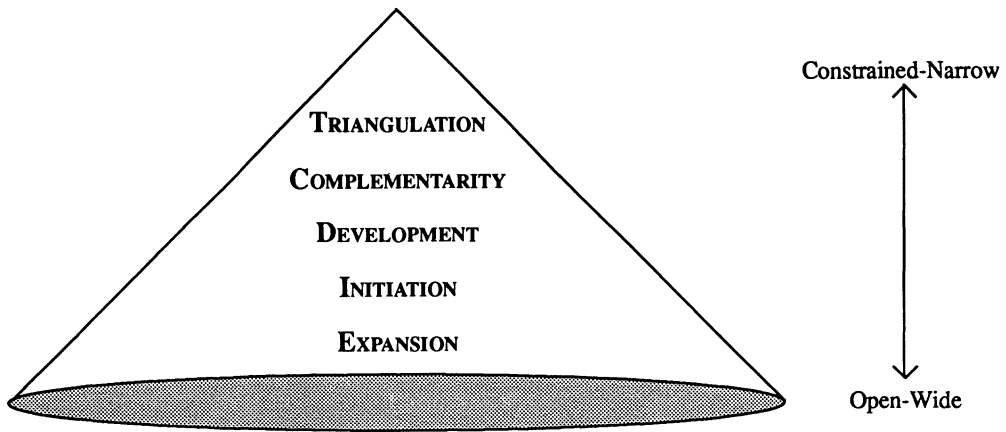


FIGURE 3. Flexibility of design options for mixed-method purposes

mixed-method purpose is as follows: triangulation, complementarity, development, initiation, and expansion.

Results for Mixed-Method Data Analyses

Our empirical review also assessed the nature and degree of qualitative and quantitative integration attained by the studies reviewed during their data analysis and interpretation/reporting stages. The results were grouped in four categories: (a) no integration—both analyses and interpretation were conducted separately; (b) analyses were conducted separately, but some integration occurred during interpretation; (c) integration occurred during both analyses and interpretation; and (d) analyses not reported. A crosstabulation of these analysis results with mixed-method purposes is shown in Table 4. These results reveal that although nearly equal numbers of the studies reviewed attained some degree of qualitative and quantitative integration as did not (23 and 25, respectively), only 5 studies achieved such integration during the analysis process itself. The results further suggest that relatively low levels of integration may characterize studies with an expansion intent, and perhaps relatively high levels may accompany studies with an initiation intent.

Toward Further Development of Mixed-Method Theory and Practice

In this analysis of selected theoretical and empirical literature, we have begun to chart the territory of mixed-method evaluation

designs. Our focus has been on clearly differentiating alternative purposes for mixing qualitative and quantitative methods. Design characteristics relevant to mixed-method strategies and appropriate for specific purposes were also explored. In addition to further refinement of these mixed-method purposes and design elements, we believe several other issues within our overall conceptual framework represent high priorities for future work. These issues include the relationship of mixed-method strategies to evaluation purpose, continuing paradigm questions, procedures for mixed data analysis, utilization, and relevant contextual factors.

With respect to the first issue, we surmised that important distinctions in mixed-method purposes and designs might arise with different evaluation intents—that is, formative versus summative or process versus product. An analysis of our own sample of empirical studies (which included 11 process studies, 12 product studies, and 30 evaluations with both process and product components) yielded no marked differences in mixed-method purpose. Nonetheless, the relationship between evaluation purposes and mixed-method strategies is an important area for further research.

To the probable dismay of purists, in this study we sidestepped the knotty paradigmatic issues involved in mixed-method inquiry. Yet, a comprehensive mixed-method framework must eventually address whether it is appropriate to mix paradigms when

TABLE 4
Crosstabulation of mixed-method analyses and purposes

Analysis category	Purpose					Totals	
	Triangulation	Complementarity	Development	Initiation	Expansion	No.	%
No integration		6	2		17	25	44
Integration during interpretation	3	8	2		5	18	32
Integration during analysis and interpretation		1	1	2	1	5	9
Not reported		3	2		4	9	16

mixing methods. Our own thinking to date suggests that the notion of mixing paradigms is problematic for designs with triangulation or complementarity purposes, acceptable but still problematic for designs with a development or expansion intent, and actively encouraged for designs with an initiation intent.

Future research also need to address the issue of data analysis strategies for mixed-method evaluations. In our empirical review, only five studies integrated qualitative and quantitative data during the analysis process. The creative and promising strategies used by these evaluators are reported more fully in a separate article (Caracelli, Greene, & Graham, 1989). In 18 additional studies that we reviewed, some measure of integration of the different data sets was attained during interpretation and reporting. Typically, in these studies, qualitative data were brought in to support or explain quantitative findings, to flesh out conclusions, or to make recommendations. However, when data mismatches occurred, there was little discussion in any of these studies about these discrepancies, nor were there efforts to resolve them. Both Trend (1979) and Jick (1983) discuss the importance and the challenge of reconciling nonconvergent findings. "When different methods yield dissimilar results, they demand that the researcher reconcile the differences somehow. In fact, divergence can often turn out to be an opportunity for enriching the explanation" (Jick, 1983, p. 143). Shotland and Mark (1987)

also underscore the importance of the "empirical puzzles" (Cook, 1985) that arise when results do not converge, and they call for a more systematic exploration of the possible causes of such inconsistent results.

An additional important area of inquiry concerns utilization specifically as it relates to mixed-method strategies. The fundamental issue here is this: In what common and different ways is quantitative and qualitative information used? And what implications do these utilization processes have for mixed-method approaches to evaluation? Further, attention to contextual factors keeps us mindful of an important question: Is the problem guiding our choice of methods, or vice versa?

These identified areas of future mixed-method inquiry—the role of evaluation purpose, paradigm issues, data analysis strategies, and utilization—as well as others of particular interest to other inquirers, are fundamental to the inherent aim of the research presented herein. Careful planning and defensible rationales must accompany the design and implementation of mixed-methods evaluations. This goal can be achieved only with a more comprehensive theory to guide use of mixed methods in evaluation practice.

Notes

¹ In order to be able to describe current mixed-method practice, we also extracted from each selected study a description of the evaluand and of the evaluation approach, purpose (e.g., formative or summative), time frame, and qualitative

and quantitative methods used. This descriptive information and reference list for our sample of mixed-method evaluation practice is available from the authors upon request, as are the complete details of our sampling and review procedures.

² Seven of the empirical studies reviewed cited as a secondary rationale the inclusion of either qualitative or, more commonly, quantitative methods, not for methodological or theoretical reasons, but rather in anticipation of study audiences' known preferences or needs for this form of information. This political *responsiveness* intent for mixing methods can be viewed, in part, as a tactical maneuver to increase the utilization of evaluation results. However, in contrast to the other mixed-method purposes, a responsiveness intent is unlikely to invoke any significant effort at integration, either at the level of methods, or, more importantly, with respect to the inferences drawn (Mark, 1988). For this reason, we view responsiveness as conceptually different from the other five mixed-method purposes.

³ An eighth design characteristic identified from the theoretical literature was deleted during the pilot testing of the empirical review guide. From Cook (1985) and Shotland and Mark (1987), this characteristic was the following: Are the criteria used to decide which phenomena to assess with multiple methods (i.e., what to make multiple) derived from theory (substantive or methodological) or from the context? Though deleted as a design element, this concern was retained in the descriptive section of the empirical review guide.

⁴ Alternatively, Mark (1988) suggested that expansion be viewed, in conjunction with complementarity and triangulation, as a continuum of mixed-method purposes. This continuum is essentially our phenomena design characteristic, representing the use of different methods to assess different, related, similar, or the same phenomena.

References

- Bower, J. C., Anderson, B. N., & Thompson, B. (1986, April). *Triangulation in the evaluation of school system interventions*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Caracelli, V. J., Greene, J. C., & Graham, W. F. (1989). *Mixed-method data analysis: Strategies and issues*. Manuscript in preparation.
- Cook, T. D. (1985). Postpositivist critical pluralism. In R. L. Shotland & M. M. Mark (Eds.), *Social science and social policy* (pp. 21–62). Beverly Hills, CA: Sage.
- Cook, T. D., & Reichardt, C. S. (Eds.). (1979). *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage.
- Denzin, N. K. (1978). *The research act: An introduction to sociological methods* (chap. 10). New York: McGraw-Hill.
- Gray, P., & Costello, M. (1987, October). *Context as determinant of methodology in multi-client service organizations*. Paper presented at the annual meeting of the American Evaluation Association, Boston.
- Greene, J. C. (1987). *Uses and misuses of mixed-method evaluation designs*. Proposal for the 1988 annual meeting of the American Education Research Association, New Orleans.
- Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. *Evaluation Review*, 9, 523–545.
- Guba, E. G., & Lincoln, Y. S. (1984). *Do inquiry paradigms imply inquiry methodologies?* Copyrighted manuscript.
- Hall, J. N. (1981). *Evaluation and comparison: Social learning curriculum and instrumental enrichment. Final report* (Contract No. G008001869, U.S. Department of Education). Nashville, TN: George Peabody College for Teachers. (ERIC Document Reproduction Service No. ED 244 484)
- Jick, T. D. (1983). Mixing qualitative and quantitative methods: Triangulation in action. In J. VanMaanen (Ed.), *Qualitative methodology* (pp. 135–148). Beverly Hills, CA: Sage.
- Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods: When stories converge. In M. M. Mark & R. L. Shotland (Eds.), *Multiple methods in program evaluation: New Directions for Program Evaluation 35* (pp. 57–75). San Francisco: Jossey-Bass.
- Knapp, M. S. (1979). Ethnographic contributions to evaluation research. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 118–139). Beverly Hills, CA: Sage.
- Louis, K. S. (1981, April). *Policy researcher as sleuth: New approaches to integrating qualitative and quantitative methods*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 207 256)

- Madey, D. L. (1982). Some benefits of integrating qualitative and quantitative methods in program evaluation, with illustrations. *Educational Evaluation and Policy Analysis*, 4, 223–236.
- Mark, M. M. (1988, October). *Discussant remarks on "Toward a theory of mixed-method evaluation designs."* Panel presented at the annual meeting of the American Evaluation Association, New Orleans.
- Mark, M. M., & Shotland, R. L. (1987). Alternative models for the use of multiple methods. In M. M. Mark & R. L. Shotland (Eds.), *Multiple methods in program evaluation: New Directions for Program Evaluation 35* (pp. 95–100). San Francisco: Jossey-Bass.
- Martin, G. O. (1987, October). *The use of qualitative and quantitative evaluation methods for program evaluation. The California International Studies Project: A case study.* Paper presented at the annual meeting of the American Evaluation Association, Boston.
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13–17.
- Maxwell, J. A., Bashook, P. G., & Sandlow, C. J. (1986). Combining ethnographic and experimental methods in educational evaluation. In D. M. Fetterman & M. A. Pitman (Eds.), *Educational evaluation: Ethnography in theory, practice, and politics* (pp. 121–143). Beverly Hills, CA: Sage.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.
- Moody, P. R. (1982). *Pitt Meadows physical education project: Evaluation of an implementation project* (Report No. 82:13). Vancouver: Educational Research Institute of British Columbia. (ERIC Document Reproduction Service No. ED 222 564)
- Moran, T. K. (1987). Research and managerial strategies for integrating evaluation research into agency decision making. *Evaluation Review*, 11, 612–630.
- Peters, M., Marshall, J., & Shaw, R. (1986). The development and trials of a decision making model: An evaluation. *Evaluation Review*, 10, 5–27.
- Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 7–32). Beverly Hills, CA: Sage.
- Reichardt, C. S., & Gollob, H. F. (1987). Taking uncertainty into account when estimating effects. In M. M. Mark & R. L. Shotland (Eds.), *Multiple methods in program evaluation: New Directions for Program Evaluation 35* (pp. 7–22). San Francisco: Jossey-Bass.
- 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, 627–643.
- Shotland, R. L., & Mark, M. M. (1987). Improving inferences from multiple methods. In M. M. Mark & R. L. Shotland (Eds.), *Multiple methods in program evaluation: New Directions for Program Evaluation 35* (pp. 77–94). San Francisco: Jossey-Bass.
- Sieber, S. D. (1973). The integration of field work and survey methods. *American Journal of Sociology*, 78, 1335–1359.
- Smith, A. G., & Louis, K. S. (Eds.). (1982). Multimethod policy research: Issues and applications [Special issue]. *American Behavioral Scientist*, 26.
- Smith, A. G., & Robbins, A. E. (1984). Multimethod policy research: A case study of structure and flexibility. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 115–130). Beverly Hills, CA: Sage.
- Smith, J. K. (1983). Quantitative versus qualitative: An attempt to clarify the issue. *Educational Researcher*, 12, 6–13.
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate. *Educational Researcher*, 15(1), 4–12.
- Smith, M. L. (1986). The whole is greater: Combining qualitative and quantitative approaches in evaluation studies. In D. D. Williams (Ed.), *Naturalistic evaluation: New Directions for Program Evaluation 30* (pp. 37–54). San Francisco: Jossey-Bass.
- Trend, M. J. (1979). On the reconciliation of qualitative and quantitative analyses: A case study. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 68–86). Beverly Hills, CA: Sage.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures* (chap. 1). Chicago: Rand McNally.

Authors

JENNIFER C. GREENE, Associate Professor, Human Service Studies, Cornell University, N136B Martha Van Rensselaer Hall, Ithaca, NY 14853. *Specializations*: program evaluation, applied social research methodology.

VALERIE J. CARACELLI, Social Science Analyst, General Government Division, U.S. General Accounting Office, 441 G St., N.W., Washington, DC 20548. *Specializations*: mixed-methods evaluation research; adolescent and adult development.

WENDY F. GRAHAM, Director of Planning and Information Systems, Cornell University, College of Human Ecology, Martha Van Rensselaer Hall, Ithaca, NY 14853. *Specializations*: program evaluation, educational administration, survey research methods.