

## The Merits of Mixing Methods in Evaluation

JENNIFER C. GREENE

*University of Illinois at Urbana-Champaign, USA*

LEHN BENJAMIN

*Cornell University, USA*

LESLIE GOODYEAR

*Cornell University, USA*

Evaluators are challenged to understand human behavior in all of its natural complexity and individuality. Our work is conducted in natural settings, where history and context matter, where human behavior traces complex patterns of influence and relationship, where what is meaningful to those in the setting is both phenomenological and structural, arising from both lived experiences and the societal institutions that frame and shape those experiences. Engaging this complexity requires *not* a privileging of just one way of knowing and valuing, but rather a marshalling of *all* of our ways of understanding in a framework that honors diversity and respects difference. This framework is advanced in this article as a mixed-method way of thinking. A presentation of key concepts in the framework is followed by case examples from the US.

**KEYWORDS:** better understanding in evaluation; evaluation practice; mixed-methods inquiry; qualitative and quantitative methods

Evaluators strive to contribute to *social betterment* (Henry et al., 1998; Mark et al., 2000). We work in the complex world of real people living in tangled and ever-changing contexts, of real programs designed with hope and promise to fulfill ambitious policies. Our work has the potential to make a difference, to improve life quality or life chances for at least some people in some places, to make a better program, to inform a better policy. The social phenomena that we study 'on the ground' in the real world are unarguably complex, dynamic, and contextually diverse. The need to understand these phenomena, to make sense of contemporary social problems and to find promising solutions to them remains pressing, if not urgent. We therefore need to use *all* of our methodological expertise and skills

### *Evaluation 7(1)*

in this endeavor for contemporary understanding of social issues. We need to marshal *all* of our multiple ways of knowing, and their associated multiple ways of valuing, in the service of credible and useful understanding. We need to adopt a *mixed-method way of thinking* about evaluation, especially social and educational program evaluation. This article offers our view of this mixed-method way of thinking, through a set of conceptual ideas and examples from the field.

Specifically, we will (a) provide a historical context for mixed-method thinking; (b) offer conceptual ideas on mixed-method paradigmatic frameworks, purposes, and design and analysis strategies; (c) provide case examples of these ideas; and (d) conclude with some reflections on both the potential and limitations of mixed-method approaches to evaluation.

### **Historical Notes**

In a world where one way of conducting research was universally considered to be correct, scientific practice would be easy . . . It is the current absence of total certainty about what constitutes correct practice that leads to the advocacy of multiplism in perspectives and methods. (Cook, 1985: 22)

According to Thomas Cook, the impetus for using multiple and mixed methods in evaluation<sup>1</sup> is two-fold: (a) the experience of applied policy researchers during the 1960s and 1970s, and (b) their concomitant appreciation of the fundamental uncertainty of science and the insufficiencies of experimental methods (Cook, 1985: 22). Both call attention to the complexity of social phenomena and the limitations of using just one research approach for capturing this complexity.

Experience during the 1960s and 70s 'War on Poverty' in the US highlighted both the inadequacy of social science knowledge for informing program design and the limitations of scientific methods for assessing program effectiveness (Cook, 1985; Marris and Rein, 1982). Not only did social science theories fail to specify under which conditions a given relationship did and did not hold, but the theories were simply not comprehensive enough to use as blueprints for action (Cook, 1985: 30–2).

Furthermore, experimental scientific methods proved problematic in applied settings, particularly in the area of program evaluation. Notably, defensible experimental research required consistent strategies in a controlled context, but good practice required constantly adjusting strategies to changing circumstances (Marris and Rein, 1982: 206). In programs like Community Action Agencies, this mismatch was clear. Marris and Rein explain,

As soon as the staff, through experimentation or trial and error, discovered a better way of serving trainees they adapted their procedures, methods and techniques accordingly. It was impossible to be inventive, flexible and expedient on the one hand and at the same time do careful, scientific, controlled research on the other. (1982: 198)

As a result, the experimental evaluation models used during the War on Poverty were challenged. These models, which had been adopted from physics and which had dominated social science practice since the 1930s, focused on explanations of observable phenomena through quantitative and objective

measurement (Cook, 1985: 26–9). Yet by mid-century, these models were increasingly under attack from philosophers and scientists alike (Cook, 1985; House, 1980). For example, critics argued that confining social inquiry to observable phenomena constrained, even biased, research on social programs and problems to the finite ability of research methods to capture such phenomena (Cook, 1985: 26). And, challenging the detached objectivity of these research models, Thomas Kuhn argued that objective knowledge is impossible because we cannot divorce our ‘selves’ from our context (Kuhn, cited in Cook, 1985: 26). In sum, Cook explains:

We can see in the attack on positivist methods a rejection of the primacy of observation over introspection, quantification over understanding, micro-level over macro-level analysis, control over naturalism, theory testing over discovery, and crucial experiments conducted on select parts of nature over more tentative probing of all of nature. (p. 29)

In essence, the exclusive use of classic experimental scientific methods by the social sciences missed the fact that good scientific practice is a trial and error process, a process of looking at a problem through multiple lenses, a continual process of action and reflection (Greenwood and Levin, 1998). Science is uncertain. The best response to this uncertainty is not to abandon experimental and quantitative methods altogether, in favor of now-legitimate alternatives from the interpretive and critical sciences, for this would be simply repeating the misguided myopia of social science practice in the 1960s. Rather, we believe that the fundamental uncertainty of scientific knowledge – especially about complex, multiply-determined, dynamic social phenomena – can be better addressed through the multiple perspectives of diverse methods than through the limited lens of just one.

Today’s researchers have a more realistic picture of the complexity of the social problems they study and the limitations of all methods in understanding social interventions. Many program evaluators routinely use a mix of different methods in their work. However, neither program evaluators nor policy researchers more broadly utilize mixed-method designs to their fullest potential. Our objective in this article is to advance more thoughtful and effective mixed-method practice by encouraging attention to several mixed-method conceptual ideas and frameworks.

### **Frameworks for Mixing Methods**

As noted, at the technical level, there has been increasingly widespread agreement that mixing methods is a useful way to collect a variety of data, for example, employing interviews to understand the nuances of respondents’ experiences of a program and questionnaires to obtain quantitative information on household income change. Clearly, ‘different kinds of methods are best suited to learning about different kinds of phenomena’ (Greene and Caracelli, 1997: 7). Supporters of mixed-methods evaluation exist outside as well as inside academia, with recent promotions emerging from the federal government and the national foundation

### *Evaluation 7(1)*

community (Annie E. Casey Foundation, 1995; Frechtling and Sharp for the National Science Foundation, 1997).

In contrast to the support that mixing methods receives on the technical level, there is continued debate at the level of philosophy. Greene and Caracelli describe this as the 'paradigm issue' in mixed-method evaluation (1997). In theory, a method does not have to be linked to a particular paradigm. However, in practical terms, it often is and therefore frames the discussion on whether or not it is sound (or even possible) to mix methods that are linked to different philosophical traditions, such as postpositivism and interpretivism.

The dialogue between and among researchers and practitioners on the paradigm issue in mixed-method evaluation has generated a number of stances that currently serve as distinct conceptual frameworks for mixing methods (Greene and Caracelli, 1997; Guba and Lincoln, 1989; Krantz, 1995; Patton, 1988). The most prominent include the 'pragmatic' and 'dialectical' views, both of which are bounded by existing philosophical traditions. More recent views feature the displacement of historical philosophical differences by a substantive framework, such as a program theory, or by an alternative inquiry paradigm, such as realism. Each of these four frameworks for mixing methods is briefly described next.

#### ***Pragmatic View***

Evaluators and policy researchers who work as pragmatists believe that it is the natural limitations and opportunities of a given context that matter when making mixed-methods decisions, rather than a consideration for the philosophical compatibility of different inquiry traditions. While fundamental differences may exist between different paradigms, they are 'descriptions of, not prescriptions for, research practice' (Greene and Caracelli, 1997: 8). Michael Patton, whose work clearly illustrates the pragmatist stance, acknowledges that paradigm differences are real; however, he argues that paradigm descriptions over-emphasize dualisms such as objectivity–subjectivity. 'Descriptions of alternative paradigms represent ideal types that contrast opposing ends of what are actually methodological continua' (Greene and Caracelli, 1997: 9). The bottom line is that the pragmatist's view helps to overcome 'seemingly logical contradictions' (Patton, 1988: 127) because resolving paradigm differences is not required for using a variety of methods. Rather, in this approach it is responsiveness to context that grounds methodological decision making.

#### ***Dialectical View***

The dialectical stance holds that differences between paradigms exist and are important. Because differences between different philosophical assumptions cannot be reconciled, it is important to value each paradigm. The notion is that a 'conversation' between different paradigms and their related methods within one study leads to better understanding. Greene and Caracelli explain that using different paradigms in an integrative manner leads to 'more comprehensive, insightful and logical results than either paradigm [interpretivist or postpositivist] could obtain alone' (1997: 10). The key dialectic tenet they advance is 'the need

to move beyond debating paradigmatic differences that may well be irreconcilable and to focus instead on joining the critical features of our evaluative claims that represent distinct traditions. Joining such critical features can help to generate more relevant, useful and discerning inferences' (1997: 19).

### ***Substantive Theory View***

In program evaluation circles, substantive theory – specifically, program theory or the intended connections among inputs, activities and outcomes – has recently re-emerged as an important framework for our work (Bickman, 1987, 1990; Chen, 1990; Chen and Rossi, 1983; Weiss, 1998). The argument is that our work should not be driven by our methods, but rather by the substantive issues and programs being investigated (House, 1994). The argument is also that evaluation should serve primarily to contribute to conceptual and practical knowledge regarding how best to address our social problems. Evaluators should concentrate on understanding the meaningfulness and effectiveness of a given program design and implementation in a given context, toward better understanding of that programmatic response to that social problem. Given the complexity of social problems, sound program theories are similarly complex and multi-dimensional. Thus, like the alternative paradigm stance (described next), using program theory as a framework for evaluation not only invites but actively requires a diverse mix of methods. However, this framework is distinct from the others in that the impetus behind using mixed methods is more program- and policy-driven rather than anchored in epistemology or methods.

### ***Alternative Paradigm View***

Finally, 'the paradigm issue' in mixed-method inquiry is addressed by proponents of alternative, 'newer' paradigms, such as some forms of critical social science (Fay, 1987); action science (Greenwood and Levin, 1998); realism (Henry et al., 1998; Pawson and Tilley, 1997); and many genres of postmodern thought. These alternative paradigms dissolve old dualisms principally by embracing both sides (or all sides). Within many of these paradigms, elements of the previous three frameworks surface, including a valuing of the pragmatic character of context, of the importance of engaging different methods in respectful conversation, and of the broader inquiry goal of contributing to conceptual and practical knowledge. Further, from the perspective of these alternative paradigms, mixed-method approaches are advocated because mixing methods constitutes good practice, which in turn generates better understanding that results in informed action. Realism, for example, seeks contextualized causal understandings of social phenomena. Such understandings require lenses of both generality and particularity, both objectivity and subjectivity, both patterned regularities and idiosyncratic stories, both policy and advocacy value stances. Such understandings therefore not only invite but actually require a diverse mix of methods.

A major premise of our mixed-method way of thinking is that good mixed-method practice is achieved by thoughtful mixed-method planning. An important initial step in such planning is thus to figure out one's stance on the 'paradigm



issue' in mixed-method inquiry (or, less grandiosely, for a given mixed-method evaluation) and to adopt an organizing framework that is consonant with that stance. Consciousness of this organizing framework and adherence to its guidance for inquiry practice, we argue, underlies good mixed-method work.

### **Mixing Methods Purposefully**

For all organizing frameworks, a second important step in mixed-method planning is to give considered attention to the reasons for mixing or the purposes to be achieved. As noted, the overall purpose for mixing methods is to afford a greater reduction in uncertainty and to attain a better understanding of the social phenomena being studied. Here, drawing from Cook (1985) and Greene et al. (1989), we provide a more detailed explication of what is meant by 'better understanding.'

1. **Enhanced validity and credibility of inferences.** This is illustrated by the classic mixed-method *triangulation* design, in which different methods – ideally with offsetting biases – are used to measure the same phenomenon with intended convergence of results and thus the ruling out of various threats to validity.
2. **Greater comprehensiveness of findings.** More complete accounts of social phenomena, 'more comprehensive picture[s] of how policies impact on the social world' (Cook, 1985: 47), are attained when different methods are used to offer different lenses on a social program (for example, the perspectives of the administrators and the experiences of the participants); to capture different dimensions of the program (for example, cross-site average effects and contextualized individual stories); and to enable more facets of the program to be studied (for example, implementation and outcomes or both standardized and phenomenological outcomes).
3. **More insightful understandings.** Some mixed-method designs yield findings that do not converge, but rather challenge or even conflict with one another, and thus must be reconciled through further analytic questioning and probing. Cook (1985) suggests that this can lead to greater specificity about the contributing factors that control the relationship under study. It can also lead to a conceptual reframing of this relationship or to the discovery of new concepts that constitute important theoretical or practical insights.
4. **Increased value consciousness and diversity.** Cook argued that 'multiplism promises to make policy research more intellectual, value conscious, and debate-centered' (1985: 47). Repeated exhortations to include diverse value stances and perspectives – via multiple stakeholder views, multiple theoretical stances, multiple analytic strategies, and so forth. – permeate Cook's presentation of multiplism as a framework for policy research. Moreover, because different methods themselves advance different values and interests (Greene, 1997), mixed-method evaluation invites value pluralism and thus critical commentary, public debate, and deliberative dialogue (House and Howe, 1999, 2000) about diverse perspectives on important public issues.

We acknowledge that these various purposes for mixing methods overlap, and that multiple purposes are often attained, or at least intended, within one evaluation design. Nonetheless, we contend that because different mixes of methods are required for different purposes (see Greene et al., 1989), attending to purpose is a critical step in thoughtful mixed-method practice.

### **Mixing Methods: Design and Analysis Strategies**

After framework and purpose comes mixed-method design. In mixed-method evaluation, the different methods can be designed and implemented relatively separately from one another in a *coordinated* design, or more intentionally interactively in an *integrated* design (Caracelli and Greene, 1997).

In a coordinated design, the primary 'mixing of methods' happens at the end of the study, at the stage of drawing inferences and conclusions. With this design, one set of findings characteristically illustrates, enhances, helps to explain, or refines the other set of findings (as illustrated by Mark et al., 1997). In an integrated design, there are planned, scheduled points of interaction between the different evaluation methods throughout the duration of the study, and so the 'mixing of methods' is iterative and ongoing. Such points of interaction might include instrument design, sample selection, initial field contact, key data gathering activities, and several points during data analysis. With an integrated design, the different methods are blended, united, dialectically combined into a coherent whole, so that evaluation results are more a synthesis of all study data and less a report of findings from each method separately. Integrated designs are generative, yielding new insights, fresh perspectives, or redirected questions.

Relatedly, analysis strategies for mixed-method inquiry can follow (from Marquart et al., 2000) 'parallel tracks', where each data set is analyzed separately and comparisons and connections made at the stage of drawing conclusions and inferences. Or, analysis strategies can be more interactive, where interim results from one method are clustered, summarized, or transformed and then crossed over to the other analytic track for further processing. This latter strategy is dubbed 'crossover tracks' analysis.

Logically, coordinated mixed-method designs and parallel track analyses fit well with a pragmatic framework for mixing methods. And integrative designs and crossover tracks analyses fit well with a dialectic mixed-method framework. In actual mixed-method practice, most inquiry designs and analyses incorporate both coordinated-parallel and integrated-crossover strands, or at times the different methods are quite separate and at other times more interactive. That is, mixed-method practice is much more complex and dynamic than these theoretical constructs can capture. These conceptual handles – of coordination and integration – and all other mixed-method constructs presented herein, are thus intended less as prescriptions of practice and more to advance the mixing of methods as a way of thinking about evaluation practice.

## **A Mixed-Method Way of Thinking: Reprise**

In sum, we believe today's demands on evaluation can be well addressed, if not best addressed, by inquiries that include diverse ways of knowing and valuing. Using multiple, diverse methods in our work, however, requires the same thoughtfulness and rigor in planning that we expect of work conducted within a single methodological tradition. Among useful planning constructs are the various mixed-method organizing frameworks, purposes, designs and analysis strategies presented above. (See also work by Brewer and Hunter, 1989; Creswell, 1994; Ragin, 1989; Tashakkori and Teddlie, 1998; among others.)

Beyond thoughtful planning, good mixed-method inquiry requires an openness to other views and perspectives, not just to rival explanatory hypotheses but more profoundly to rival ways of thinking and valuing. Good mixed-method evaluation actively invites diverse ways of thinking and valuing to work in concert toward better understanding. In good mixed-method evaluation, difference is constitutive and generative.

In the remainder of this article, we present and analyze three case examples of mixing methods. We connect these examples to our mixed-method constructs, and we use them to reflect on the promise and limitations of mixing methods in evaluation and policy research.

## **Case Examples of Mixed-Method Social Inquiry**

We selected three case examples to represent three of the four different frameworks for mixing methods presented earlier: pragmatic, dialectic, and substantive theory. We excluded the fourth alternative paradigm framework primarily because the diversity of these alternative paradigms themselves makes any one example insufficiently illustrative. We selected case examples that we perceive to be good instances of mixed-method practice, because in each the mixing of methods contributed significantly to 'better understanding'. For each example, we briefly describe the inquiry context, delineate the mix of methods used, and assess how 'better understanding' was achieved, with reference to the mixed-method purposes and design and analysis strategies offered above.

### ***Being Pragmatic in Mixing Methods***

Phase I of an evaluation of New Hope, an anti-poverty program in Milwaukee, Wisconsin, provides an excellent example of pragmatic mixed-method inquiry.

**The Context** New Hope was a voluntary anti-poverty initiative offered to residents in selected inner-city neighborhoods in Milwaukee. In exchange for working 30 hours a week, New Hope participants received, as needed, a wage subsidy, health insurance, and a child care allotment. During the two years of this initiative, Wisconsin was experiencing a booming economy, which affected many aspects of New Hope's implementation and evaluation.

**The Mixing of Methods** The evaluation of New Hope began as a randomized experiment (Bos et al., 1999), focused on causal explanations of intended program



outcomes, especially poverty reduction, employment, and family and child well being. At baseline and then again at the end of the 2-year program, administrative records and extensive family and teacher surveys were used to assess treatment and control group status on these intended outcomes. At the two-year point, the ethnographic component of the evaluation design was initiated (Weisner, 2000). This component was intended primarily to contribute phenomenological understanding regarding the meaningfulness of the program experience over the first two years, and in anticipation of a planned 5-year follow-up study (Phase II of the evaluation). A random sample of 45 families, half from the treatment group and half from the control group, were interviewed ethnographically to ascertain the meaningfulness of the New Hope intervention, or comparable interventions for the control group, in terms of families' activities of daily living.

For this first phase of the evaluation then, the two data sets were analyzed mostly separately, and comparisons made at the level of findings.

### ***Better Understanding as More Comprehensive, Enriched, and Contingent Findings***

In this evaluation, there were several important instances of findings from the two data sets together offering a more comprehensive and enriched understanding of study phenomena than either data set alone. One example relates to intended effects of the intervention on work and income. The qualitative interviews suggested important heterogeneity among program families with respect to their ability to benefit financially from the New Hope intervention. Some families had so many challenges – like drug dependence, children with severe behavioral problems, or family members with serious chronic illness – that New Hope's particular package of benefits was unlikely to make difference in their ability to secure a job. Other families had no apparent challenges of this kind and were already sustaining employment on their own. Yet, a middle group of families, with only one or two such challenges, appeared poised to benefit from the New Hope package. In the quantitative data, an index of 'potential employment barriers' was constructed using past employment, education, and arrest records, and the data were re-analyzed by whether families faced zero, one, or two or more such barriers. The results dramatically confirmed the qualitative trends. There were no program effects on earnings for families with zero or multiple barriers to employment, but significant positive effects (averaging about \$1250 per year) for families with just one such barrier.

### ***Mixing Methods Dialectically***

A 5-year national study of preschool inclusion, conducted by the Early Childhood Research Institute on Inclusion which involved researchers from five universities,<sup>3</sup> well illustrates the dialectic potential of mixed-method inquiry (Li and Zercher, 1997; Marquart, 1997; Marquart et al., 1997, 2000).

**The Inquiry** The Early Childhood Research Institute was funded to inform policy and program initiatives regarding the inclusion of preschool-aged children with disabilities in educational and child-care settings with typically developing

### *Evaluation 7(1)*

peers. The Institute's research was designed to examine barriers to inclusion and to identify promising strategies for overcoming those barriers and thus promoting inclusive preschool programs. Under the assumption that inclusion can only be understood by examining the many contexts in which children and their families participate, these researchers adopted the ecological systems model of Urie Bronfenbrenner (1979) as the substantive framework for their work.<sup>4</sup>

This study was conducted on a sample of 16 inclusive preschool programs, four each in four different sites around the country. The programs were identified as 'inclusive' by key informants in each site. They were also purposively selected to demonstrate diversity on program model for inclusion and child/family demographics. Within each program, seven children were purposively selected for study – five children were disabled and had an Individualized Educational Plan and two were typically-developing peers in the same classroom – for a total of 112 children. In addition, these children's families and teachers, and the administrators and policy makers for each program were studied.

**The Mixing of Methods** The inquiry included a variety of qualitative and quantitative assessments, selected intentionally to gather both emic, contextual meanings of key study constructs 'in that the perspectives of the individuals involved in the inclusion process were examined . . . first-hand' (Marquart et al., 1997: 3); and etic, standardized definitions and data. A summary of the family and classroom measures used in this research is provided in Table 1.

In this extensive study of preschool inclusion, the researchers self-consciously mixed methods to attain *all* of the mixed-method purposes identified above. Their mixed-method design and analyses were intentionally interactive, including, for example, planned iterations of 'qualitative-then quantitative-then qualitative' data gathering and inventive transformations of data from one type to the other, enabling highly integrative analyses. This research thus clearly illustrates a dialectic framework, in that not only were different methods mixed but different paradigmatic assumptions and characteristics were also embraced, explicitly as means to better understanding.

**Better Understanding as More Insightful Findings** One example from this study will be shared (Marquart et al., 2000). This example involves the development of case summaries for each of the 16 programs, which were 'to provide an in-depth analysis of [the meanings of] inclusion in the programs, to determine barriers and facilitators of inclusion, and to describe idiosyncratic issues as well' (p. 13). In developing these case summaries, the researchers used a highly interactive, 'cross-over analytic approach [which] is characterised by a concurrent analysis of both qualitative and quantitative data. Data analysis moves back and forth constantly between both data sets throughout the stages of data [description and reduction,] transformation, comparison, and integration' (p. 13).

Specifically, relevant data of both types were first analyzed descriptively and reduced in standard ways – quantitative data to graphs and tables, and qualitative data to descriptive themes, narrative sums, and vignettes. Second, selected sets of these reduced data were transformed – the most salient points in quantitative

Table 1. Family and Classroom Measures in the Preschool Inclusion Research

Domain	Qualitative measures	Quantitative measures
Classroom	Participant observation	CASPER (Code for Active Student Participation and Engagement-Revised)
	Post-CASPER notes	
	Teacher survey about child's friendships	Peer Rating Sociometric Assessment
	Teacher interviews	Battelle Developmental Inventory
	Administrator and policy maker interviews	Curriculum Activities Survey
	Analysis of program documents	
Family	Individual interviews	Telephone survey
	Family survey about child's friendships	

Source: Marquart et al., 1997: 4

tables and graphs were summed in narrative form, and ordered matrices were developed to capture the main points in narrative sums. Third – and this is the heart of this analytic approach – the two sets of tables-graphs-matrices (the original quantitative set and the transformed qualitative set) were carefully compared for instances of convergence, complementarity, and discordance. Parallel comparisons were conducted on the two sets of narrative sums (the original qualitative set and the transformed quantitative set). Congruence in these comparisons led to stronger, more valid inferences, incongruence to further probing and generative insights.

An example of one generative insight for one case summary follows.

1. The **descriptive quantitative graphs** of observation data in this site revealed that one child with disabilities 'spent a longer time . . . by herself outside a group setting than her peers, and the least time in circle time, which was the main instructional time each day' (p. 16).
2. A **comparison** of this graph with the (transformed and ordered) matrices of qualitative observation notes indicated that **all** of this child's negative behaviors 'had something to do with her difficulties in staying with a group, [suggesting] that her problem was not confined to circle time . . . but might occur in any of the teacher-directed group activities' (p. 16).
3. **Further probing** of field notes and teacher interview data revealed a discrepancy in inclusion philosophy and child expectations between the regular teacher and the special education teacher, a discrepancy that was complex, nuanced and multi-faceted. Notably, the special education teacher gave this child tacit permission to leave the group whenever she

### *Evaluation 7(1)*

felt like it, believing that 'all children should try their best, but if they still could not follow the rules [for example, sitting still in group time], they should be given time to learn' (p. 16). The general education teacher, in contrast, said that high and consistent expectations are needed for all children, so that they can get what they need from the educational program. When a child is exempted from the rules, like participating in group time, 'she is not getting what she should because she is not staying with the group. She doesn't sit down long enough to pick up what we are doing . . . I feel she is not getting what she really needs as far as an education is concerned' (p. 16).

This discrepancy of views was interpreted as a major barrier to inclusion in this particular program. The generation of this discrepancy illustrates what we believe is the dialectic potential of mixed-method inquiry. By intentionally honoring multiple ways of knowing – in this example, both contextualized and standardized, both particular and general – and by actively seeking them out via the implementation of diverse methods and the planned interactions and comparisons of different data, a more insightful, better understanding of the program was attained. This interactive, 'conversational' design importantly characterizes a dialectic mixed-method framework.

### ***Mixing Methods in Service to a Substantive Program Theory***

A dissertation study, which investigated how well theorized benefits of performance assessment were actually manifested in practice (Sivalingham, 1997), illustrates our third mixed-method framework – one guided by substantive theory.

**The Inquiry** The context for this study was an evaluation of a high school science reform, a reform that included a shift from multiple-choice assessments to performance-based assessments. This particular component of the evaluation empirically examined claims from the literature that performance assessments:

1. Enhance positive student attitudes (perceived ability in science, interest and motivation for science, meaningfulness of science);
2. Influence instruction toward less emphasis on facts and more on concepts and problem solving;
3. Can measure higher order thinking skills; and
4. Yield more equitable performance across student groupings of gender, race, and class.

This study thus enabled an evaluation of the 'consequential validity' of these performance assessments. For present purposes, Sivalingham's investigation of the second theoretical claim above – that performance assessments influence instruction – will be discussed.

Briefly, much assessment theory discusses the ways in which assessment can influence, even 'drive', teachers' instructional decisions and practices (Popham, 1999). Because judgments about teacher competence (by authorities and the

general public) are linked to student test performance, teachers experience pressures, both explicit and implicit, to generate strong test performance among their students. And teachers commonly respond to these pressures by teaching the kinds of knowledge and skills that are tested. In this context, performance assessment theory contends that use of performance assessments can promote instructional attention to, and thus student mastery of, complex cognitive tasks, specifically, understanding scientific concepts and principles, problem solving skills, and higher-order thinking skills. (For further elaborations of this aspect of performance assessment theory, see the literature review in Sivalingham, 1997.) This theoretical claim then, rather than the inquiry context or paradigmatic or methodological considerations, was the driving framework for the design and selection of methods in this evaluative study.

**The Mixing of Methods** For her study, Sivalingham chose a sample of courses from the high school science curriculum that clearly contrasted multiple-choice versus performance-based assessments on the end-of-year examples. She selected for her study four 'new' courses with final exams that were 60 to 100 percent performance-based, and which encompassed 13 classes of 250 students. And she selected two 'traditional' courses with final exams that were 85 to 100 percent multiple-choice, and which included 7 classes of 116 students. All science teachers in the school were represented in this sample.

From this sample, Sivalingham gathered data from both teachers and students on the kinds of activities featured in science instruction. In individual interviews, teachers were asked about their science class activities, particularly the activities they used to promote understanding concepts, problem solving and higher-order thinking (the three cognitive tasks identified in the literature as particular benefits of performance assessment). A questionnaire also asked students to identify the kinds of activities they did in science class (for example, build a model, do a lab) and to note their frequency.

**Better Understanding as More Integrated Findings** Sivalingham undertook an interactive analysis of these interview and questionnaire data. Like Marquart et al., she sought to integrate the two streams of data into one overall portrayal. This analysis proceeded in three main stages.

1. First, Sivalingham inductively analyzed the teacher interview data to better understand (a) how these science teachers conceptualize understanding concepts, problem solving, and higher-order thinking, and (b) what activities they use to promote student mastery of these cognitive tasks in their classrooms.
2. Second, Sivalingham conducted descriptive and factor analyses on the student questionnaire data, looking for clusters of classroom activities as reported by students. *Interpretation of these factor results was importantly aided by the teacher data.* The results, displayed in Table 2, indicated four factors:



*Evaluation 7(1)*

- I. Participative-collaborative activities that require active student participation and that teachers use to promote problem solving and higher-order thinking.
  - II. Activities that advance students' individual construction of knowledge and that teachers use to promote understanding of scientific concepts.
  - III. Activities that require students to take responsibility for their own learning.
  - IV. Teacher-directed activities.
3. Third, Sivalingham compared the different courses in her sample on these four factors of instructional activities. Collapsing across classes taught by the same teacher, Sivalingham computed mean factor scores for each course and then displayed the results graphically (see Figure 1).

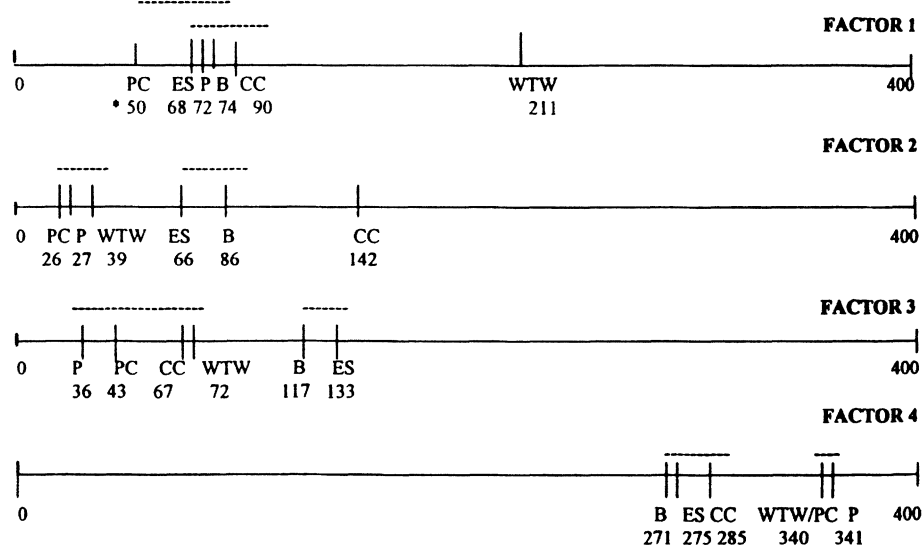
The results for Factors 1 and 2 indicate that students in the most performance-based courses – the Way Things Work (WTW) and Chemistry in the Community (CC) – reported significantly more frequent use of classroom activities that teachers use to promote complex cognitive outcomes than students in the most multiple-choice courses – Physics (P) and Physical Chemistry (PC). Yet, the results for Factor 4 also indicate that students in all courses reported substantially more teacher-directed activities than any other kind. In short, these findings support the theoretical claim that performance assessment can influence instruction toward the inclusion of activities designed to promote higher-order cognitive skills, but only modestly, at least in this context. Instructional activities in all science classes in this school remained largely teacher-directed.

This integrated portrayal – derived from a mix of methods and an inventive

*Table 2.* Factors Representing Clusters of Activities in the Student Data

<i>Factor 1: Participative activities, promoting problem solving and higher-order thinking</i>	<i>Factor 2: Individual activities, promoting knowledge and concept understanding</i>	<i>Factor 3: Activities requiring student responsibility</i>	<i>Factor 4: Teacher-directed activities</i>
Open-ended lab (group)	Concept map	Open-ended lab (individual)	Structured lab (group)
	Library research		Write lab report
Defend design of lab expt/model (oral/written)	Write own notes from readings	Structured lab (individual)	Copy notes given by the teacher
Model	Poem	Write own notes from readings	Teacher demonstration
Oral presentation	Research paper/term paper	Field visits	Worksheet
Project		Journal	
		Portfolio	

Source: Sivalingham, 1997: 144



\*Mean scores of courses with decimal omitted; broken lines denote no significant difference at 0.05 level.

Figure 1. Comparison of Science Courses on the Four Factors  
 ES - Earth Science; B - Biology; CC - CHEMCOM; P - Physics; PC - Physical Chemistry; WTW - Way Things Work  
 Source: Sivalingham, 1997: p. 163.

analysis – offers a parsimonious and theoretically meaningful lens on the evaluative questions about performance assessment in practice.

### Mixing Methods: Promise and Limitations in Practice

In conclusion, we acknowledge that while mixed-method approaches have great promise for enhanced understanding of social programs and policies, they are not a panacea. In addition to the practical limitations of time and resources that constrain any evaluation study, we have identified four issues that must be afforded particular attention when planning mixed-method evaluation. We present these issues not as the ‘downsides’ of mixed-method inquiry, but rather to reinforce the essential importance of purposeful, thoughtful planning that must accompany all such inquiry.

#### The Balance of Certainty and Complexity

As we have discussed, mixed-method approaches can enhance validity and offer greater certainty when results from two approaches reveal similar conclusions. Yet, because methods ‘take a different picture’ of the same social phenomenon, results may converge but offer a substantially more complex and contextual portrayal of this phenomenon. Mixed-method designs may thus require the evaluator to think carefully about representing heterogeneous findings to users of

evaluation results, and to proactively engage in follow-through, helping program staff and policy makers to understand and use the complex pictures that result from such designs. In turn, such follow-through could importantly provide practitioners and policy makers with opportunities to think about social problems differently and with spaces to consider ideas for innovative practice.

### ***Selecting What to Make Heterogeneous in Methods Choices***

Methods vary on multiple dimensions. Which dimensions are most important to make heterogeneous in a given mixed-method study? This question warrants considerably more research. Useful starting points are offered by Cook (1985), Shotland and Mark (1987), and Greene et al. (1989). Cook, and Shotland and Mark emphasize the importance of heterogeneity or balance in methodological bias, so that multiple methods yield a less biased estimate of program effects than one method alone. This reasoning is clearly most relevant to triangulation mixed-method designs. For other mixed-method purposes, Greene et al. offer some preliminary ideas about designed method variation. Further work on this issue should definitely include well defended instances of good mixed-method practice.

### ***Are All Methods Created Equal?***

Numerous authors have attended to the weight accorded to the different methods, different paradigms, or both in mixed-method inquiry (Creswell, 1994; Greene et al., 1989; Smith, 1997; Tashakkori and Teddlie, 1998). Designs in which one method or paradigm is dominant (Creswell, 1994) are recognized as different from designs in which the various methods or paradigms have relatively equal status and voice. Like the question of what to make heterogeneous in method choices, this question of balance could also benefit from additional research. Meanwhile, we believe that evaluators should decide what is the most important question or problem being addressed by the study and prioritize the method/paradigm appropriately. Even so, a key benefit of using a mixed-method approach, and especially one that includes mixing paradigms, is that while one world view might be privileged, it is constantly challenged by other approaches in the same study.

### ***The Increased Importance of Data Quality***

Although mixing methods seems a creative and useful way to include different points of view on the same issue, the demands of a mixed-method design are different than the demands of a single-method study. Because methods and underlying paradigms are being compared and contrasted through the mixing of methods in one study, there needs to be a heightened awareness on the part of the evaluator of the assumptions that undergird the methods as well as the epistemological assumptions that shape knowing within a paradigm. In a single-method study, these assumptions may not be brought to the fore, because they are not challenged by the collection of data using other methods or paradigmatic assumptions. In a mixed-method study, these assumptions need to be foregrounded and wrestled with in a thoughtful and ongoing manner.

## Conclusion

In sum, we believe that mixed-method evaluation offers much promise for increased understanding of social programs, which in turn helps to improve practice. But designing and implementing a mixed-method evaluation is not merely choosing from a smorgasbord of methods available. Whether an evaluator is mixing methods practically, dialectically, within the boundaries of program theory or within alternative paradigms, this approach demands a heightened reflexivity and responsiveness that may not be necessary in single-method studies. This reflexivity fundamentally rests on openness to diversity, acceptance of difference, tolerance of uncertainty – a mixed-method way of thinking. The possible gains achieved by mixing methods in evaluation are great: increased validity, more comprehensiveness of findings, more insightful understandings and greater value consciousness and diversity. To fully realize this potential, we have offered a set of mixed-method constructs, several examples, and a strong exhortation for mixed-method evaluators to be purposeful, thoughtful and reflexive in their practice.

## Notes

An earlier version of this article, with additional co-author Staci Lowe of Cornell University, was presented in November 1999 at the annual APPAM meeting in Washington DC.

1. Multiple methods refers to using more than one method, and mixed methods refers to using two or more *different* kinds of methods in the same study. While Tom Cook's discussion is on multiple methods and multiplism, he emphasizes the importance and value of difference and diversity in methodological perspective throughout.
2. Having experienced the benefits of diverse data sources and information, the New Hope inquiry team (Tom Weisner and others, personal communication, 2000) is thinking more interactively and dialectically for the 5-year follow-up phase of this evaluation. Already, findings from one study component have informed the design and sampling for the methods being used in the other.
3. The universities were San Francisco State University, Vanderbilt University, University of Maryland, University of North Carolina, and University of Washington.
4. Based on this framework, this 5-year study investigated preschoolers within their classrooms, homes, and community programs (microsystem), the relationships between families and program professionals and among professional across settings (mesosystem), the program administrative structures and policies (exosystem), and the larger socio-political/cultural contexts for inclusion (macrosystem). This national research program thus could also illustrate the third mixed-method framework identified previously, in which a strong conceptual framework invites, even insists upon multiple methods.

## References

- Annie E. Casey Foundation (1995) *Getting Smart, Getting Real: Using Research and Evaluation Information to Improve Programs and Policies*. Report of the Annie E. Casey Foundation's September 1995 Research and Evaluation Conference. Baltimore, MD.

*Evaluation* 7(1)

- Bickman, Leonard (ed.) (1987) *Using Program Theory in Evaluation. New Directions for Evaluation No. 33*. San Francisco, CA: Jossey-Bass.
- Bickman, Leonard (ed.) (1990) *Advances in Program Theory. New Directions for Evaluation No. 47*. San Francisco, CA: Jossey-Bass.
- Bos, J., A. C. Huston, R. Granger, G. Duncan, T. Brock and W. C. McLoyd (1999) *New Hope for People with Low Incomes: Two-Year Results of a Program to Reduce Poverty and Reform Welfare*. New York, NY: Manpower Research Demonstration Corporation.
- Brewer, John and Albert Hunter (1989) *Multimethod Research: A Synthesis of Styles*. Thousand Oaks, CA: Sage.
- Bronfenbrenner, U. (1979) *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.
- Caracelli, Valerie J. and Jennifer C. Greene (1997) 'Crafting Mixed-Method Evaluation Designs', in J. C. Greene and V. J. Caracelli (eds) *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms, New Directions for Evaluation No. 74*: 19–32. San Francisco, CA: Jossey-Bass.
- Chen, Huey-Tsyh (1990) *Theory-Driven Evaluation*. Thousand Oaks, CA: Sage.
- Chen, Huey-Tsyh and Peter H. Rossi (1983) 'Evaluating with Sense: The Theory-Driven Approach', *Evaluation Review* 7: 283–302.
- Cook, Thomas D. (1985) 'Postpositivist Critical Multiplism', in R.L. Shotland and M. M. Mark (eds) *Social Science and Social Policy*, pp. 21–62. Thousand Oaks, CA: Sage.
- Creswell, John W. (1994) *Research Designs: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.
- Fay, Brian (1987) *Critical Social Science*. Ithaca, NY: Cornell University Press.
- Frechtling, Joy and Laurie Sharp (eds) (1997) *User-Friendly Handbook for Mixed Method Evaluations*. Arlington, VA: National Science Foundation, Directorate for Education and Human Resources.
- Greene, Jennifer C. (1997) 'Evaluation as Advocacy', *Evaluation Practice* 18: 25–35.
- Greene, Jennifer C. and Valerie J. Caracelli (1997) 'Defining and Describing the Paradigm Issue in Mixed-Method Evaluation', in J. C. Greene and V. J. Caracelli (eds) *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms, New Directions for Evaluation No. 74*: 5–17. San Francisco, CA: Jossey-Bass.
- Greene, Jennifer C., Valerie J. Caracelli and Wendy F. Graham (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis* 11: 255–74.
- Greenwood, Davydd and Morten Levin (1998) *Introduction to Action Research: Social Research for Social Change*. Thousand Oaks, CA: Sage.
- Guba, Egon G. and Yvonna S. Lincoln (1989) *Fourth Generation Evaluation*. Thousand Oaks, CA: Sage.
- Henry, Gary T., George Julnes and Melvin M. Mark (eds) (1998) *Realist Evaluation: An Emerging Theory in Support of Practice. New Directions for Evaluation No. 78*. San Francisco, CA: Jossey-Bass.
- House, Ernest R. (1980) *Evaluating with Validity*. Beverly Hills, CA: Sage.
- House, Ernest R. (1994) 'Integrating the Quantitative and the Qualitative', in C. S. Reichardt and S. F. Rallis (eds) *The Qualitative-Quantitative Debate: New Perspectives. New Directions for Evaluation No. 61*. San Francisco, CA: Jossey-Bass.
- House, Ernest R. and Kenneth R. Howe (1999) *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.
- House, Ernest R. and Kenneth R. Howe (2000) 'Deliberative Democratic Evaluation', in K. R. Ryan and L. DeStefano (eds) *Evaluation as a Democratic Process: Promoting*



Greene et al.: The Merits of Mixing Methods in Evaluation

- Inclusion, Dialogue, and Deliberation. New Directions for Evaluation No. 85:* 3–12. San Francisco, CA: Jossey-Bass.
- Krantz, D. L. (1995) 'Sustaining Versus Resolving the Quantitative-Qualitative Debate', *Evaluation and Program Planning* 18: 89–96.
- Li, S. and C. Zercher (1997) 'Data Analysis Strategies and Issues in Mixed-Method Research', paper presented at the annual meeting of the American Evaluation Association, San Diego, CA.
- Mark, Melvin M., Irwin Feller and Scott B. Button (1997) 'Integrating Qualitative Methods in a Predominantly Quantitative Evaluation: A Case Study and Some Reflections', in J. C. Greene and V. J. Caracelli (eds) *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms, New Directions for Evaluation No. 74:* 47–59. San Francisco, CA: Jossey-Bass.
- Mark, Melvin M., Gary T. Henry and George Julnes (2000) *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco, CA: Jossey-Bass.
- Marquart, J. M. (1997) 'Mixed-Methods Studies: Design and Analysis Dilemmas and Solutions', paper presented at the annual meeting of the American Evaluation Association, San Diego, CA.
- Marquart, Jules M., S. Li and C. Zercher (1997) 'Making the Whole More than the Sum of the Parts: Challenges in a Mixed-Method Study of Inclusion', paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Marquart, Jules M., S. Li and C. Zercher (2000) 'Conceptual Issues and Analytic Strategies in Mixed-Method Studies of Preschool Inclusion', *Journal of Early Intervention* 23: 116–32.
- Marris, Peter and Martin Rein (1982) *Dilemmas of Social Reform: Poverty and Community Action in the United States*, 2nd edition. Chicago, IL: University of Chicago Press.
- Patton, Michael Q. (1988) 'Paradigms and Pragmatism', in D. M. Fetterman (ed.) *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*, pp. 116–37. New York: Praeger.
- Powson, Ray and Nick Tilley (1997) *Realistic Evaluation*. London: Sage.
- Popham, W. James (1999) *Modern Educational Measurement: Practical Guidelines for Educational Leaders*. New York: Allyn and Bacon.
- Ragin, Charles C. (1989) *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- Shotland, R. Lance and Melvin M. Mark (1987) 'Improving Inferences from Multiple Methods', in M. M. Mark and R. L. Shotland (eds) *Multiple Methods in Program Evaluation, New Directions for Evaluation No. 35:* 77–94. San Francisco, CA: Jossey-Bass.
- Sivalingham, Vanaja (1997) 'Examining Claims Made for Performance Assessments Using a High School Science Context', doctoral dissertation, Department of Education, Cornell University, Ithaca, NY.
- Smith, Mary Lee (1997) 'Mixing and Matching: Methods and Models', in J. C. Greene and V. J. Caracelli (eds) *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms, New Directions for Evaluation. No. 74:* 73–86. San Francisco, CA: Jossey-Bass.
- Tashakkori, Abbas and Charles Teddlie (1998) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.
- Weisner, Thomas (2000) 'Understanding Better the Lives of Poor Families: Ethnographic and Survey Studies in the New Hope Experiment', *Poverty Research News* 4(1): 10–12. Joint Center for Poverty Research, Northwestern University, IL.
- Weiss, Carol H. (1998) *Evaluation*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.

*Evaluation 7(1)*

JENNIFER C. GREENE is a professor at the University of Illinois, Urbana-Champaign. Her evaluation research and practice features interpretive, participatory and mixed-method approaches, primarily in the domains of education and social programs for families and children. Please address correspondence to: Department of Educational Psychology, 260E Education, 1310 S. Sixth St., Champaign, IL 61801, USA. [email: [jcgreene@uiuc.edu](mailto:jcgreene@uiuc.edu)]

LEHN BENJAMIN is a doctoral candidate in the Department of City and Regional Planning at Cornell University. Her research interests include evaluation in the community development field, with a particular focus on evaluation as democratic dialogue. Her dissertation research focuses on how performance measurement is shaping community development. Please address correspondence to: Department of City and Regional Planning, Cornell University, 106 W. Sibley Hall, Ithaca, NY 14850, USA. [email: [lmb18@cornell.edu](mailto:lmb18@cornell.edu)]

LESLIE GOODYEAR recently received a PhD in Human Service Studies from Cornell University, and is currently a research associate and managing editor of *The Evaluation Exchange* at the Harvard Family Research Project. Her evaluation and research interests include the application of interpretive, participatory and mixed-method approaches in the areas of HIV/AIDS prevention education and educational and social programs for families and children. Please address correspondence to: Harvard Family Research Project, 38 Concord Avenue, Cambridge, MA 02138, USA. [email: [leslie\\_goodyear@harvard.edu](mailto:leslie_goodyear@harvard.edu)]