2007

Merging Qualitative and Quantitative Data in Mixed Methods Research: How To and Why Not

David L. Driscoll
Senior Public Health Scientist, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle park, NC 27709, DDriscoll@virginia.edu

Afua Appiah-Yeboah
Health Analyst, RTI International

Philip Salib
Health Analyst, RTI International

Douglas J. Rupert
Research Associate, RTI International

Articles

Merging Qualitative and Quantitative Data in Mixed Methods Research: How To and Why Not

David L. Driscoll[1], Afua Appiah-Yeboah[2], Philip Salib[2], and Douglas J. Rupert[3]

This study assesses the utility of mixed methods designs that integrate qualitative and quantitative data through a transformative process. Two strategies for collecting qualitative and quantitative datasets are described, and processes by which they can be merged are presented in detail. Some of the benefits of mixed methods designs are summarized and the shortcomings and challenges inherent in quantitizing qualitative data in mixed methods research are delineated.

KEYWORDS: study design, analytic design, mixed methods, program evaluation, quantitizing

Introduction

Researchers seeking associations between primarily quantitative biophysical and primarily qualitative sociocultural data, including environmental and natural resource anthropologists, can look to mixed method research designs for structured and tested integrative processes. Such designs have been used to augment traditional methods for assessing and monitoring the impacts of recreation and tourism on the physical environment (Mackay 2004). In a larger sense these designs could aid ecological and environmental anthropologists in their efforts to overcome lack of public engagement in, or denial of, linkages between human activities and their physical environments (Schmidt 2005).

We use the term *mixed methods research* here to refer to all procedures collecting and analyzing both quantitative and qualitative data in the context of a single study (*sensu lato* Tashakkori and Teddlie 2003). Our objectives are to describe how and why we conducted two mixed methods research designs, and to discuss some of the benefits and challenges of mixed method research. We hope to inspire further investigation and informed application of such designs.

Background

Researchers have been conducting mixed methods research for several decades, and referring to it by an array of names. Early articles on the application of such designs have referred to them as multi-method, integrated, hybrid, combined, and mixed methodology research (Creswell and Plano Clark 2007: 6). The basis for employing these designs are likewise varied, but they can be generally described as methods to expand the scope or breadth of research to offset the weaknesses of either approach alone (Blake 1989; Greene, Caracelli, and Graham 1989, Rossman and Wilson 1991).

The prospective mixed methods researcher will find a variety of classificatory metrics by which mixed methods research designs can be described. The designs have been differentiated by the level of prioritization of one form of data over the other, by the combination of data forms in the research process (such as during the collection or analysis phases), and by the timing of data collection, such as whether the quantitative and qualitative phases take place concurrently or sequentially, and if so, in what order (Creswell, Fetters, and Ivankova 2004; Datta 2001; Johnson and Christensen 2004; Tashakkori and Teddlie 2003). Some researchers have integrated several different metrics to create mixed methods

[1] Senior Public Health Scientist, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle park, NC 27709
[2] Health Analyst, RTI International
[3] Research Associate, RTI International

classificatory systems (see for example Creswell, Plano Clark, Guttman, and Hanson 2003; Johnson and Onwuegbuzie 2004). In sum, there is as of yet no discrete list of mixed methods design options, and so researchers should plan to develop a design that answers their own research questions within the constraints and boundaries of the study context (Johnson & Onwuegbuzie 2004: 20).

Some researchers have taken issue with the term *mixed methods* to describe research designs that consciously blend both approaches within or across the stages of the research process (Johnson and Onwuegbuzie 2004). They suggest the term *mixed model* be used to differentiate research designs integrating qualitative and quantitative data from those who merely employ both types of data. These include transformative designs that change one form of data into another (most often qualitative to quantitative data) so that the data collected by mixed methods designs can be merged (Caracelli and Green 1993; Onwuegbuzie and Teddlie 2003).

The term *quantitizing* has been coined to describe the process of transforming coded qualitative data into quantitative data and *qualitizing* to describe the process of converting quantitative data to qualitative data (Tashakkori and Teddlie 1998: 126). While some recent studies have explored the utility of research that integrates qualitative and quantitative data (e.g., Adamson et al. 2004; Sandelowski 2000; Weisner 2005), there remains a need for systematic information on how to actually carry out such transformative analytic designs.

This paper describes two transformative mixed methods research designs. The two designs fall on somewhat different ends of the mixed methods design spectrum related to when the data are collected. The first is a relatively simple design in which qualitative and quantitative data are collected concurrently. The other is a fairly complex sequential design. We draw on examples from a recent evaluation of a federal policy regarding safe immunization practice to describe how these designs have been applied in practice. Contractual limitations preclude a detailed description of that study, and the objectives and findings of and from this research are described only as they relate to design implementation.

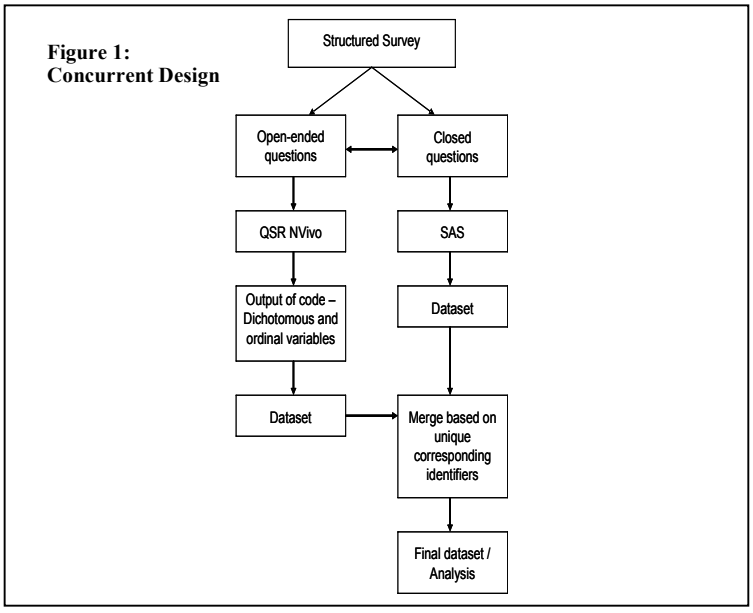
We begin by describing the data collection process employed for each design, then summarize the transformative analytic process we used for these designs, and finally describe the relative benefits and shortcomings of these designs and transformative mixed methods approaches in general.

DATA COLLECTION

Concurrent Design

Concurrent mixed method data collection strategies have been employed to validate one form of data with the other form, to transform the data for comparison, or to address different types of questions (Creswell & Plano Clark 2007: 118). In many cases the same individuals provide both qualitative and quantitative data so that the data can be more easily compared.

Figure 1:
Concurrent Design



This design was employed in a recent study (Figure 1) to collect and compare perceptions of vaccine safety among an extensive and varied set of stakeholder groups. The research questions involved levels of familiarity and agreement with various vaccine safety guidelines. Although the structured response categories

were identified in close consultation with vaccine specialists and agency officials, questions remained regarding the applicability of the response options for all advocacy groups. At the same time, the number and variety of survey respondents challenged pre-testing of the survey items. For these reasons we chose to employ a concurrent mixed methods research design involving a Web-based instrument to collect both structured and unstructured data. Each topic-specific set of structured questions in the survey instrument was followed by at least one open-ended and unlimited comment field, which was explicitly linked to the question set immediately preceding it. In most cases, the open-ended question asked: "What additional information would you like to provide to explain these responses?"

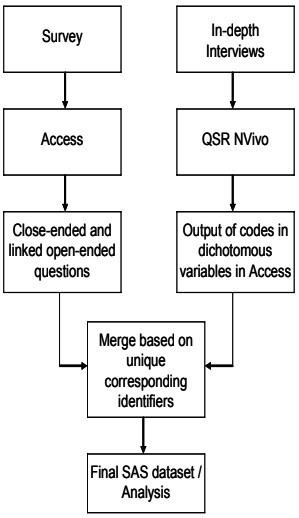
This data collection strategy has several advantages for mixed-methods applications. First, they can be fairly intuitive for participants. In the study described, the Web-based format was easy to understand and the open-ended response fields were unlimited, so many respondents took advantage of the resource to post extensive comments. Also, these fields were overtly linked to the preceding structured responses, facilitating linkage both by the participant during data collection and by the research team in relating the structured and unstructured responses. However, concurrent data collection designs preclude follow-up on interesting or confusing responses. In our study we relied entirely on respondents to augment their survey answers by following up on such issues. Many respondents did provide such follow-up, as described below, but some did not.

Sequential Design

Sequential mixed methods data collection strategies involve collecting data in an iterative process whereby the data collected in one phase contribute to the data collected in the next. Data were collected in these designs to provide more data about results from the earlier phase of data collection and analysis, to select participants who can best provide that data, or to generalize findings by verifying and augmenting study results from members of a defined population (Creswell & Plano Clark 2007:121). Sequential designs in which quantitative data are collected first can use statistical methods to determine which findings to augment in the next phase.

This design was employed in a recent study (Figure 2) to collect perceptions and attitudes regarding the utility of vaccine-safety guidelines from staff of several federal agencies with vaccine-safety missions. The study participants had various roles and disciplinary backgrounds and were associated with various federal agencies. Further, the prospective participants had very limited time available to respond to the study. For these reasons we chose to employ a flexible and iterative data collection strategy consisting of two data collection phases. In the first phase, we collected survey data; in the second phase, in-depth interview data. The survey questions were entirely close-ended, and the response categories were developed in consultation with representatives of the various federal agencies. The subsequent in-depth, semistructured interview instruments consisted of individualized questions intended to explore particularly interesting or ambiguous survey responses as well as standard questions exploring general perspectives on the purpose and future utility of vaccine safety guidelines.

Figure 1:
Sequential Design



This two-phased approach allowed study participants to respond to the survey on their own time and reduced the time required for in-depth discussions of emergent themes. It provided members of the research team with the opportunity to review and analyze the survey results and tailor the subsequent in-depth interview instrument to follow-up on confusing or significant responses. This iterative analytic approach also simplified subsequent attempts to integrate the coded qualitative data collected in in-depth interviews with survey data. A primary disadvantage of this strategy is the time required to design and conduct separate tailored instruments for each key informant. A second complicating factor is the lack of overt linkages between the structured and unstructured responses compared to the concurrent design.

Data analysis

There are several strategies by which qualitative data collected using the designs described can be quantitized to create a single comprehensive dataset. One of the more common strategies counts the number of times a qualitative code occurs. Some qualitative data analysis software programs (such as Atlas or NVivo) can generate these reports. Such quantitized frequencies can indicate particularly influential codes, but can also be prone to confounding by repetitive respondents who fix on a certain concept or theme. Other approaches to quantitizing qualitative data include enumerating the frequency of themes within a sample, the percentage of themes associated with a given category of respondent, or the percentage of people selecting specific themes (Onwuegbuzie & Teddlie 2003). In all these cases, the quantitized data can be statistically compared to the quantitative data collected separately.

Yet another strategy for quantitizing qualitative data enumerates whether or not qualitative responses included certain codes. In other words, rather than seeking to understand how many times a certain code was provided by each participant or the frequency with which they appeared, this strategy quantitizes the presence or absence of each code for each participant.

This was the strategy employed in the studies described earlier, and we will detail how the process was conducted. The application and transformation of qualitative to quantitative data owes some impetus to the development of software programs that allow qualitative researchers to process a large volume of qualitative data (Bazeley 1999). We used QSR NVivo2 to transform individual responses to our open-ended survey and interview questions into a series of coded response categories that were, in turn, quantified as binary codes and integrated into the associated survey responses. This process involved four analytic steps:

1. The survey data were entered into an Access database (Figure 3). This process was fairly straightforward and similar to that used to manage any structured database.

2. The qualitative data were analyzed for codes or themes using NVivo. These codes were then developed into qualitative response categories that were entered into a second Access database (Figure 4).

3. These two databases were linked by key informant identification numbers to ensure that each record contained both the survey and in-depth interview data.

4. The coded qualitative data were then quantified into dichotomous variables 0 or 1 based on absence or presence of each coded response.

5. Associations were analyzed using SAS.

The quantitization and data entry in Step 2 involves transition of actual codes into dichotomous variable: either 0 or 1, corresponding to absence or presence of prospective coded responses to each question. Many qualitative data analysis software packages quantify participant attributes, such as demographic data or response frequencies, but few allow for the quantitization of coded response data

Figure 3: Access Database for Survey Data

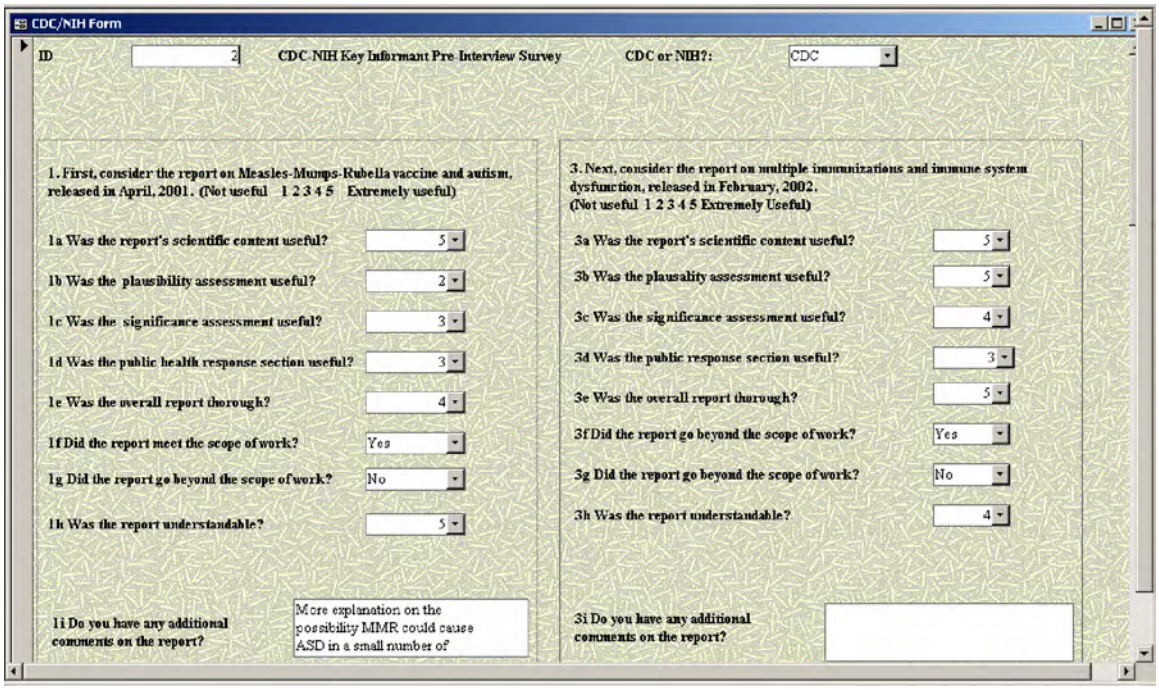
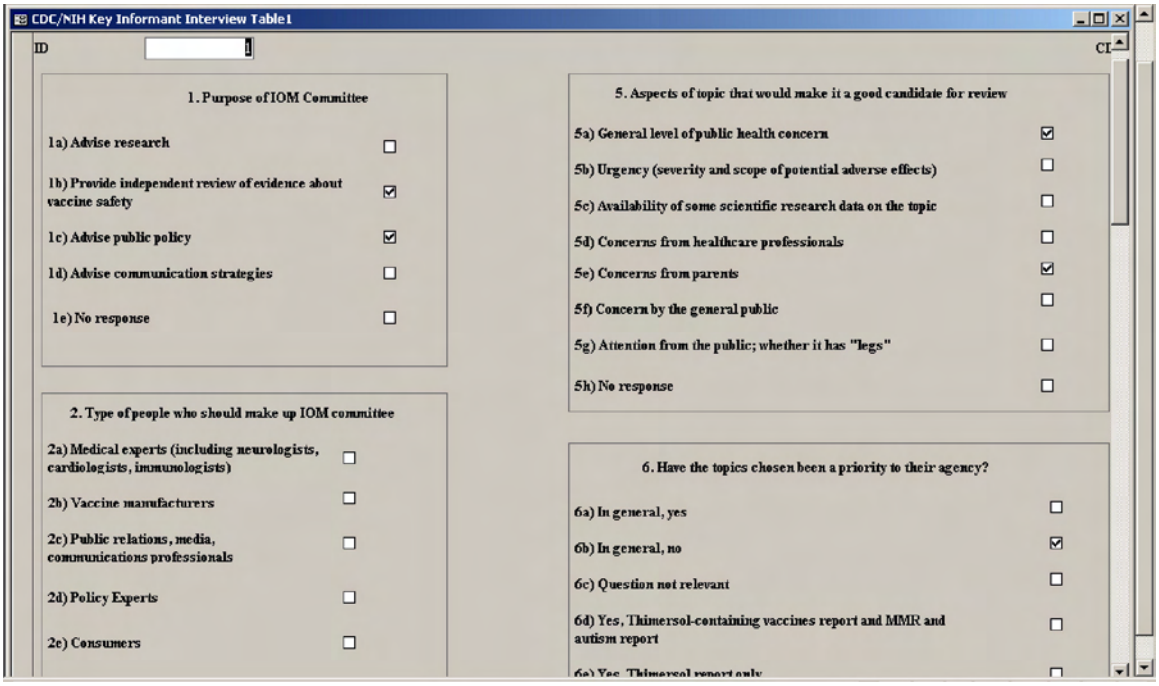


Figure 4: Access Database for Qualitative Data Codes



Results

While many studies have described transformative designs, few have focused on their advantages and disadvantages. Our discussion of this matter here should not be considered an argument against the use of such designs. We are interested in providing information that allows prospective researchers, specifically those working in the environmental arena, to make informed decisions about whether or not to apply these designs in their research.

Advantages

Concurrent Design. The collection and analysis of embedded qualitative responses can augment and explain complex or contradictory survey responses. For example, structured responses to our survey of stakeholders indicated strong support for specific immunization guidelines. In the survey responses more than 90% of respondents either "agreed" or "strongly agreed" with a statement that the process by which the guidelines were determined was objective and rigorous. An approximately equal proportion affirmed that "there is a continuing need" for such assessments and guidelines. However, more than 30% of respondents either "somewhat disagreed" or "strongly disagreed" with many of the immunization safety guidelines resulting from this process.

Figure 5. Evidently Contradictory Survey Findings	**Ordinal Survey Responses**		
	Strongly Agree or Agree	**Neither Agree no Disagree**	**Strongly Disagree or Disagree**
Vaccine Safety Review Process is Scientifically Rigorous and Objective	90%	5%	5%
Continuing Need Exists for the Vaccine Safety Review Process	90%	10%	0
Recent Vaccine Safety Decisions Are Accurate	45%	25%	30%

This evident contradiction in survey responses was explained by the qualitative data. The coded open-ended responses revealed that the primary reason for disagreement with specific guidelines centered on the information used in the review process. Respondents who disagreed with the guidelines believed them to be based on minimal or limited scientific evidence. Thus, the evaluation highlighted a disagreement with the inclusion criteria used in the review process. This finding was not anticipated by the structured responses provided in the structured instrument despite extensive review by vaccine specialists and agency representatives.

Sequential Design. The collection and analysis of structured survey and open-ended key informant interviews in an iterative analytic process can provide important information on emergent and unexpected themes. For example, a statistical analysis of the combined survey and quantitized interview responses in our sequential design revealed a significant, and heretofore unrealized, association between the perceived utility of the vaccine guidelines and their audience. Initial analysis of the survey responses demonstrated significant differences in satisfaction with the readability and utility of the guidelines by agency. Many used the open-ended response fields to detail their concerns or satisfaction with the format and content of the reports, but these fields did not explain the disparity by organizational affiliation.

We explored this disparity in subsequent in-depth interviews, and the resulting themes were quantitized and integrated with the survey responses. Subsequent analysis revealed that those agency officials most satisfied with the reports viewed the scientific community as the major audience. Those who felt that the general public and policy makers were the main audiences felt that the reports were too dense or complicated to be readily understandable. The use of simple statistical measures of association, in this case chi square, thus allowed for the discovery of an important difference of opinion that may have been missed without the iterative combination of structured and unstructured data.

Disadvantages

While there are demonstrated benefits to the transformative mixed methods designs, there are several limitations and challenges as well. We will start with a disadvantage commonly voiced by qualitative researchers: the loss of depth and flexibility that occurs when qualitative data are quantitized. Qualitative codes are multidimensional, meaning they can and do provide insights into a host of interrelated conceptual themes or issues during analysis (Bazeley 2004). Codes can also be revisited during analysis in an iterative analytic process to allow for the recognition of emergent themes and insights. Conversely, quantitized data are fixed and one-dimensional; that is, they are composed of a single set of responses prospectively representing a conceptual category determined prior to data collection. They cannot change in response to new insights in analysis. In short, reducing rich qualitative data to dichotomous variables renders them single dimensional and immutable.

This is a serious challenge to transformative mixed methods designs. In theory, mixed method researchers who quantitize qualitative data need only to avoid focusing on the quantitative dataset to the exclusion of the original qualitative data to avoid this problem. In practice, however, this can be difficult. Analyzing, coding, and integrating unstructured with structured data is a complex and time-consuming process. The prospect of reconsidering and potentially reconfiguring the coding scheme can be unappealing after the team has begun statistical analyses with the existing dataset. We found it helpful to return to discrete and topically-bounded qualitative responses associated with significant statistical findings rather than to the entire qualitative dataset. For example, we reviewed agency representative responses about the utility of vaccine guidelines to verify the coding structure before accepting the association between such responses and agency affiliation.

A second broad category of challenges to mixed methods research, commonly leveled by quantitative researchers, concerns the limitations of quantitized qualitative data for statistical measurement. First, these data are vulnerable to the problem of collinearity, wherein response categories are themselves linked as a consequence of the coding strategy (Roberts 2000). Second, the need to collect and analyze qualitative data can force researchers to reduce sample size, which can curtail the kinds of statistical procedures that might reasonably be used, particularly the more rigorous parametric measures of association, such as t-tests and analyses of variance.

Collinearity is a problem for even the most forgiving measures of statistical association including the nonparametric tests chi-square, a workhorse for bivariate tabular analysis of a wide variety of research data. Mixed methods researchers, however, can largely avoid collinearity associated with quantitization by identifying and separating dichotomized codes derived from a single open-ended question in subsequent statistical analyses. In the studies described above we employed simple statistical measures of association only for response categories collected in different phases and with different questions. The problem related to sample size, however, is a serious challenge for mixed methods studies involving quantitization. Prospective mixed methods researchers should be aware of the sample size required to provide sufficient statistical power for the study question, and whether the study parameters will allow for the inclusion of quantitized qualitative data. If not, they might consider mixed methods designs not requiring data transformation.

Conclusion

Mixed methods designs can provide pragmatic advantages when exploring complex research questions. The qualitative data provide a deep understanding of survey responses, and statistical analysis can provide detailed assessment of patterns of responses. However, the analytic process of combining qualitative and survey data by quantitizing qualitative data can be time consuming and expensive and thus may lead researchers working under tight budgetary or time constraints to reduce sample sizes or limit the time spent interviewing. Ultimately, these designs seem most appropriate for research that does not require either extensive, deep analysis of qualitative data or multivariate analysis of quantitative data.

This study demonstrates some techniques for and outcomes from mixed methods research designs involving quantitizing qualitative data. The strategies employed had some commonalities. For example, open-ended survey responses and in-depth interview data were coded using an analytic software package. The data collected for each task were integrated using a data management software package. These combined data could then be assessed using simple measures of frequency to explain apparent discrepancies by providing contextual data on what survey responses actually meant (i.e., stakeholders' disagreement with guidelines) and reveal determinants of various responses (i.e., key informants' conceptions of the major audience for guidelines).

This form of sequential data collection may be of use for environmental researchers involved in descriptive studies of readily-structured biological or environmental measures as patterns of resource use or activity and of social metrics that may defy easy categorization, such as potentially-related perceptions, attitudes, or beliefs. The opportunity to provide additional qualitative information augmenting structured responses can provide key insights into unexpected relationships between local resource use patterns and community factors. This is but one example, and researchers interested in applying this design could revise the structure as necessary to be responsive to their particular study objectives and parameters.

Acknowledgements

The authors gratefully acknowledge Asta Sorensen, MA, for her comments and word processing contribution.

References

Adamson, J., R. Gooberman-Hill, G. Woolhead, and J. Donovan 2004 Questerviews: Using Questionnaires in Qualitative Interviews as a Methods of Integrating Qualitative and Quantitative Health Services Research. Journal of Health Services Research and Policy 9(3):139-145.

Bazeley, Pat 1999 The *Bricoleur* With a Computer: Piecing Together Qualitative and Quantitative Data. Qualitative Health Researcher 9(2):279–287.

Bazeley, Pat 2004 Issues in Mixing Qualitative and Quantitative Approaches to Research. *In* Applying Qualitative Methods to Marketing Management Research. R. Buber, J. Gadner, eds. Pp 141–156. Hampshire, United Kingdom: Palgrave Macmillan.

Blake, R. 1989 Integrating Quantitative and Qualitative Methods in Family Research. Families Systems and Health 7:411–427.

Caracelli, V. J., and J. C. Greene 1993 Data Analysis Strategies for Mixed-Method Evaluation Designs. Educational Evaluation and Policy Analysis 15(2): 195-207.

Creswell, J. W., V. L. Plano Clark, M. Gutmann, and W. Hanson 2003 Advanced Mixed Methods Research Designs. *In* Handbook of Mixed Methods in Social and Behavioral Research. A. Tashakkori and C. Teddlie, eds. Pp. 619-637. Thousand Oaks, CA: Sage.

Creswell, John, Michael Fetters, and Nataliya Ivankova 2004 Designing a Mixed Methods Study In Primary Care. Annals of Family Medicine 2:7–12.

Creswell, John W., and Vicki L. Plano Clark 2007 Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage Publications.

Datta, Lois-ellin 2001 The Wheelbarrow, the Mosaic, and the Double Helix. Evaluation Journal of Australasia 1(2):33–40.

Greene, J., V. Caracelli, and W. Graham 1989 Toward a Conceptual Framework for Mixed-Methods Evaluation Designs. Educational Evaluation and Policy Analysis 11:255-274.

Johnson, Jeffrey C., ed. 1990 Selecting Ethnographic Informants. Thousand Oaks, CA: Sage Publications.

Johnson, B., and L. Christensen 2004 Educational Research: Quantitative, Qualitative, and Mixed Approaches (2nd Edition). Boston, MA: Pearson Education, Inc.

Johnson, R. B., and A. J. Onwuegbuzie 2004 Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educational Researcher 33(7):14-26

MacKay, Kelly J., and J. Michael Campbell 2004 A Mixed-Methods Approach for Measuring Environmental Impacts in Nature-Based Tourism and Outdoor Recreation Settings. Tourism Analysis; 9(3):141-152.

National Institutes of Health, Office of Behavioral and Social Science Research 1999 Qualitative Methods in Health Research. Opportunities and Considerations in Application and Review. Washington, DC: National Institutes of Health

Onwuegbuzie, A. J., and C. Teddlie 2003 A Framework for Analyzing Data in Mixed Methods Research. *In* Handbook of Mixed Methods in Social and Behavioral Research. A. Tashakkori and C. Teddlie, eds. Pp. 351-383. Thousand Oaks, CA: Sage.

Roberts, C.W. 2000 A Conceptual Framework for Quantitative Text Analysis. Quality and Quantity 34:259–274.

Rossman, G., and B. Wilson 1991 Numbers and Words Revisited: Being "Shamelessly Eclectic." Evaluation Review 9(5):627–643.

Sandelowski, Margaret 2000 Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Methods Studies. Research in Nursing and Health 23:246–255.

Schmidt, G. 2005 Ecology & Anthropology: A Field Without a Future? Ecological and Environmental Anthropology 1(1):13-15.

Tashakkori, A., and C. Teddlie 1998 Mixed Methodsology: Combining Qualitative and Quantitative Approaches. Thousand Oaks, CA: Sage Publications.

Tashakkori, A., and Teddlie C. 2003 Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks, CA: Sage.

Weisner T.S. 2005 Discovering Successful Pathways in Children's Development: New Methods in the Study of Childhood and Family Life. Chicago: University of Chicago Press.