Copyright 2006 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2006, Vol. 13, No. 1, 84-92

Enhancing the Policy Process Through the Use of Ethnography and Other Study Frameworks: A Mixed-Method Strategy[1]

Valerie J. Caracelli
U.S. Government Accountability Office

This article examines the current context of federal program evaluation and the prominence given to evaluation for program accountability purposes. Mixed-method studies have been recognized for their capacity to strengthen evaluation evidence. While ethnographic studies can be crafted as mixed-method designs, it is also the case that they can be added as another component or integrated into a different study framework (such as a randomized experiment or survey) creating other mixed-method designs. Several examples illustrate how together the contrasting method frameworks can complement and strengthen the evaluation evidence, increase accuracy, provide important information about context, and enhance explanation and confidence in findings. A pragmatic approach to mixed-method evaluation that weighs the resource issues and potential benefits in using such designs is recommended.

The ushering in of the 21st century brought unexpected challenges to the United States with 9/11, changing population demographics burdening the health care system, new technologies, program efforts focused on combating terrorism, and issues related to coordination and cooperation among agencies in disaster assistance (e.g., the federal response to Hurricane Katrina). Tensions proliferate in current evaluation practice. For example, today the capacity for electronic record linkage, while offering potential benefits in tracking program outcomes and informing policy debates, can also create new information at the level of the individual that has the potential to violate an individual's right to privacy. Current demands to demonstrate clear evidence of program effectiveness fosters a competitive programmatic environment given diminished evaluation resources (U.S. Government Accountability Office [U.S. GAO], 2001; U.S. GAO, 2005a; U.S. GAO, 2005b; Lane, 2006).

It is in this tumultuous context that evaluation takes place. Evaluation studies are expected to produce information that can be used to make critical decisions about programs so that scarce resources are invested in programs that benefit the country and its citizenry, and problematic programs are improved or replaced. Federal evaluations frequently rely on multiple and mixed-method studies capable of providing defensible evidence coupled with a sophisticated understanding of the context. Ethnography is particularly suited for providing a framework for examining not only purposeful results-based use, conceptual use, or political or persuasive use of evaluation results but also sources of influence that emanate from the process of conducing the evaluation. (See Kirkhart, 2000, for an explication of an expanded understanding of evaluation use that is modeled as an integrated theory of influence.) Further, process-based influence has a political dimension that can reveal dynamics of power and privilege in the evaluand, provide a forum for all voices and opportunity for constructive dialog, or, at the least, provide the potential for raising issues and drawing attention to social problems (Greene, 1988).

The present paper examines the shift in evaluation toward accountability, the concomitant preference for designs—such as randomized controlled field trials—that have the potential to determine whether desired effects, if obtained, are attributable to the program, and examines the utility of mixed-method studies that include ethnographic methods to assure contextual understanding.

Shift in Federal Evaluation Toward Evaluation for Accountability Purposes

Evaluations may be conducted for a number of different purposes that are generally characterized as falling within three general perspectives: (a) evaluation for *development* purposes emphasizing building capacity and improving institutional performance; (b) evaluation for *accountability* purposes involving the measurement of results or efficiency, and (c) evaluation for *knowledge* purposes to develop a deeper

Correspondence for this article should be addressed to Valerie J. Caracelli, U. S. Government Accountability Office, 2727 29th St. N.W Apt. 512, Washington, DC 20008.
Email: CaracelliV@gao.gov

understanding of the factors underlying social problems and the theory and logic inherent in programmatic solutions (Chelimsky, 1997). At the federal level, the prominence given to evaluation for program accountability purposes was already underway in the 1990s when Congress and the executive branch put in place a statutory and management framework for strengthening government performance (U.S. GAO, 1998; U.S. GAO, 2004b). The various reform initiatives such as the Government Performance and Results Act of 1993 (commonly referred to as GPRA or the Results Act; P.L. 103-62) shifted the emphasis from the characteristics of program constituents and the services they receive (process issues) to an increased demand for information on program effectiveness. This shift in perception about what counts as useful information for decision-making was mirrored at state and local levels (U.S. GAO, 1994).

The accountability and program results orientation of federal program evaluation has been reinforced over the last several years with the use of the Office of Management and Budget's (OMB) Program Assessment Rating Tool (PART). PART comprises a standard series of questions used by budget examiners to allow for a consistent approach to assess federal programs in the budget formulation process. PART draws on program performance and evaluation information, scores agencies as to the effectiveness of agency programs and recommends improvements in program design, management, and assessment (U.S. GAO, 2004a; U.S. GAO, 2005c). Further, OMB's guidance to agencies specifies a desire for strong evidence of a program's effectiveness noting this is best achieved through randomized controlled trials (RCTs). Even though the revised guidance acknowledged that RCTs are not always possible or practical to carry out, it reinforced the description of RCTs as the "gold standard" for assessing an intervention's effectiveness in such fields as medicine, welfare and employment, psychology, and education (OMB, 2004).

However, contentious views about the practical capabilities and limitations of RCTs, compared to other designs, have been expressed (Brass, Nunez-Neto, & Williams, 2006). Without clear guidance on alternative methods and the situations for which they are suited, the perception that RCTs were preferred and that strong evidence is necessarily quantifiable evidence was manifested in agencies' responses to accountability for results. In fact, a vitriolic debate among experts within the American Evaluation Association (AEA) took place after the U.S. Department of Education's Institute of Education Sciences took a stance on prioritizing funding for experimental and some types of quasi-experimental designs over other methods. (For a recap of the debate on whether RCTs should be privileged in funding competitions see Donaldson & Christie, 2004.) In another voice, Chatterji (2004) advocated for the use of multiple methods, effectively implemented, in order to secure quality evidence from studies providing causal explanations of "what works" in education.

The U.S. Government Accountability Office is a legislative branch agency that conducts evaluations, among other types of studies, to improve federal programs and to assure their accountability to Congress and the U.S. Public. A recent study was initiated to determine agencies' progress in implementing OMB's PART-related recommendations for evaluation. The study reported that some program managers disagreed with OMB on the scope, purpose, and quality of their evaluations as well as the usefulness of evaluations by independent third parties unfamiliar with their programs. Several program officials noted that, in designing their evaluations, they were more concerned with learning how to improve their programs than in meeting an OMB checklist (U.S. GAO, 2005d). Federal Evaluators, an informal association of evaluation officials in the U.S. Government, believed that OMB guidance materials reflected too narrow a range of rigorous evaluation designs. Federal Evaluators shared presentations of alternative designs with OMB staff better suited to various program purposes and situations where randomized field trials may not be appropriate (Bernholz et al., 2005).

Holding Evaluation Designs Accountable

At the federal level, randomized field trials have a long history of use in assessing program effectiveness (Boruch, 1997, 2005); equally illustrious is the history of the use of ethnography in government studies. Both types of designs have been criticized because they are resource intensive, yet both have also been used productively together as a particular type of mixed-method design. The randomized experiment is notably strong in internal validity but constrained by experimental conditions that may limit generalization; moreover, such studies are not always immune to sources of bias such as differential reactivity and biasing social interactions (see Droitcour & Kovar, in press; Fetterman, 1982). The randomized experiment can be strengthened through cross-design techniques that, for example, may use other sources of data to adjust for bias (Droitcour, 1997). Adding or embedding an ethnographic component can provide a more comprehensive study that uses the ethnography to respond to particular study questions, to detail

unfolding explanatory processes during program implementation, or to offer a means by which a more contextualized understanding and thorough explanation of experimental findings can be achieved (Cook, 2001; Cronbach, 1982; see also Onwuegbuzie & Johnson, this volume, for a range of quality issues specific to mixed-method evaluation research).

Crafting Mixed-Method Evaluations Using Ethnography and Other Study Frameworks

Mixed-method designs are now recognized for their capacity to serve a number of purposes that strengthen evaluation evidence. Many examples of such designs have appeared in the literature, but there is no hierarchy of preferred designs yielding the strongest evidence. Nor would this be a reasonable expectation. It is precisely the flexibility to craft the best design options, including both qualitative and quantitative forms of evidence, for specific problems and questions that is the strength of mixed methods designs. Such designs may be intra-disciplinary or inter-disciplinary. As Morse (2003) pointed out, an ethnography with methodological congruence can include both qualitative forms of evidence (e.g., in the form of interviews, observations, diaries) and quantitative evidence (e.g., psychometric tests and scales, biological measurements) to answer a research question or enrich explanation. Studies rooted in other social science disciplines that call on methods not characterized as ethnographic (e.g., studies in health, child development, education, criminal justice) can add ethnography to their methods-base to allow for a more in-depth understanding of the phenomena under study. So, although an ethnographic study can be crafted as mixed-method in its design and execution, such a study can also be added as another component or integrated into a different study framework (e.g., randomized experiment, survey).

In previous mixed-method work, the benefits of ethnographic case studies as a complement to quantitative approaches has been illustrated. Ethnographic methods were used to deeply probe quantitative findings and increase understanding and confidence in them (Caracelli & Greene, 1997; Greene, Caracelli, & Graham, 1989). Examples included examining extreme cases that were revealed through regression analysis (Fry, Chantavanich, & Chantavanich, 1981), embedding ethnography into a study using a path-analytic framework in order to examine the influences of culture and environment (Jacob, 1982), and nesting one design within another to illuminate causal processes and increase the interpretability of experimental outcomes (Maxwell, Bashook, & Sandlow, 1986).

For example, in the context of cross-cultural educational research in Thailand, schools depicted as outliers when educational outcomes were analyzed in a multiple regression analysis were further studied using ethnographic case studies for the purpose of clarifying the initial theoretical model (Fry et al., 1981). Ethnographers blind to the quantitative findings were sent to "extreme case" schools to study their educational processes. Ethnographic variables derived from insights not considered in the initial analysis, such as differences in teaching methods, principal characteristics, community support, or other factors important in assessing school effectiveness were incorporated back into the regression model to increase its explanatory power. The technique of "ethnographic residual analysis" allowed for a better understanding of the complex network of relationships among educational inputs, processes, and outcomes (Fry et al., 1981). The potential of such a mixed-method study and integrated analysis strategy is in its capacity to refine, recast, or elaborate initial theoretical assumptions (Caracelli & Greene, 1993).

Providing Policy Relevant Information Using Multiple Methods and Ethnographic Case Studies

Today, with an emphasis on program performance and results, it has become incumbent upon agencies to provide evaluation results that demonstrate how well programs are working and/or how to improve agency programs, policies, and procedures. Datta (1997) illustrated how the GAO evaluation of the H-2A Farmworkers program required a mixed-method approach with an ethnographic component in order to be fully responsive to questions needed for a hearing on changes in immigration law. The H-2A program permitted U.S. growers of perishable crops to hire temporary farmworkers from other countries so long as no U.S. citizens or legal immigrants were available to pick the crops. Among other questions, the evaluation focused on whether the H-2A program successfully protected U.S. citizens and legal immigrant farmworkers' employment opportunities, wages, and working conditions. A multidisciplinary team and several bilingual staff were assembled to implement a multiple methods study consisting of ethnographic case study, historical analysis, field test of worker availability, technical reviews, and secondary analysis of wage data. (The complete methodology can be found in U.S. GAO 1998b.)

The evaluators found only surface compliance with the H-2A law. The ethnographic case studies indicated that in one county the local workers were forced to look for other employment owing to a disparity in working conditions intentionally favoring the seasonal farmworkers. In another county, lower

rates of H-2A farmworkers were not the result of good working conditions favoring jobs for U.S. and legal immigrant workers but instead resulted from illegal fieldworkers allegedly being employed. Importantly, the ethnographic case study and participant observer methodology was viewed as essential to the deep understanding acquired on how the law was being carried out.

Although ethnographic methods have had a long history of use in the federal government, efforts to document and analyze how such methods have contributed to program improvement are still needed. Recently, there has been an acknowledgement about how ethnographic studies can inform agency actions and how it can be used to study culture in organizations. As Patton (2005) noted, the organizational culture is often the context for the program culture; thus, organizational effectiveness and program effectiveness are often interdependent.

Using Ethnographic Methods to Illuminate Organizational Culture and Transformation

A transformative shift in agency culture can be captured using ethnographic methods. Under study at GAO was the Department of Veterans Affairs (VA) Patient Safety Program introduced in 1999 to fix system flaws that could harm patients (U.S. GAO, 2004c). Reports about unnecessary deaths and expenditures from accidents in hospitals, the attendant media coverage, and public outcry gave rise to a program intended to minimize the catastrophic effects of physician and staff error. The VA program relied on staff openness to report close calls and adverse events. In the past such reports resulted in hospital employees being held responsible for adverse patient outcomes even when mistakes were of a systemic nature. Success in implementing the new safety initiative required a cultural shift from fear of punishment to mutual trust and comfort in reporting adverse events.

The challenge of determining whether an organizational cultural shift supporting the program had occurred was undertaken at four medical facilities. Ethnographic methods (site visits, including in-depth interviews with VA staff, observations of daily routines and meetings, and document review) were used for the purpose of providing knowledge of organizational culture from the perspective of VA's physicians, nurses, and an array of others responsible for patient care. GAO found that progress in program implementation varied at the different facilities, and this stemmed from significant differences in clinicians' familiarity with and participation in the program. Also, while mainly positive, the levels of cultural support for the program differed among physicians, and at one facility the culture blocked physician participation. Still, three of the four facilities had a supportive culture so that staff were trusting and comfortable about reporting close calls and adverse events.

In its recommendations, GAO encouraged the VA to set goals for increasing staff familiarity with the program's major concepts (close call reporting, confidential reporting program, root cause analysis) and mutual trust, measure goals by facility, and develop interventions when goals are not met. The use of ethnographic methods to study culture change surfaced a process model for how the Patient Safety Program would lead to desired outcomes through (a) clinicians with cultural support for reporting adverse events and close calls, (b) teams that investigate root causes, (c) systems change, and (d) feedback and reward systems to encourage reporting, resulting in (e) patients who are safer.

In addition to the actual VA study, also of interest is a description of how the study team was able to create a sense of "buy-in" by GAO managers by demonstrating how an ethnographic approach can be as valuable as more typical performance audits used in GAO engagements and could, if appropriate, complement and enhance such audits (Goodman, Trainor, & Divorski, 2005). First, critical to accepting their methodology was whether the ethnographic component could meet stringent audit documentation requirements. The multidisciplinary team cross-validated coding definitions, clarified their interpretations through interrater reliability techniques, and provided audit-trail documentation that was reviewed by a referencer independent of the team for quality assurance purposes. Second, time constraints attend congressional requests and these constraints create challenges considering that ethnography is characterized by prolonged engagement at sites. However, rapid ethnographic assessment is gaining popularity, making ethnographic methodology more amenable to policy time frames. Rapid assessment process (RAP) used by a team offers a team-based ethnographic inquiry that relies on iterative data analysis, including data triangulation, and additional data collection to develop a preliminary "emic" understanding of a situation relatively quickly (Beebe, 2001).

A Coming of Age in the 21st Century: The Diverse Practice of Ethnography

Potential benefits and limitations of a rapid ethnographic assessment in comparison to traditional ethnography for examining organizational culture,

organizational change, and interrelationships among organizational levels is an area ripe for research. Hammersley (2006), in describing current practice, highlights areas of tension and conflict in ethnography in the social sciences, including educational research, notes the movement from the older anthropological model of ethnographic fieldwork to more recent forms that may attempt only a partial study of people's lives over a relatively short time frame (e.g., months rather than years), with part-time participant observation. These changes reflect increasing pressure to demonstrate productivity, shorter contract time-frames, and uses of new technology. Hammersley, while acknowledging the integral place ethnographic case studies have occupied in conjunction with quantitative methodology, in the study of schools in the U. K., cautions that this shift in practice requires conscientious attention to potential problems of sampling, generalization, cyclical variability, and fundamental patterns of change. Nevertheless, technological changes that allow for the use of portable audio and video recording devices and data analysis software packages that have been developed to assist in the analysis of large quantities of data collected are now part of the evaluator's tool chest and contribute to variation in how ethnography is practiced. Bazeley (this volume) describes the use of qualitative data analysis software packages that both facilitate the analysis of qualitative data and have the capacity for combining and integrating qualitative and quantitative data collected in studies that rely on mixed and/or multiple methods.

These new developments are likely to be looked upon with favor in the policy arena that favors receiving relevant information for decision-making in shorter time frames than traditional ethnographic methods can provide. As illustrated, an ethnographic approach has utility in studying cultural changes in organizations; however, ethnography also can be used concomitantly or sequentially with other study frameworks, within a mixed-method strategy, to secure information in complex study environments.

Responding to Federal Program Issues: Enhancing Mixed-Method Studies Through Inclusion of Ethnographic Methods

In 2003, GAO examined the range and scope of the use of ethnography in the federal government. Although it was recognized that ethnography has had a long history of application in the federal government, information about the past and present uses of this methodological approach to improve federal programs had not been systematically gathered or analyzed. While not exhaustive, GAO found that in 10 federal departments or agencies that employed ethnographic methods, the use of ethnography focused on understanding a group's sociocultural life with respect to an important federal program issue (U.S. GAO 2003). Ethnography was a method of choice when the program's operation or outcomes depended on the actions of a definable cultural community. Some of the benefits cited were the collection of new information, increased understanding of issues or problems important to a program, support for conclusions across sites, reliable identification of recurring themes, and the capacity to integrate ethnographic information with economic or other quantitative data. Several study examples used a mixed-method framework linking an ethnographic study to other quantitative data collection strategies to enhance the quality and accuracy of findings.

For example, the Bureau of the Census has used ethnography to conduct alternative enumerations of urban and rural hard-to-count populations. The Census is the premier source of information about U.S. population and is the basis for apportioning seats in the House of Representatives. Yet, impediments to participation resulting in undercounts, particularly among certain racial and ethnic groups, pose continuing challenges. Ethnography, in use by the Bureau since the 1960s, has offered a means by which the Bureau increases its understanding of why certain groups are undercounted so as to allow an alternative enumeration in the sites studied. Through the Statistical Research Division, the Bureau shed light on critical problems of data quality with ethnographic studies that document undercounts and other enumeration errors, factors contributing to them, and recommendations for improvement.

During the 1988 census "dress rehearsal," ethnographic evaluation studies were conducted in five sites documenting the day residence of all persons enumerated and comparing their observed count to the "official" count and later returned to the field to examine discrepancies. In 1990 ethnographic alternative enumerations in 29 rural and urban sites were undertaken by ethnographers with established relationships with the sites. A complete list of all housing and people in the area, recorded behavioral observations about the neighborhoods, and other factors that could prevent complete counts were studied. After matching ethnographers' counts with the Census count from returned Census forms, the ethnographers returned to the field to reconcile the differences. The ethnographic evaluation found that disparities in the accuracy of the counts for low-income minority populations were attributable to irregular housing units missed in the census, residential mobility, limited English proficiency, distrust of government, and ambiguous housing units (e.g., unrelated individuals or households with two or more

nuclear families) that made census residency rules difficult to apply. The combination of centralized control and standardized methods complemented by the ethnographers' local knowledge, expertise, and experience improved the accuracy of the estimation.

In another example, the Administration for Children and Families of the Department of Health and Human Services used ethnographic techniques in two national Early Head Start (EHS) evaluation sites to illuminate ways in which the families served accepted or rejected the program's Montessori Intervention. Early Head Start, a national evaluation with a traditional random assignment design with quantitative measures of process and outcome, included ethnographies at two sites to document more fully the program's socio-cultural context, inform the story of program implementation, and detail more fully what the intervention meant to parents and children living in poverty. One ethnographic study included (a) extensive participant-observation in the classrooms and (b) case studies chosen to illuminate patterns seen in the quantitative data on the impact of the program for 12 families. Preliminary results indicated, in contrast to what may have been believed about Montessori before the program's experience, that low-income parents appreciated and valued the Montessori approach and the growing independence and facility in daily routines that they saw the program produce in their children. This was important because typically the Montessori curriculum had been administered to mid-upper class rather than the study population of low-income children and also because the study confirmed the program's theory that families were changed by first changing the children (Spicer, McAllister, & Emde, 2001).

The other ethnographic study was a nested design in which community and policy developments influencing the operation of EHS were explored. In this study, shifts in the program's theory of change were examined by participant observation of the program activities and focus groups with program staff; ethnographic interviews with program families provided information about their program experiences and their understanding of key program contexts. Relationships among community context, program implementation, and family perspectives were clarified. The ethnography provided a greater understanding of family culture and elaborated the program's theory of change in changing community contexts, which resulted in expanded home-visiting services to informal child care providers and partnering with formal care providers to ensure quality and improve access. This was particularly important given the national evaluation had somewhat limited ability to assess the community context; yet, the EHS programs were required to tailor services to meet community needs. In both cases, the ethnographies provided insights into aspects of program process that were unanticipated in the design of the randomized trial. (A detailed description of methodology and results related to the study can be found in Mathematica, 2002.)

Concluding Considerations

As noted at the beginning of this article, federal evaluation is taking place in a policy environment that emphasizes evaluation for accountability purposes. Agencies' performance information is examined by OMB through the PART process, reinforcing the government's focus on program results. PART assessments summarize key performance data and findings, and agencies are scored on program purpose and design, strategic planning, program management, and program results (which account for 50% of the overall score). OMB then makes recommendations to improve performance and may request program assessments (e.g., to develop new measures or conduct program evaluations). As previously stated, the OMB guidance specifies a preference for randomized controlled trials when asking for a demonstration of the program's effectiveness but acknowledges that randomized controlled trials may not always be able to be carried out. As Greene (this volume) points out, an evaluator's sociopolitical commitments can influence the type of knowledge produced and the interests served by a particular approach to inquiry. Yet, the political arena has an influence of its own. Amid claims of scarce resources, the policy environment exerts an influence on the evaluation community through the types of evaluations that receive funding, the skills that are valued, and the request for proposals that are generated.

It is likely that, despite a preference for experimental evidence, policy makers use multiple criteria when trying to maximize the potential of policies and programs to ameliorate persistent problems faced by society, assure the accountability of programs, and effectively serve the needs of the citizenry. This article has elaborated on different, yet valued information that ethnography can provide. While it is the case that an ethnographic study can be designed as a mixed-method study, it is also the case that ethnography can serve as a component or can be integrated into a particular study framework that generates primarily quantitative or multiple forms of evidence (such as, a field experiment or longitudinal survey). Several examples have been provided that demonstrate how the contrasting types of evidence

complement each other and provide in some cases more accurate findings, in others more contextualized information, and in yet others explanation or generation of new insights.

The analytic frameworks and methods depicted here are rarely singly applied in an evaluation. Instead, they are arrayed in different configurations depending upon theories about the program and problems the program is intended to address, the particular questions that guide the evaluation, the skill set and resources of those conducting the study, the user needs for timely information, and other factors. As a result, Datta (1997) wisely advocated a pragmatic approach to the selection of design frameworks and incorporation of mixed-methods that considers the practical consequences of such decisions given the demands, opportunities, and constraints that the situation may place on carrying out a quality evaluation, optimizing resources, and assuring timely, useful findings. This is the approach recommended here as well.

The lead editors for this article were R. Burke Johnson and Anthony J. Onwuegbuzie.

References

Beebe, J. (2001). *Rapid assessment process: An introduction*. Walnut Creek, CA: Alta Mira Press.

Bernholz, E., et al. (2005). *Evaluation dialogue between OMB staff and federal evaluation leaders: Digging a bit deeper into evaluation science.* Washington, DC. Retrieved June 23, 2006, from http://www.fedeval.net

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation.* Thousand Oaks, CA: Sage.

Boruch, R. F. (2005). Preface: Better evaluation for evidence-based policy: Place randomized trials in education, criminology, welfare, and health. *Annals of the American Academy of Political and Social Science, 599*, 6–18.

Brass, C. T., Nunez-Neto, B., & Williams, E. D. (2006). *Congress and program evaluation: An overview of randomized controlled trials (RCTs) and related issues* (Congressional Research Service Report Order No. RL33301). Washington, DC. Available at http://www.opencrs.com/rpts/RL33301_20060307.pdf

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 15,* 195-207.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. In J.C. Greene & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (pp. 19-32). San Francisco: Jossey-Bass.

Chatterji, M. (2004). Evidence on “what works”: An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher*, *33*(9), 3-13.

Chelimsky, E. (1997). The coming transformations in evaluation. In E. Chelimksy & W. Shadish (Eds.), *Evaluation for the 21st century* (pp. 1-26). Thousand Oaks, CA: Sage.

Cook, T. (2001). A critical appraisal of the case against using experiments to assess school (or community) effects. *Education Next.* Retrieved June 23, 2006, from http://www.educationnext.org/unabridged/20013

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Datta, L-E. (1997). A pragmatic basis for mixed-method designs. In J.C. Greene & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (pp. 33-46). San Francisco: Jossey-Bass.

Donaldson, S. I., & Christie, C. A. (2004). *The 2004 Claremont debate: Lipsey vs. Scriven: Determining causality in program evaluation and applied research: Should experimental evidence by the gold standard?* Retrieved June 23, 2006, from http://www.cgu.edu/include/SBOS_2004_Debate.pdf (See also, JMDE, 1556-8180, NO. 3, Oct. 2005)

Droitcour, J. (1997). Cross-design synthesis: Concept and application. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 360–372). Thousand Oaks, CA: Sage.

Droitcour, J., & Kovar, M. G. (in press). Multiple threats to the validity of randomized studies. In P. Brandon & N. Smith (Eds.), *Fundamental issues in evaluation*. New York: Guilford.

Fetterman, D. M. (1982). Ibsen’s baths: Reactivity and insensitivity (a misapplication of the treatment-control design in a national evaluation). *Educational Evaluation and Policy Analysis*, *4,* 261–279.

Fry, G., Chantavanich, S., & Chantavanich, A. (1981). Merging quantitative and qualitative research techniques: Toward a new research paradigm.

Anthropology and Education Quarterly, 12, 145-158.

Goodman, C., Trainor, B., & Divorski, S. (2005). Using ethnographic methods to evaluate the Department of Veterans Affairs Patient Safety Program. *NAPA Bulletin, 24*(1), 57-70.

Greene, J. C. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, *12*, 91-116.

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

Hammersley, M. (2006). Ethnography: Problems and prospects. *Ethnography and Education, 1*(1), 3-14. Oxford, UK: Routledge, Taylor & Francis.

Jacob, E. (1982). Combining ethnographic and quantitative approaches: Suggestions and examples from a study on Puerto Rico. In P. Gilmore & A. A. Glatthorn (Eds.), *Children in and out of school: Ethnography and education* (pp. 124-147). Washington, DC: Center for Applied Linguistics.

Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V.J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use: New directions for evaluation, No. 88* (pp. 5-23). San Francisco, CA: Jossey-Bass.

Lane, J. (2006, May 17). *Optimizing the use of microdata: A new perspective on confidentiality and access* (abstract). Retrieved June 23, 2006 from http://www.scs.gmu.edu/~wss/Sem2006.html

Mathematica. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of early head start. Vol. I. Final technical report.* Washington, DC: Department of Health and Human Services, Administration on Children, Youth, and Families. Retrieved June 23, 2006, from http://www.acf.hhs.gov/programs/opre/ehs/ehs_resrch/reports/impacts_vol1/impacts_vol1_title.html or http://www.mathematica-mpr.com/earlycare/ehstoc.asp

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.

Morse, J. M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 189-208). Thousand Oaks, CA: Sage.

Office of Management and Budget. (2004). *What constitutes strong evidence of a program's effectiveness.* Retrieved June 23, 2006, from http://www.whitehouse.gov/omb/part/2004_program_eval.pdf

Patton, M. Q. (2005). The view from evaluation. *NAPA Bulletin, 24*(1), 31-40.

Spicer, P., McAllister, C. L., & Emde, R. (2001). Ethnography and the Early Head Start evaluation: Contributions from local research to understanding program processes. In Administration for Children and Families (Ed.), *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families, Vol. II: Technical report, appendixes* (pp. A.77-A.80). Washington, DC: U.S. Department of Health and Human Services.

U.S. Government Accountability Office. (1994, December). *Managing for results: State experiences provide insights for federal management reforms* (Publication No. GAO/GGD-95-22). Washington, DC: Author.

U.S. Government Accountability Office. (1998a, January). *Managing for results: The statutory framework for performance-based management and accountability* (Publication No. GAO/GGD/AIMD-98-52). Washington, DC: Author.

U.S. Government Accountability Office. (1998b, October). *The H-2A program: Protections for the U. S. farmworkers (Publication No.* PEMD-89-3). Washington, DC: Author.

U.S. Government Accountability Office. (2001, April). *Record linkage and privacy: Issues increating new federal research and statistical information* (Publication No. GAO-01-126SP). Washington, DC: Author.

U.S. Government Accountability Office. (2003, March). *Federal programs: Ethnographic studies can inform agencies' actions* (Publication No. GAO-03-455). Washington, DC: Author.

U.S. Government Accountability Office. (2004a, January). *Performance budgeting: Observations on the use of OMB's program assessment rating tool for the fiscal year 2004*

budget (Publication No. GAO-04-174). Washington, DC: Author.

U.S. Government Accountability Office. (2004b, March). *Results-oriented government: GPRA has established a solid foundation for achieving greater results* (Publication No. GAO-04-38). Washington, DC: Author.

U.S. Government Accountability Office. (2004c, December). *VA Patient Safety Program: A cultural perspective at four medical facilities* (Publication No. GAO-05-83). Washington, DC: Author.

U.S. Government Accountability Office. (2005a, February). *21st century challenges: Reexamining the base of government* (Publication No. GAO-05-325SP). Washington, DC: Author.

U.S. Government Accountability Office. (2005b, September). *Hurricane Katrina: Providing oversight of the nation's preparedness, response, and recovery activities* (Publication No. GAO-05-1053T). Washington, DC: Author.

U.S. Government Accountability Office. (2005c, October). *Performance budgeting: PART focuses attention on program performance, but more can be done to engage congress* (Publication No. GAO-06-28). Washington, DC: Author.

U.S. Government Accountability Office. (2005d, October). *Program evaluation: OMB's PART reviews increased agencies' attention to improving evidence of program results* (Publication No. GAO-06-67). Washington, DC: Author.

Notes

[1] The opinions in this article are the author's and should not be construed as the policy or position of the U.S. Government Accountability Office.