The Alberta Journal of Educational Research *Vol. XLV, No. 3, Fall 1999, 304-313*

Lyn M. Shulha

Robert J. Wilson
Queen's University

and

John O. Anderson
University of Victoria

Investigating Teachers' Assessment Practices: Exploratory, Non-Foundationalist, Mixed-Method Research

Audiences for educational research are rarely invited to review the decision-making processes that permeate systematic inquiry. This account of three researchers' efforts to make sense of preservice teachers' assessment practices reveals how methodologically diverse investigations emanated from the context of a single study. The developments, findings, and warrants that characterize this inquiry are presented, as well as arguments why in retrospect this study can be described as exploratory, mixed-method, non-foundationalist research. Consistent with the general themes of pragmatism, the reader is invited to subject the account of novice teachers' assessment practices to intense scrutiny and to judge the success of the researchers' efforts to build at least temporary certainty and primacy for claims about novice teachers' thinking and behavior. The same process can also be used with the suggested consequences for classroom assessment practice, which conclude both this article and the series.

Il est rare que l'on invite les publics de la recherche pédagogique à revoir le processus de prise de décisions qui imprègne les enquête systématiques. Cet article, qui explique les démarches qu'ont entreprises trois chercheurs pour comprendre les pratiques d'évaluation de stagiaires, révèle comment des enquêtes à méthodologies diverses sont sorties du contexte d'une seule étude. On présente les démarches, les résultats et les vérifications qui caractérisent cette enquête ainsi que des arguments qui affirment pourquoi, a postiori, cette étude peut se définir comme une recherche exploratoire, à méthodes mixtes et non-fondationaliste. Conformément aux thèmes généraux du pragmatisme, on invite le lecteur à examiner minutieusement le compte-rendu des pratiques d'évaluation des enseignants débutants et à juger si les chercheurs ont réussi à établir une certitude et une primauté - même temporaires - quant à leurs conclusions sur le comportement et les réflexions des enseignants débutants. La même approche peut servir pour la pratique d'évaluation en salles de classes. Cet article, ainsi que la série dont elle fait partie, termine par des suggestions quant aux implications des résultats.

We believe the study of preservice teachers' classroom assessment practices reported in this series of articles is significant in terms of substance, form, and possible extensions (Anderson, in press; Shulha, in press; Wilson & Martinus-

---

Lyn Shulha is an associate professor in the Faculty of Education. She specializes in program evaluation, planning and assessment, and curriculum.
Robert Wilson is a professor in the Faculty of Education. He specializes in assessment and evaluation.
John Anderson teaches measurement and evaluation in the Department of Educational Psychology and Leadership Studies, of which he is also the Chair.

sen, in press). The assessment of student achievement is a structural element of all formal classrooms. It is an activity that directly affects everyone who comes in contact with its requirements, its processes, and its outcomes. An understanding of how novice teachers go about making assessment decisions can help to target foundational notions that may anchor later classroom assessment practices. Unravelling these notions could have profound positive consequences for teachers, students, and for the broader educational community.

Our efforts at understanding novice behavior featured the implementation of a language arts portfolio. In asking 147 preservice teachers to track a student named Chris over an entire reporting period (as an apparent service to a regular classroom teacher), we immersed our participants in a simulation that mirrored professional practice. At the same time, the portfolio worked as an ecologically valid research instrument that allowed for the controlled variation of Chris's characteristics and responses over time.

The dialectic surrounding the design and implementation of the portfolio as a research tool focused on how to observe assessment behavior and how to access the meanings embedded in that behavior. An early decision allowed the research problem and not a particular research paradigm to give form to the questions, designs, strategies for data management, and analyses. What follows is a retrospective on the developments, findings, and warrants that would eventually characterize this as an exploratory, mixed-method, non-foundationalist approach to inquiry.

Reviewing Our Study

Prioritizing Research Questions

Our primary motivation as researchers and educators was an uneasiness with notions that classroom assessment practices could be "fixed" if teachers had appropriate training in the principles of measurement and assessment. We had some evidence that classroom teachers' predominant routines are not consistent with the highly structured and objective practices endorsed by the educational measurement community even when these behaviors are encouraged by school policies (Wilson, 1996). Added to this were our own theories about how the attendant conditions for assessment interact with teachers' expectations about growth and achievement to influence subsequent observations and judgments. It was during the task of transforming our concerns, speculations, and partial knowledge of the problem into specific questions that we first noticed how easily the research process could become captive to the paradigm debates: "Would this be a quantitative study or a qualitative one?"

We purposefully sidestepped this question. Instead we considered the information we would need for a more informed discussion of our problem, the implications and tradeoffs of conducting a study with preservice teachers, and the resources and constraints of our own research context. Staging the research in a faculty of education would extract much of the complexity that is normally found in classroom. Yet the same unalterable conditions that prevented us from launching a classroom-based research design also provided us with some effective experimental controls. A structured context for assessment and reporting would allow us to observe and analyze variability more directly as it occurred in our participants.

Confident that a carefully orchestrated simulation would inform an understanding of classroom assessment practices, we settled on the first two research questions: What is the nature of thinking and decision-making when novice teachers assume the assessment and reporting responsibilities for a grade 8 student? To what extent will our current hypotheses explain the behaviors of novice teachers who are confronted with a structured set of assessment and reporting tasks?

The breadth and depth of these questions spanned the boundaries of both traditional research paradigms (Fielding & Fielding, 1986). To restrict ourselves only to questions answerable through a single paradigm would be limiting to our investigation and our analysis. In selecting research questions we allowed ourselves to be guided by two beliefs: (a) that understanding how individual participants contextualized their assessment and reporting tasks would inform, and be informed by, a more distanced analysis of regularities in their collective behavior; and (b) that knowledge generated in this context could help educators at various levels to think about classroom assessment practices.

A later critique of our own behaviors revealed that by giving the logic and purposes of our work precedence over the tenets of orthodox methodology we had forged an alliance with the broad themes of the pragmatic tradition.[1] "Pragmatic choices about what to research and how to go about it are conditioned by where we want to go in the broadest sense. Values, aesthetics, politics, and social and normative preferences are integral to pragmatic research, its interpretation and its utilization" (Cherryholmes, 1992, p. 13).

The early decision to pursue questions that were both feasible and central to our more general research problem was not trivial. The promise of both quantitative and qualitative data meant that we could not anchor our behaviors in any single analytic framework. In exchange, however, we were free to consider new possibilities. Inherent in mixed-method designs is the opportunity to integrate questions about behavior and meanings (Hammersley, 1992), large-scale structural features, and microprocesses (Bryman, 1992). They also make it possible to build bridges between these often separately researched facets of social science. This latter assurance confirmed the appropriateness of our decision to add two dissimilar but relevant questions as the study proceeded. Why did virtually all the participants feel it necessary to offer unsolicited comments to Chris when marking the writing assignments? Is it possible to derive an empirical model that could represent how achievement was actually determined by this group of preservice teachers?

Quilting a Research Design

This pragmatic orientation continued into the overall design of the study. Specifically, the use of the portfolio exploited an increasingly popular assessment format. The incremental addition of information and some variation of Chris's work mirrored some of the rhythms and activities evident in many elementary and middle-school classrooms. It also allowed for the control of many of the factors known to influence marking: the type of work to be assessed; some of the characteristics of the student being evaluated; and the location, sequence, and timing of the assessment activities.

The design also featured some control of the assessment context. The notion of control traditionally implies a quantitative agenda. But the controls were strict only in the sense of the information they afforded participants about their Chris and Chris's classroom. We realized that a practice teaching experience during the simulation would encourage participants to compare their work in the Faculty with the task of assessing real students in real classrooms. These novice teachers were thus in an ideal position to analyze the contextual information offered by the design and to judge its adequacy and utility.[2]

Such sensitivity to context and setting is normally considered in the domain of the naturalistic or interpretive paradigm. Hammersley (1992) argues that the distinction between natural and artificial settings is spurious. He maintains that researchers are always trading off the efficiency of collecting of highly relevant data with the danger of influencing people's reactions to the collection of that data. "Much depends on whether the reactivity affects the results in ways that are relevant to the research topic and in ways that cannot be allowed for" (p. 44). In this study the fit between the experimental materials and the experiences the candidates were receiving in their placements may have limited the reactivity and indeed may have contributed to the results.

In our study descriptions of the simulated setting came from real schools, real classrooms, and a real teacher. Expectations were established using the words and voices of real children. Finally, all the tasks that Chris completed represented activities that could be a part of any grade 8 curriculum. Although the researchers fabricated Chris's responses to the tasks in accordance with the three experimental levels of growth, independent reading by graduate students in education, many of them teachers, confirmed these responses to be both distinctive and appropriate for a grade 8 student.

Producing the portfolio involved meticulous attention to the collection, verification, and integration of genuine contextual and assessment materials. Without this credibility it would be erroneous to assume that the individual assessment and reporting tasks would elicit attempts at a professional response from our novice teachers. It did not seem important during the design phase to be concerned whether resulting data would be counted or read. Our priority was to gather information that would help us understand the problem, both at the micro level (behavior attributable to individual characteristics) and at the macro level (action governed by some collective structure). This preference reflected our own theoretical position that teachers' assessment practices are best understood as a composite of these two worlds.

To suggest that the design process was only a conceptual exercise or that the research questions emerged independent of concerns about feasibility or their perceived value to educators and learners would misrepresent the process. In practice, the year previous to the implementation of this study was spent learning about preservice teachers' reactions to our proposed simulation and to the individual tasks that would be incorporated into the portfolio.

The pilot study involved 40 participants similar in professional experience and program choice to the group targeted for the full study. The portfolio was implemented using the same materials and sequences that we expected to use. In addition, however, three participants volunteered to meet with us weekly to talk aloud about the tasks they had performed, the quality of the contextual

data, and about the challenges of assessment in general. Results from the pilot study suggested that our hypotheses were worthy of further testing. They also helped to refine the simulation. Specifically, more context data were added to the portfolio, adjustments were made to the structure of the final exam, and parental involvement with all its supporting documentation was created as a new variable. It was also rewarding to hear the pilot group talk about the portfolio task as a valuable professional learning exercise.

Throughout the preparation of the research design, we adopted what we now call a non-foundationalist stance (Phillips, 1992). By not anchoring the study in any one research paradigm with its concomitant set of interlocking premises and procedures we committed ourselves to a continuous dialogue about the meaning and significance of the data we would collect. We also risked criticism from those who tend to judge the quality of research findings by the degree to which each step in the process is coherent with a set of assumptions about what is possible to know and how that knowledge must be garnered.

We have since found that non-foundationalist, mixed-method designs have a rich tradition.[3] A particular interest in this form of inquiry is evident in the field of program evaluation. Using Greene and Caracelli's (1997) framework, it is possible to describe our efforts as an "integrated" study with an "embedded or nested design" (pp. 23-24). That is,

> one methodology located within another, interlocking contrasting inquiry characteristics in a framework of creative tension.... The data resulting from dovetailing these methodologies convey both the meaning of naturally occurring behaviors in their social contexts and the frequencies representing macrolevel relationships. (Greene & Caracelli, 1997, p. 24)

Researchers who have spent significant energy examining the nature of inquiry and knowledge tend to agree that research methods are not intrinsically linked to any particular epistemological stance (House, 1994; Reichardt & Rallis, 1994; Robinson, 1998). Howe and Eisenhart (1990), in their attempt to refocus the qualitative-quantitative debates onto deliberations about the "logic of use" associated with various methodologies asserted, "Failing to follow a given theoretical perspective or methodological convention does not necessarily diminish the warrant of the conclusions drawn" (p. 6). For us the logic of use dictated that each method for data collection be integrated into the logic and rhythms of classrooms in general and the use of portfolios in particular. As Huberman (1987) predicted, we relied heavily on common sense, prior experience, and the logic inherent in the proposed simulation in crafting instruments that would serve both quantitative and qualitative data collection.

Multiple methods of data gathering, often referred to as triangulation, also strengthen possible conclusions. Mixed-method designs expand the traditional virtue of triangulation to include not only multiple measures, but also multiple investigators, multiple datasets, and multiple theories about the research problem itself (Brannen, 1992). The four researchers in this study represent differences in theoretical backgrounds, work in public education, age, and gender. These differences shaped the final set of research questions, our individual involvement with the study's implementation, and how the data were even-

tually treated. The collection of data also differed in type (quantitative and qualitative), in duration (over 10 weeks), and in context (both in class and at the participants' home work space). The personal theories we each brought to the analytical phase contributed to the current form of our collective learnings. Personal and pragmatic theories about the purposes of research also compelled us to test the methods and the findings of this study with both practitioner and research audiences.[4]

Managing and Analyzing Data

The design of this study resulted in the acquisition of multiple forms of data. Although each required a particular analytic approach, no analysis was conceptualized independent of the others. Fielding and Fielding (1986) argue that "an intimate back and forth testing, critiquing, and syntheses" of approaches stands "the best chance of specifying powerful solutions" to important problems (pp. 12-13). An example of such a dialogue in this study was our attempt to understand how Chris had been graded.

The finding that Chris's final grade was not well explained by marks on assignments and the exam was only a starting point in our analysis. Our mixed-method design allowed us to probe the meaning of this finding. One strategy involved the design of a structural equation model (Anderson, in press). The resultant good fit between the causal model and the data gathered about participants' behaviors supported the assumption that background, context, and perceived growth did indeed shape final grades. This analytic lens also allowed us to observe how growth and achievement competed for attention during the marking process and how some assessment instruments were more conducive to one consideration over the other.

A related analysis of variance of the different experimental groups provided evidence that participants constructed grades not only from observations of Chris provided by the design, but also from some expectations of performance. These expectations seemed to emerge from participants' understandings of Chris as this profile evolved over the 10 weeks of the simulation (Wilson & Martinussen, in press). Even when evidence was available to suggest Chris was not achieving on instruments designed to measure performance in language arts, this condition was not necessarily reflected in Chris's grades. No Chris received a failing grade. As well, teachers in some levels of Chris seemed intent on using grades as a way to reward a student they had decided should do well (Wilson & Martinussen, in press, Figure 1).

The reasons for this type of teacher behavior became more clearly understood through the analysis of the qualitative data (Shulha, in press). Both the comments placed on Chris's written work and the dendrogram that was constructed from participants' talk about their efforts demonstrated a view that assessment is an interactive process. It is expected that students will contribute to the quality of this interaction through formal responses to assessment tasks and through responses to instruction and teacher feedback. One role of the teacher in this interaction is to establish the conditions in which students can learn and demonstrate that learning. (According to the dendrogram there are at least 22 conditions under the influence of the classroom teacher.)

In this study there is evidence from all three distinct analyses that conditions beyond Chris's performance impinged on participants' judgments of achievement and therefore Chris's grade. Geertz (1979) would probably describe this way of coming to understand the grading phenomena as dialectical tacking between experience distant (nomothetic) and experience near (idiographic) data.

Assigning Value to the Study

Mixing methods in this way produced inferences that were certainly broad in scope. The design also gave us multiple ways to confirm these findings. Still the question remains: Did we select the best way to construct new understandings of assessment given the data? Phillips (1992) provides some guidance in this deliberation. His challenge to researchers is to adopt a *critical tradition;* that is, to subject the inquiry to intense scrutiny. Drawing on the work of Dewey (1957) and Popper (1976), Phillips proposes a set of inspection standards or "warrants" to judge the adequacy of knowledge derived from formal research.

According to Philips (1992), warranted knowledge is characterized by arguments based on evidence. This evidence is used not only to build a good case, but to set the stage for a good challenge. Arguments of evidence are apparent if (a) decisions have been carefully reasoned, (b) rules for classifying and counting the important data have been made transparent, (c) the logic and explanatory power for attaching significance to data have been accessible, (d) the value premises of the researchers have not been hidden, (e) there has been sensitivity to disconfirming evidence, and (f) early inferences have not been overly influential in the final analysis.

Similarly, Howe and Eisenhart (1990) argue that warrants need be applied only to questions of importance to education. For them warranted knowledge demonstrates: (a) a fit between the research questions, the data collection, and the analytic techniques; (b) the effective application of the data collection techniques; (c) an alertness to and coherence with background assumptions; (d) an "overall warrant" that includes being able to employ knowledge and evaluative arguments from outside the particular tradition of the research problem and being able to explain theories emerging from the data that were rejected; and (e) evidence of value (likely consequences) and worth (internal integrity).

It is not enough for the researchers to claim they have addressed the standards for warranted knowledge. Judgments using these criteria by an interested audience are still required. The role of the current discussion is to assist readers with this critical task. The invitation to critique is not issued as a professional courtesy. By claiming this work to be non-foundationalist research we deny the view that "grounded meaning and truth can be determined once and for all" (Cherryholmes, 1992, p. 15). For Phillips (1992) and others who hold a non-foundationalist stance, the criticism and debate arising when these standards are imposed on any piece of work will in the short term help differentiate between carefully conducted and shoddy research. In the long term it is proposed that warranted knowledge will contribute to the systematic demystifying of a problem, because such knowledge remains highly accessible and transparent enough continually to invite critical dialogue.

Consistent with this stance we would characterize our research processes and the subsequent findings as exploratory in nature. As such they should point to future efforts. Continuing to identify various student characteristics and context variables that influence teacher practices may lead to a fuller understanding of the implicit goals of classroom assessment. A first challenge, however, appears to be to help teachers to make the implicit explicit. Only when teachers are sanctioned to examine the complexity of their practices can they begin to monitor the accuracy and adequacy of the information they integrate into their judgments. Currently many formal policies mandate that teacher judgments about growth and achievement be traceable to formally designed instruments. The potential of newer types of assessments to capture the rich understandings that teachers and students develop about each other through their multidimensional classroom interactions remains unclear. Our data suggest that further investigation of three factors, *student attitude and effort, teacher disposition toward teaching and learning,* and the *nature of planning* would be important to understand further the interactions that take place under the label of assessment.

Another important feature of this study was the ambiguous findings generated about the influence of parental expectations on assessment practices. It is unclear whether the methods we used to introduce the needs and interests of parents into the simulation were inadequate or whether, despite talk to the contrary, teachers really do draw the boundaries for their practice only around themselves and their students. The increasing emphasis in educational policy on home and school partnerships, including parent councils, makes continued exploration of this interaction essential.

Yin (1994) points out that even those who represent competing research paradigms agree that quality inquiry should have significant implications beyond the immediate work. What we learned from the behaviors and reports of our participants has a direct bearing on teacher education. Our preservice teachers seemed hungry for strategies that would help them face the varying demands that the portfolio served up. Questions about how to involve students in their own assessment were as common as queries about how to develop the more traditional instruments used to assess Chris. When formal assessment instruments seemed appropriate for gathering information, our participants wanted these assigned tasks matched up with clear instructional goals and appropriate criteria for marking. For this group of novice teachers optimal classroom assessment had to include the skilled observation and tracking of individual learning over time. It is likely that our participants' enthusiasm for learning about assessment was rooted in the practical problems fashioned throughout the simulation. This lends support to the notion that preservice and inservice professional development is most effective when it is anchored in the real problems inherent in classroom contexts (Shulha & Wilson, 1997).

Conclusion

House (1994) warns against the use of dichotomies: "The reaction to the mistakes and excesses of positivism was interpretivism with its own excesses. Overemphasis on method led to definition by opposition: If one method was

quantitative, the other qualitative: if one was objective, the other was subjective" (p. 20). In practice this study demonstrates a conscious determination to link methodological decision-making more directly to the features of the research problem.

Along the fuzzy edges of the paradigms are problems requiring a type of systematic inquiry that is sensitive to the tensions that exist between individual action and more global summaries. Educational practice is filled with these kinds of problems (Robinson, 1998). Researchers with a disposition and a facility for mixed-method designs wish to explore the intimacy of individual experience while testing hypotheses about the tacit structures that shape experience. The warrants to which we have ultimately appealed in framing, implementing, and reporting our study are intended to foster at least temporary certainty and primacy for our knowledge claims about novice teacher assessment practices. The pragmatic and exploratory intentions of the study compel us to link our findings carefully to classroom teachers' assessment practices. How long these claims and links have utility will be a function of the amount of activity that will be aimed at critiquing and illuminating them further.

Notes

1. Cherryholmes (1992) acknowledges the current versions of pragmatism and provides a list of authors that could be referenced for a more in depth understanding of the nuances of the tradition.
2. Participants were not asked to talk directly about how assessing Chris compared with assessment as it occurs in the context of a regular classroom. Instead they were asked to analyze the strengths and limitations of portfolio assessment in the simulation. The resulting data highlighted the features of classroom contexts that participants tried to attend to during their task.
3. In reviewing the history of the paradigm wars, Datta (1994) presents a chronology of researchers whose work and writings promote the blending of qualitative and quantitative approaches. This list begins with the work of Campbell and Fiske (1959) and includes researchers such as Cook and Reichhardt (1979) and Yin (1989).
4. The methods and findings that shape this set of articles were first critiqued at the annual meeting of the Canadian Society for Studies in Education in June, 1996. Since then they have been reviewed by practitioners and students enrolled in graduate classes in measurement, evaluation, and research methods. Feedback from these audiences has led to the re-analysis of data and the rewriting of our work.

References

Anderson, J.O. (in press). Modeling the development of student assessment. *Alberta Journal of Educational Research.*

Brannen, J. (1992). Combining qualitative and quantitative approaches: an overview. In J Brannen (Ed.), *Mixing methods: Qualitative and quantitative research.* Brookfield, VT: Ashgate.

Bryman, A. (1992). Quantitative and qualitative research: Further reflections on their integration. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 57-78) Brookfield, VT: Ashgate.

Campbell, D.T., & Fiske, D.W. (1959). Convergent discrimination validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Cherryholmes, C.L. (1992). Notes on pragmatism and scientific realism. *Educational Researcher, 21*(6), 13-17.

Cook, T.D., & Reichardt, C.S. (Eds.). (1979). *Quantitative and qualitative methods in evaluation research.* Newbury Park, CA: Sage.

Datta, L. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. *New Directions in Program Evaluation, 61,* 53-70.

Dewey, J. (1957). *Reconstruction in philosophy.* Boston, MA: Beacon.

Fielding, N.G., & Fielding, J.L. (1986). *Linking data. Qualitative research methods, Series 4.* Thousand Oaks, CA: Sage.

Greene, J.C., & Caracelli, V.J. (1997). Crafting mixed-method evaluation designs. *New Directions in Program Evaluation, 74*, 5-18.

Geertz, C. (1979). From the Native's point of view: On the nature of anthropological understanding. In P. Rabinow & W. Sullivan (Eds.), *Interpretive social science* (pp. 225-242). Berkeley, CA: University of California Press.

Hammersley, M. (1992). Deconstructing the qualitative-quantitative divide. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 39-56). Brookfield, VT: Ashgate.

House, E.R. (1994). Integrating the qualitative and quantitative. *New Directions in Program Evaluation, 61*, 13-22.

Huberman, A.M. (1987). How well does educational research really travel? *Educational Researcher, 16*(1), 5-13.

Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research. *Educational Researcher, 19*(4), 2-9.

Phillips, D.C. (1992). *The social scientist's bestiary: A guide to fabled threats to, and defenses of, naturalistic social science*. New York: Pergamon Press.

Popper, K.R. (1976). The logic of the social sciences. In T.W. Adorno, H. Albert, R. Dahrendorf, J. Habermas, H. Pilot, & K.R. Popper (Eds.), *The positivist dispute in German sociology*. New York: Harper and Row.

Reichardt, C.S., & Rallis, S.F. (1994). The relationship between qualitative and quantitative research traditions. *New Directions in Program Evaluation, 61*, 5-11.

Robinson, V.M.J. (1998). Methodology and the research-practice gap. *Educational Researcher, 27*(1), 17-26.

Shulha, L.M. (in press). Understanding novice teachers' thinking about student assessment. *Alberta Journal of Educational Research.*

Shulha L.M., & Wilson, R.J. (1997, June). *A learning partnership in professional development: Building a better bridge*. Paper presented at the annual meeting of the Canadian Society for the Study of Education, St. John's.

Wilson, R.J. (1996, June). *Factors affecting the assessment of student achievement*. Paper presented at the annual meeting of the Canadian Society for Studies in Education, Brock University, St. Catharines.

Wilson, R.J., & Martinussen, R.L. (in press). Factors affecting the assessment of student achievement. *Alberta Journal of Educational Research.*

Wilson, R.J., & Martinussen, R.L. (in press). Factors affecting the assessment of student achievement. *Alberta Journal of Educational Research.*

Yin, R.K. (1989). *Case study research: Design and methods* (3rd ed.). Newbury Park, CA: Sage.

Yin, R.K. (1994). Evaluation: A singular craft. *New Directions in Program Evaluation, 61*, 71-84.