Whose Story Is This, Anyway? Navigating an Ethical Conundrum

Sharon F. Rallis
QUALITATIVE INQUIRY IN EVALUATION
From Theory to Practice

LESLIE GOODYEAR, JENNIFER JEWISS, JANET USINGER, AND ERIC BARELA

2014

Jossey-Bass
A Wiley Brand
CHAPTER

9

Whose Story Is This, Anyway?

Navigating an Ethical Conundrum

Sharon F. Rallis

Key Ideas

- Ethical dilemmas can arise when those involved in a qualitative evaluation have conflicting norms around practice. Evaluators must consider their own ethical practice as well as the ethical practice of stakeholders and others who are peripherally connected to the program.
- Evaluators often have to deal with surprises, particularly when stakeholders fall short of the evaluator's expectations or standards.
- Ethical challenges can arise when qualitative evaluators are navigating the following:
  - Creating the context for evaluation
  - Negotiating multiple evaluator roles
  - Engaging program participants
  - Clarifying project purposes, benefits, and risks
  - Confronting a lack of methodological expertise among stakeholders
  - Representing participant voice
Maintaining relationships with stakeholders and others peripherally connected to the program

On August 14 in the early afternoon, the director of one participating school in *Bridging the Gap: Innovation for Achievement* (pseudonym of a federally funded initiative to study and disseminate findings regarding successful innovative schools) sent this e-mail to the executive director of the sponsoring agency:

Subject: instr/concl chapters from [the book]

Norman, Please do what you need to do to remove Horizon School from the [research team’s] study. We have consensus that we want to pull out. Do what you have to do to ensure that our rights to do so are respected. Please reply as soon as possible as to my request. TC

The request in the e-mail represents an ethical dilemma that emerged from conflicting norms between two cultures in a quasi-ethnographic study of innovative schools that was one component of the larger initiative. The conflicting norms were the researchers’ culture and the cultures of the participant schools. Specifically, the researchers valued the norms of academia and scholarship (such as following stipulated protocols and publishing results), whereas the schools valued practical action to meet students’ needs. As in the case of sociologist Carolyn Ellis’s study of the Virginia Tidewater Guinea (Ellis, 1986), the two cultures differed on what defines the *truth* and on the ethics underlying the discovery of any truth. Ellis’s judgments that the fisher culture was deficit (defined as a lack or impairment in a functional capacity) were based on her western white worldview; that is, she saw a culture lacking social mechanisms she considered necessary to function and prosper in the modern world. Similarly, the judgments made by the university researchers regarding aspects of the schools in the study were based on their own academic values, not the values that drove schools’ operations. And, like Ellis’s fisher folk community, the schools had opened doors and taken the research team into their lives with the expectation that their voices would be heard and respected. However, unlike Ellis, the members of the research team were novice ethnographers who did not practice the *caring reflexivity* (see Guillemin & Gillam, 2004; Railis & Rossman, 2010) necessary for a qualitative study to be ethical and thus trustworthy.
As is often the case in federally funded education initiatives that include research as well as dissemination, evaluation is required. As the project evaluators, we found ourselves entangled in ambiguous role delineations while also addressing conflict and inexperience. At times we saw ourselves playing the guide for ethical decision-making. This chapter tells the story of the project through the eyes of the evaluator, using a moral lens and organized by the challenges we faced. We follow the development of our role as evaluators and explore the implications of our actions within multiple contexts. That both the methods used by the research team and the primary tools used by the evaluators were qualitative holds potential for dual insights into what contributes to ethical qualitative evaluation practice.

THE CONTEXT: WHO'S DOING WHAT?

Bridging the Gap (BTG) was a replication and dissemination grant project funded by the US Department of Education (DOE) through a nonprofit State Association of Innovative Schools (henceforth to be called the association). Specifically, the association received the funding to carry out the project, which brought together a variety of players. BTG began with three questions:

- What are the common elements of success (as measured by standardized tests and graduation rates) in high-quality innovative schools serving students at educational risk in high-need communities?
- How do we share this information with a broad audience of educators and policy makers?
- How can states disseminate and replicate these common elements to create more highly successful schools?

To find answers, BTG put forth goals to: identify high-quality innovative schools in the state that serve students at risk of educational failure in high-need communities; research and document the common elements of success in these schools; disseminate findings nationally through a variety of media; and conduct trainings to replicate best practices. To meet these goals, a mix of people interacted: the association director and project...
coordinator, innovative school leaders and staff, university researchers, a filmmaker, personnel from the professional development organization that would support dissemination activities, and two evaluators. Each stakeholder group brought to the project its own agenda, particular interests, and individual perspectives.

The first task was to identify and select five innovative public schools in which student performance on a variety of academic scales (i.e., state high-stakes standardized tests, graduation rates, college acceptance) outperformed the demographics of each school's surrounding district. Each of the selected schools wrote whole school papers explicating the philosophy and practices that they believed contributed to the high achievement of their students. The association contracted a university researcher specifically designated by the funder to research the schools and document what these schools did to produce such positive results. The research team (one faculty member and five graduate students) was to collect data and produce a book for dissemination. The project also contracted an independent filmmaker to produce a film illustrating what he saw as elements contributing to the high performance of the students; the film was to be suitable for dissemination through public broadcasting and other public venues. An agency with expertise in providing professional development was contracted to design and lead study tours of the schools in the final year of the project as a means of demonstrating best practices. Finally, the association chose my colleague Andy and me, of the Education Evaluation and Policy Research Group, to evaluate the BTG project.

Regarding evaluation of the project, the association wrote in the proposal: "We believe that staffing for [the evaluation] tasks requires a combination of sophisticated evaluation design skills (benchmarks, protocols, etc.), high-level personal interaction skills with senior personnel (school leader interviews, review panel recruitment and facilitation, and research and project team), and regular, ongoing availability to manage logistics, attend meetings, schedule interviews, collect and write up data, provide editorial support, synthesize findings, etc." Indeed, in our roles as evaluators, we drew on all those skills and more. BTG's goal was to produce knowledge for dissemination and replication, so throughout we found ourselves challenged
ethically and logistically with these enduring questions: Whose knowledge is captured? How is it captured and represented? Does the story that unfolds realistically relate participants' experiences?

THE MORAL LENS ON QUALITATIVE INQUIRY: PRINCIPLES AND ETHICAL THEORIES

As a program evaluator who relies on qualitative methods to inform questions about the program, I recognize that I am the means through which the data are collected, analyzed, and interpreted. Essentially, I am constructing new or confirming extant knowledge; as such, I continually and consciously make decisions about conceptualizing, designing, conducting, interpreting, and writing up findings. To ensure that my decision-making is both competent and ethical, I establish my moral position—that all evaluation should have a goal of improving some social circumstance (Rossman & Rallis, 2003), specifically to “improve policy and programming for the well-being of all” (Weiss, 1998, p. ix)—and I draw on moral principles that direct me to act as I would want everyone else to act in any given situation (Kant, 1788/1956): justice (fairness and equity), beneficence (risks and benefits, reciprocity), and respect for persons (participants are seen as autonomous agents) (see Hemmings, 2006).

Nearly all aspects of qualitative inquiry are governed by ethics, which are standards of conduct based on moral principles. “The point of moral principles is to regulate interactions among human beings” (Strike, Haller, & Soltis, 1998, p. 41). Thus, my ethical keystone is relationships, not procedures. A social activity, evaluation is naturally interactive and dialogic; ethical evaluators seek out and listen carefully to the voices embedded in the social context (Taylor, 1994). “Good qualitative [evaluators] are adventurous in their pursuit of thick descriptive data on how [actors] make sense of their worlds: . . . and do their best to forge rapport with participants in order to capture their ‘emic’ viewpoints on various social phenomena” (Hemmings, 2006, p. 12). We engage in the reflexive reasoning that is central to any trustworthy evaluation. Reflexive reasoning is relational; it is multifaceted and iterative, moving back and forth between
the details and the whole, between the evaluator's questions, the participants' knowledge, and the evaluator's interpretation of that knowledge. We ask: What do I see? How and why do I see this? How might others see it? What does it mean to me? To others? Put simply, reflexivity is "looking at yourself making sense of how someone else makes sense of her world" (Rossman & Rallis, 2003, p. 49).

Reflexivity recognizes that the evaluator and the participants are involved in continual and changing interaction. The evaluator asks: What is happening in this relationship? Is it fair? What might be possible consequences of my actions? Of my interpretations? What rights of participants might I violate or potential harms might result? Are the participants likely to benefit in any way—and are they likely to agree on the benefit? What burden might my inquiry lay upon participants? Equally important is that we question our own actions: Do I act according to my principles?

Building an ethical relationship with participants may not be easy or happen immediately, but I do believe that it becomes easier when we articulate (at least explicitly to ourselves) the moral principles we intend to use. Not all moral principles are compatible, and contradictions even lie within some. Moral principles form ethical theories that can be grouped into two broad categories: consequentialist and non-consequentialist. Consequentialist ethical theories focus on the results of actions in determining their rightness or wrongness. Any particular action is neither intrinsically good nor bad; rather, it is good or bad because of its results in a particular context—its consequences. From a consequentialist perspective, the evaluator considers the results over the process; means are less important if the end has value; for example, utilitarianism advocates that one's behavior should result in the greatest good for the greatest number. Using this moral principle, an evaluator's work would be more summative; she might document the impact of the program on a majority of participants.

In the other major category, non-consequentialist ethical theories derive from the moral principle that universal standards exist to guide human behavior and interaction. Two non-consequentialist theories on which we as qualitative evaluators claim to operate are the ethic of individual rights and responsibilities and the ethic of justice (see, for example, House
& Howe, 1999). The first upholds the unconditional worth of the individual and judges actions by the degree to which they respect a person, not by its outcomes or consequences. The protection of these rights may not be denied, even for the greatest good for the greatest number. Ethics of justice go beyond individual rights and responsibilities to espouse the redistribution of resources and opportunities to achieve equity above equality. The goal is to ensure that everyone is better off, even though the allocation of some benefits may differ; rightness or wrongness of actions is judged through principles of fairness and equity. The moral principles of non-consequentialist ethics argue that people must be treated as an end in themselves, not as a means to an end. According to Rawls (1971), the benefit or welfare of the least advantaged, not that of the majority or average, must drive any action. His view maintains that improving the welfare of the least advantaged ultimately benefits everyone because the communal resources of society and future generations will grow.

Principles of justice and individual rights consider questions of power and representation: Who defines what is right in a given situation? Whose values are used in rating on criteria? Are all voices given opportunity to be heard? Other perspectives situate morality within the context and relationship. Communitarianism (MacIntyre, 1981) acknowledges that communities differ on what is morally good or right. Evaluators often find that not all stakeholders share fundamental values, and that those values may conflict with the evaluator's own values. So whose values define the research, guide decisions, and shape interpretations? A postmodern response would posit that power and dominant versions of "truth" shape the relationships and thus the research (Foucault, 1970).

Knowing ethical theories and the principles on which they rest can help evaluators analyze and understand their choices for action, but theory is abstract and therefore may not be practically helpful. The AEA Guiding Principles for Evaluators codifies ethical theories into a collection of principles to guide evaluation practice; for example, evaluators will: "conduct systematic, data-based inquiries"; "provide competent performance to stakeholders"; "display honesty and integrity"; "respect the security, dignity, and self-worth" of all involved; and "take into account the diversity of general and public interests and values"
(www.eval.org). Not standards or directives, each principle must be explicated and applied to the relevant situation in practice. From a moral perspective, the danger in applying any set of externally generated theories to real-life settings, that is, translating abstract principles into practical actions, is the establishment and imposition of one-size-fits-all rules. Thus, evaluators are still challenged not to simply follow procedures but to consider what any action may mean for the relationships present in the evaluation setting.

The ethic of care is an alternative perspective that offers a potentially powerful and practical way to operationalize the moral and ethical aspects of the evaluation process. Caring argues that the core of morality must be the relationship itself. “Ethical decisions must be made in caring interactions with those affected by the discussion. Indeed, it is exactly in the most difficult situations that principles fail us. Thus, instead of turning to a principle for guidance, a carer returns to the cared-for” (Noddings, 1995, p. 187). This approach emphasizes concrete circumstances over abstract principles: What does this person need in this moment? Care theory emphasizes the moral interdependence of people: “Our goodness and our growth are inextricably bound to that of others we encounter. As researchers, we are as dependent on our [participants] as they are on us” (Noddings, 1995, p. 196).

From these ethical theories and their corollary moral principles emerge two practical tensions that offer a framework for analyzing our ethical and logistical challenges: harm versus care and burden versus benefit. As evaluators of BTG, our overall goal was to contribute to improving aspects of the initiative so that the initiative could, in turn, contribute to the improvement of the social circumstance: education. In achieving this goal, we sought to avoid doing harm, both in our own work and in the work of the various other players; to encourage care and respect for all participants; and to minimize burden while promoting benefit. We knew that in qualitative evaluation potentials for harm “are often quite subtle and stem from the nature of the interaction between researcher and participant” (Guillemin & Gillam, 2004, p. 272). Thus we strove for conscious reflexivity, which included reflection both on and during action, as antidote to harm. Our reflexivity, driven by our ethical stance, drew both
from principles and from reciprocal and caring relationships unique to the contexts. We believe we practiced caring reflexivity to conduct ethical evaluation and to monitor for ethical research. To summarize, in all decisions and interactions with participant individuals and groups, we considered principles of justice, beneficence, and respect for persons.

**CHALLENGES: ETHICAL AND LOGISTIC**

**Challenge 1: Creating the Context for Evaluation and Negotiating the Evaluator's Role**

As an evaluator, especially one who uses primarily qualitative tools, I am used to negotiating my role. This project was no exception. As noted earlier, federal research grants often require an independent evaluation of the research, thereby relegating the evaluator to a role of compliance monitoring (e.g., did the project meet benchmarks/do what they promised?). As an evaluator, I prefer to inhabit a role that works with the program people to consider the merit or worth of the program or of the study being conducted about the program. Defining the scope and activities of the evaluation emerged as a challenge during early conversations among ourselves in the Evaluation Group and between us and the association. At the outset, we wondered why we were not conducting the research on (i.e., evaluating) what made these schools effective. While evaluators might seem the natural choice to study the schools with the aim of identifying elements that contributed to school success, for later dissemination through book, film, and tours, the federal request for proposal (RFP) separated the research role from the evaluation role, and the grant officer specified that the researchers were to be recruited from a specific institution. Thus, the association had already contracted “researchers” to complete that task as well as to produce the book.

So what was our role to be? The grant defined evaluation as monitoring to ensure that tasks were completed as described in the proposal, but we demurred because we could not see ourselves as police. Thus, we re-negotiated the purpose and conditions of the evaluation. First, we established our approach to evaluation: interactive and participatory, grounded in an
ethic of care (Noddings, 1984). Although we recognize that outcomes are important, we choose to focus on what happens to reach those outcomes (a more non-consequentialist stance). We proposed that we do more than monitor benchmark achievement and instead suggested the role of evaluator as *critical friend* (see Rallis & Rossman, 2000, 2001), that is, observing, interacting across the players, providing feedback, and asking multilayered questions about intents and impacts. This enactment of our role, while drawing on mixed-methods tools, places strong emphasis on the more dialogic reporting strategies common in qualitative or participatory studies.

While still stipulating the monitoring role, the association warmed to our proposed expanded role, so the contract between the group and the association defined the evaluation scope of work as the following:

- Work with program staff and research team to clarify research goals and project benchmarks
- Develop protocols for collecting qualitative and quantitative data on process, product quality, and project outcomes
- Document research and project progress according to benchmarks
- Independently review data used in the selection process (including US Census, state education department files, and school records)
- Interview school leaders to ensure that the research accurately represents the schools profiled
- Assemble, convene, and synthesize feedback from a review panel to evaluate quality of products
- Monitor and document progress on book, video, website, ads, institutes, and other documentation, dissemination, and replication products
- Provide editorial feedback on research, documentation, and dissemination products
- Provide data and feedback to program staff on a regular basis, for reporting and program improvement purposes
Produce a quantitative and qualitative summative evaluation report

Thus began a three-year relationship with the many players involved with BTG. Initially our efforts were directed toward building a trustworthy relationship with the association director (Norman) and project coordinator (Alice), based on the mutual goal of ensuring that the project be the best it could possibly be. Our part of the bargain was to provide evidence-based feedback, facilitate interpretations, summarize findings, and raise questions. One of the first e-mails between the project coordinator, Alice, and me illustrates the tone of our interactions. Note that we clarified how the knowledge produced is theirs, not ours:

Alice and Andy,

Great—albeit brief—talk just now. To summarize, the point is that we need to modify the task/activity-timeline chart to add two columns: completion and quality. We can then check off and add footnotes to explain or inform any deviations (to quote you, Alice!) in the completion column. For quality we agree that to check off we need to be sure we have discussed and agreed—and reference any formal assessments leading to judgments of quality. Note that we, as evaluators, are not the determinants of quality, but serve to help you make quality judgments (i.e., are you pleased and why or why not?)

The three of us may choose to meet to review—or we may agree that face-to-face meeting is not necessary.

Thanks, Alice, for your persistence and patience.

Sharon

At each step along the way, we had a set of questions that guided our behaviors. The questions we pondered are boxed near each of our challenges.

**Challenge 2: Engaging Participants and Clarifying Project Purposes, Benefits, and Risks**

Our second challenge arose in ensuring that the “right” innovative schools were chosen and brought on board. This task was tricky because the schools did not initiate the project and initially saw project activities as potentially more intrusive than beneficial. However, were project involvement to become seen as
Box 9.1. Questions We Used to Guide Our Qualitative Inquiry and Actions (Challenge 1)

- What moral principles do we expect will guide our practice?
- What approach and tools will best serve this evaluation? How do we describe these to stakeholders?
- Can all parties agree on our role and expectations for the evaluation? Can all reach consensus on expectations for activities and events?
- What is the nature of the relationships we are building (e.g., critical friend)? How will we interact with the players?

advantageous or selection as an honor, competition to be chosen would result. Thus, the association realized that both school selection and entry/access would be complicated and difficult. Andy and I helped Alice, the project coordinator, specify and promulgate selection criteria and then verified that the selection process was transparent and matched the criteria. For this task, we used primarily qualitative methods: interviews and reviewing descriptive documents about the schools, both to select and to verify.

Once the five schools had been selected and agreed to join, questions of power, authority, and representation surfaced (questions qualitative researchers learn to notice). Andy and I brought evidence of the schools’ interest and their concerns to Alice’s attention. The schools expressed reasons why they were willing to participate: recognition; promised resources; embedded professional development; opportunity to “tell our story.” They also intimated the kinds of knowledge they hoped the research would produce: their strengths; their struggles (e.g., long hours, burn-out); their successes. The first general meeting was held at the university where the researchers were located; the schools appeared impressed to be part of the “research agenda” of a professor at a prestigious institution. They were excited about being filmed and about the concept that they would serve
as "laboratories for innovative developers and district schools identified for improvement, corrective action, or restructuring." They expressed concerns about the time that might be required of them and of their staff and students. Although BTG made clear that involvement would entail additional time, promises were made that the schools' needs and integrity would be respected.

Gradually, what became clear to us was that not all parties shared common perspectives of participation and product. In hindsight, we now realize that as evaluators we should have asked that both sides make more explicit and formally agree on their expectations and promises. The ethical challenges were those of consequence (would the intrusion be worth the result?), just-ness (would rights and expectations of individuals in the school be honored?), and care (what relationships would develop and what would they mean to the various parties?).

The first activity to engage the schools, the writing of the Whole School Papers by a team of school members, both set the tone for future engagement and provided an integral way for us to connect with people at the schools. These papers were to be based on a school-designed self-analysis, and the evaluators were supposed to ensure that this self-analysis occurred in each school, that it involved faculty and staff as well as administration, that it was conducted somewhat systematically, and, finally, that the results of the analysis were incorporated in the papers. This role provided us an opportunity to talk with the school writing team about their goals for the paper and how they could use their learning from the self-analysis to meet those goals. We also facilitated conversations between the authors and other staff members to ensure agreement on how the authors represented the school in the papers.

Our next activity was to review and comment on the professional development experts' initial plans for the study tours. The study tours (a dissemination activity) were meant to offer other schools opportunities to visit the participant schools and learn from witnessing the research findings (best practices) in action. The plan revealed that the professional development experts drew on their expertise and incorporated practical input from the participating schools and the association leadership. In an e-mail to Alice, we commented on the evidence that the experts were developing their plan in conjunction with others in the schools and thus
would likely provide a strong foundation for the “replication” process. For example, we highlighted these excerpts from the draft:

- The principals noted that involving (district school) teachers throughout the dissemination process would ensure their investment in the project.
- In addition to ensuring that district school teachers are involved in all steps of the dissemination process, it might be useful to include them in meetings/outreach when initially presenting the project to prospective schools.
- Find out not only what specific district schools want to learn about/change, but what state and district mandates have been given to them.
- Approaching district leadership first will ensure that district leaders are aware of the project and its potential benefits.

Now, having come to know the school people and having seen their early interactions with other players, we felt that access to the five schools was negotiated; indeed, the host schools opened their doors to the research team, the professional development experts, and the filmmaker. As evaluators we contributed to reaching this point through drawing on multiple roles that included convening conversations and facilitating dialogue; lending our expertise to the whole school paper teams regarding case study development; reviewing and commenting on documents; and informal content analyses of work to date.

**Challenge 3: Cooperating with the Research Team: Issues of Expertise, Experience, and Representation**

The most extensive component of the BTG project was the work conducted by the research team to “identify common elements in the five schools that could be described for the benefit of others wishing to learn from our experience” (from the proposal). Each of the five graduate students would take the lead for one school: they chose an ethnographic approach, spending time on-site to observe actions and interactions; interviewing teachers to capture what these interactions meant to participants; and discovering
patterns. The team’s findings would serve as a basis for all dissemination and replication activities, so Andy and I felt a strong sense of responsibility for ensuring that the research process was both rigorous and ethical.

The proposal described our charge under Ongoing Evaluation of Outcomes, Goal 2 (Documentation and Research):

1. Evaluate research findings.

2. Evaluator will develop a review panel consisting of innovative school authorizers, innovative school and district school leaders, legislative education staff, and higher education faculty. Researcher will present findings to review panel through oral presentation, as well as print materials (summary of conclusions and spreadsheet). Review panel will provide verbal input and written feedback; specifically, panel members will evaluate the research.

3. Evaluator will summarize verbal feedback, compile written evaluation responses, and share these with researcher and association.
4. Researcher will use input to review and refine research, determine next steps—more research, a change in approach, etc.

We became aware of potential challenges to evaluating the research activity—perhaps playing a broader role than we had planned—when Howell, the professor leading the team of five graduate students, took me aside to say, “I’m really glad you are the evaluator because I’m not really a researcher—I don’t know a lot about doing this.” Shocked, I was unsure of how to deal with the ethical dilemma I felt I faced: the knowledge of the research team’s lack of expertise and experience that might cause harm conflicted with promises made by the project to protect the schools. A non-consequentialist moral principle would ask: What is fair and just? A caring perspective would ask: What do the participants need in this relationship? Three of the American Evaluation Association (AEA) Guiding Principles for Evaluators mandate that I “adhere to the highest technical standards,” ensure that I “provide competent performance to stakeholders,” and that I “display honesty and integrity in my behavior” (www.eval.org).

Howell suggested that, given my knowledge of qualitative methods, I could provide his team with an overview. Because I teach, have written about, and have offered workshops on qualitative research, I agreed to meet with the team to do a crash course, but I wondered whether it would be enough. Practically, I did not think such quick coverage of qualitative methods would suffice, but I felt I had no choice but to work with them because their contract was established in accordance with the specifications of the granting agency. My decision was driven by a hope to meet the needs of the graduate student research team (ethic of care) and by our charge to evaluate the research process and product quality (consequentialist moral principles). Still, I wondered, what implications might arise from my adding methods instructor to my role?

Evaluators are often “asked” (in various ways) to take on unanticipated and additional roles. Sometimes these roles either complement or replace a designated role. Other times, they become a burden, demanding more than resources allow. In the worst cases, they may contradict what was promised. The request to help the research team get off to an informed start seemed
doable and even complementary to our role as critical friend. I even felt that Howell's remark underestimated his knowledge and skills in research. In retrospect, I wonder whether providing an inadequate crash course contributed in some way to later research team actions that may have been interpreted as harmful or burdensome to the school personnel.

We set a date for me to review the basics of qualitative ethnographic research with the research team. We discussed the need for a solid conceptual framework to guide data collection and analyses. We considered what rigor and ethics might mean in this study. We talked about establishing a trustworthy and caring relationship with the school people, especially because they were generous to open their doors and give their time to the researchers. One of the students had experience doing qualitative research; another had previously worked in one of the schools. They all expressed confidence in their ability to conduct the research. Off they went to collect their data; however, I remained concerned that they were unaware that qualitative research often seems simpler in theory than it is in practice. I was left to monitor their research practice.

We turned our attention to assembling and facilitating an expert review panel, which I hoped would offer additional eyes in evaluating the research process and findings. The panel was convened in January of the second year, after the preliminary research report was completed. The review panel comprised local and regional practitioners termed "experts": three leaders of innovative schools not in the study, two public school principals, three education faculty members, and the innovative school director in the state's department of education. No national experts were able to attend (one sent his comments to the preliminary report). As requested, the research team presented the themes that emerged from their preliminary analyses: organizational practices (what); organizational assumptions (why); human side of the organization (who). The expert reviewers reacted, focusing on how these themes would be represented in the book. Their recommendations included the following:

- Clarify that members of the audience are practitioners and write for them
- Present the book as a best-practices "menu" for practitioners
Describe specific practices with sufficient detail so the reader can “see” the practice.

Be careful not to oversell the differences between the innovative and district situations (“best practices are still best practices”).

The exchange between the expert panel and research team revealed significant tensions grounded in cultural differences between the two groups. One panelist said: “I want to know what exactly CIT Prep is doing around their Saturday School model. Connect the conceptual and operational.” Howell responded, “But that is not doing research.” He reminded the panel that while they were experts on practice, he and his team were the experts on research. A few panelists noted that, because the team was from a university, Howell would have the last word on what constituted research. In my notes on the meeting, I wondered about Howell’s rapid acquisition of expertise, given his confession several months earlier that he was not a researcher and knew little about research.

During the presentation and discussion, other panelists expressed additional concerns. Several questioned how any conclusions could be drawn after “only twenty-two days spent across all [five] schools?” Others wondered why students’ input was minimal; students had not been surveyed, and, according to one of the researchers, “We did each shadow a student but have not done any formal interviews with students.” One panelist said, “I need a chapter on instruction practices—curriculum, instruction, and assessment. Then I can think about how to support teachers in classrooms.” Howell replied: “We have a lot on assessment, alignment, not on instructional practice.” Another panelist responded, “Don’t try to talk about what you don’t know. If you don’t have it, you don’t have it.” Howell noted that a focus on instruction was “beyond their charge.”

Andy and I summarized the panel discussion, noting themes, questions, recommendations, and concerns. We also reiterated the purpose of the book. We forwarded this summary, along with a transcript of the panel discussion, to all who had attended—researchers and panelists—asking for a check on our interpretations of what had transpired. Our goal was transparency and agreement on representation. All agreed that the summary
accurately represented all viewpoints expressed. Revealing his acceptance of my summary, Howell e-mailed me:

Hi there, Thanks so much for sending along the summary and that set of notes from the review. We are going over them now and hope on Monday to get fairly well squared away on the format for the book. Thanks for all of your suggestions and help. When I know more I'll be in touch.

Drawing on the panel discussion and the results of a survey we conducted to collect reactions from the school personnel to the research team’s preliminary research report, I met with Howell to review his plans for the book—a case study of each school with a cross-case analysis chapter. We also discussed possible publishers and dissemination. I encouraged him to choose a publisher who would not determine what was important to emphasize; as the evaluator, my effort was to ensure that the book represented actual events and activities occurring in the schools, not what a publisher thought would be marketable. Shortly thereafter, we (Alice, Andy, and I) received Howell's revised outline and a possible publisher. I e-mailed Alice:

My thinking in response to the direction Howell seems to be taking with the book is: Whose book is this? The review panel clearly called for a practitioner-oriented book. Howell appears to be heading in a somewhat different direction (for example, it raises issues that are not illuminated by the data collected or necessarily integral to the original conceptual framework for the book). You may find [you need to confront Howell]... because you hold the grant and the responsibility to see that your purposes are met. I'm hoping that in our critical friend evaluator role, we can help facilitate a product that will meet the needs of both the association and Howell. You have a lot of good material in Howell's work—now the challenge is to get it out in a form that will be useful for practitioners. I look forward to the meeting with all of us (Andy, you, me, and Howell).

Sharon

Alice e-mailed Howell:

This project has a central focus on producing practical products. At the vetting, we consistently heard “let the reader see a practice,” “how is this operationalized?” etc. Your reaction at one point (“that isn’t research”) worried me, as did the shift you mentioned regarding publishers being considered, and to a lesser degree your
response to photographs. I think the feedback at the vetting was fairly consistent, and now I need to follow through—how is that feedback being incorporated into the next steps, into the production of the book?

Alice

Howell responded to these concerns by assuring us that “We will produce a book that you are proud of.”

We waited in anticipation for the book draft, which, according to the original agreement, was to be reviewed by Norman, Alice, Andy, and myself. The original agreement also stipulated that the association, the evaluators, and the researcher were to collaborate on book goals, editing, and design. In March, before the drafts could be vetted, the graduate students uploaded papers that were versions of their draft case studies on a website for an annual conference of their professional association. One of the schools discovered that the papers had “gone public” and was very upset. Alice asked Howell to have the papers removed, which was done.

The damage, however, had begun. School people had read the material on the web and were troubled by what they read. During the next month, full drafts of the case studies were sent out; because of earlier concerns, Norman forwarded each on to its school. The director of Connection e-mailed her reaction:

The proposal to DOE describes the process of selecting schools in urban areas who are successful in standardized testing, that success being the basis of selection for the project. We were told during the whole paper/school survey phase to look at the challenges faced by our students and at methods we employed to meet those challenges in order to close the achievement gap. Other schools, we were told, would do the same and researchers would be selected to “identify common elements in the five schools that could be described for the benefit of others wishing to learn from our experience.” Based on this understanding, then, we instructed our staff and faculty to speak with researchers about how we addressed the goal of academic success for each student.

We asked them to focus on how we diagnosed learning needs and then developed methods to address those needs in an effort to close the gap. This is what was expected by the association and the DOE.

So what happened? [In the case study], we are described as focusing on children’s deficits and somehow not respecting their
cultural heritages. Our entire community, especially our parents, is most concerned about the conclusions drawn by the researchers. We do not agree with the opinions related to our respect for the whole child. I hope you understand our concerns. TC

Alice’s e-mail to Andy and me shortly thereafter seems to recognize the relational issues and what it means to produce findings:

If you read the EQUAL paper, you will also find some very poorly written material, and there are some worrisome places in the Horizon paper too (but definitely fewer and less worrisome; the director is replying to those). I’m glad the other two papers are in very good shape—COL and CIT Prep—though after all this with Connection, I think I’ll re-review all chapters. As David at CIT Prep put it, they and their researcher were a good match; that probably goes a long way toward making this chapter business work.

What followed was a spring and summer of often vitriolic exchanges regarding the tone and content of the case studies and the cross-case analyses that came after. As independent evaluators who had been seen as critical friends to all parties, not members of either group, we became mediators and negotiators, facilitating a co-construction of the work.

**Challenge 4: Finding Balance between Perspectives: Participants’ Voice or Researchers’ Truth?**

Two cultures at odds—public school versus academic research—had faced off. The question centered on which culture would prevail in the book: What message would be sent about the schools? How would they be represented in the images and words presented in the text? As discussed earlier, Ellis’s (1986) judgments of the *Fisher Folk* were based on her worldview, which represented their culture as deficient, that is, lacking social mechanisms she saw as requisite to prospering according to western modern criteria. In the case of BTG, the judgments of the university researchers critiqued aspects of the innovative schools based on their own academic values (e.g., publishing, being recognized as experts in their field, claims of truth), not the values that drove schools’ operation. What fed the conflict was the perception that the researchers’ judgments were unbalanced and were over-interpretations of the data.
BOX 9.3. Questions We Used to Guide Our Qualitative Inquiry and Actions (Challenge 3)

- How is our role developing? Do our actions support justice, beneficence, and respect?
- Do we provide evidence with our feedback? Do we summarize findings, raise relevant questions, and facilitate interpretations?
- What expertise is needed? Did we appropriately facilitate the provision of expertise via the panel? What else might be needed?
- Have we established and promulgated criteria we will use to judge merit and worth of project activities and products? Do all parties understand and agree on these criteria?
- What evaluation tools are proving useful?
- Are our activity reviews/summaries revealing evaluator perceptions of activities, events, and products? What evidence do we have that these summaries are useful? What purposes are they serving?
- Whose voices are represented in project products? What voices are absent? Should more voices be included?
- What conflicting perspectives and moral stances may be in play in the project activities? What impact do conflicts have on justice, beneficence, and respect?

One side, the researchers, laid a claim for scientific rigor that can discover the “truth” of a situation. The research team posited that their perspective was “objective” and thus provided the true, accurate representation. (As a qualitative evaluator, I note that qualitative ethnographers do not claim to be objective, nor do they seek a single truth—we analytically describe and interpret events and interactions and make explicit our subjective perspectives.) On the other side were the schools and the association, which
claimed the researchers had exceeded the scope of the project as defined in both the proposal and the contract—and as had been explained to the schools when they gave their consent. In the United States, any institution that receives federal funds must have a review board to protect human subjects from harm or other violations of their individual rights. These institutional review boards (IRBs) require that research participants give their voluntary and informed consent. This consent must document that the participant (“subject”) fully understands the nature, purpose, and potential uses of the research; privacy and confidentiality are assured; participation will not adversely affect the rights and welfare of the individual; and participants may withdraw at any point without any negative consequence to themselves. (This agreement was what the opening e-mail referenced in asking to pull out.) Given the explanation by the association of the project purposes and goals, the schools expected that their voices would be heard, that the story told would be one they recognized as theirs, not one they felt was imposed on them.

The actual conflict was more complicated than the school and association personnel believing that the researchers had over-represented what had happened at the schools. They also questioned how it could have been possible for the researchers to represent the voice of the school with so little contact, thus challenging the research methods and rigor:

These assumptions [about our schools] are presented as based on many weeks of research in these schools. We don’t agree with that. We know they were not in schools that much. We need to have a clear statement of how many classes were seen, how many hours, and over what period of time. The researchers need to make it clear that they just saw a snapshot. (LT at COL)

The researchers’ representation of one school’s use of English as the dominant language illustrates this conflict of perspectives. The following paragraph concluded a section of one case study:

It is important to note that these administrators’ framing of their students’ backgrounds causes them to focus on deficits, rather than assets. While such an intense focus on student weakness may be, to some degree, inevitable in the standards and accountability era—and while this orientation does allow for instructional
goal-setting, concentrating on a child’s perceived deficits could inhibit a full understanding of a whole child—an understanding that honors the many facets of a child’s identity and worth, and not just those skills that are easily measured and valued by the current testing educational system. Indeed, Connection navigates what can appear to be a cultural trade-off for some students: Standard English skills for the language spoken at home or within the home country. While the school does have cultural nights and students do receive some Spanish lessons, English instruction is, without a doubt, the focus of the staff’s instructional energies, especially for students within the lower grades. Such a trade-off raises questions, questions confronting any school drawing upon a diverse student body: What are the costs of replacing the language of home with the language of school? Does the notion of “compensatory education” suggest that students must give up something worthy and valuable about their own backgrounds? (Draft of Connection chapter)

Teachers and leaders at Connection were deeply offended. They protested that the chapter did not present the whole picture. They were complying with the state’s English-first policy; indeed, parents had chosen the school because of its emphasis on English. Furthermore, they argued that topics such as “costs of replacing the language of home with the language of school” were not even in the purview of the research project. In an e-mail exchange, they remarked that this section read like a “typical academic treatise.” Andy and I worked closely with Howell and the graduate students to modify this and other offending sections; all agreed that the resulting modifications compromised neither the integrity of the school nor of the research.

Another illustration of conflict is the representation of cognitive demand that appeared in the chapter with the cross-case analyses. One section emphasized the low cognitive demand of the instruction and the schools’ focus on the state’s high stakes test (passing was required for graduation). A leader in one school reacted as follows:

Wow—This is TERRIBLE. It reads like an indictment—it seems very one-sided, meaning only negative and not balanced at all. I’m not just over sensitive here—increasing the rigor of our instruction is a constant goal and conversation for us—but the examples and analysis in this chapter doesn’t accurately represent the work
that goes on at Horizon. It's also completely inaccurate that we are solely focused on the [state tests] and use that as our main reference point in classrooms or at the leadership level. . . The tone of the chapter is also condescending, particularly with regard to the SAT, as if we train our kids like dogs for the state test and can't do that for the SAT because it requires too much higher-order thinking for us (perhaps she should compare SAT data disaggregated by race or first-generation college [students] as a more accurate comparison. . . there's a lot more going on in our kids' scores than [what she describes as "low level" instruction]. The conclusion over-reaches and seems biased. Seems to essentially say that the [state standardized test] is pathetic and it's the only measure by which any of our schools have had success (not true). (PJ at Horizon)

Whereas we understood the concerns expressed by the school and association staff, the research team was also a stakeholder group with which the association had entered into a contract. As our relationship with Howell was positive, he sent me this e-mail:

Sharon, This brings into question the integrity of our work. I can't agree to withhold information that is part of why these schools are successful (even if it is not complementary [sic]).

In reaction to the cross-case analysis chapter, one of the schools asked to withdraw from the study, a move that would remove their data from the book. Howell contacted his university's lawyers, claiming "breach of confidentiality" on the part of the association in sharing the case studies with the schools, although the researchers themselves had been the first to make the papers public on the aforementioned website. Again, Howell e-mailed me:

Norman's sharing our drafts with the schools was unauthorized and completely compromises my commitment to my IRB to only share documents about the project where I said I would. Furthermore he has, by sharing these DRAFTS essentially trashed my reputation with these schools and with others in the metropolitan area. Can you imagine someone taking your first draft and sending it to your subjects without permission? Imagine how you would feel, what it would do to your ability to work with those people in the future and your reputation. Does it make you shudder? If not, it should in my humble opinion. I intend to make sure the Program Officer knows full well of this terrible indiscretion.
Good luck on reporting about this mess created by Norman in your overall evaluation. I have every intention of making it well known in Washington funding circles including the DOE, as well as with [others in the innovative school community] and all of the other grantees supported by this initiative from the DOE. It will already have been heard of in DC and in innovative school circles by the time your report arrives.

I was surprised by Howell’s stance. After careful consideration, I sent him this e-mail on the following day:

Howell,

Let me start by repeating that the work of you and your team is quite well done. But that is why I am so surprised and concerned with your reaction—concerned because it runs counter to procedures and ethics of conducting solid qualitative research that honors the idiosyncratic and contextual nature of human experience and allows complex and dynamic interpretations of that experience. The issue is not about academic freedom—it is about rigorous and systematically conducted research and producing valid results.

You ask me to imagine how I would feel were someone to share my drafts with my “subjects.” My response is that I do not have to imagine—I always share my drafts with my participants (qualitative researchers generally do not call the people in their studies subjects) as a validity check. The process is referred to as member checks—not so that members can change the results but so that the researcher can ensure that her interpretations are ones others share. While qualitative research embraces subjectivity (individual perspectives and interpretations), Michael Scriven (in his classic Philosophical Inquiry Methods in Education for AERA, 1988, and his Thesaurus, 1991) acknowledges a form of objectivity can be approached when multiple perspectives are integrated to fully describe the phenomena—thus the need for member checks to ensure validity and a trustworthy study (see the classic text by Lincoln and Cuba, 1985, or Rossman and Rallis, 2003). Without integrating the schools’ perspectives, your interpretation is as subjective as theirs.

The issue is also ethical. The IRB exists to protect the participants/subjects, not the researcher. Qualitative researchers deeply respect the voices of their participants and acknowledge that the process is as important as the outcome (re: John Rawls means must equal ends). Ethically conducted research is transparent—that is, everyone sees clearly how you arrived at your conclusions and accepts your process—not necessarily your conclusions, but the
process must be transparent. Thus, sharing drafts with participants is de rigueur.

Yes, this takes time. Often unbelievable amounts of time. In fact, I recently finished a chapter for The Sage International Handbook of Educational Evaluation—the chapter is based on the work of the Connecticut Superintendents Network (the one Dick Elmore has been facilitating), and I must have run 5 or 6 different drafts by them, making many changes—both minor and major—never changing the findings but the way in which I represented them. The final result is one that they and I are deeply pleased with. And we did have a deadline but the publishers felt it was critical that the chapter offer a trustworthy representation of what actually happened in the network.

I repeat that in general you and the team have used systematic and ethical inquiry appropriate for the purposes of this study. Sharing drafts with participants should be a part of that systematic inquiry.

Sharon

While I received no direct response from Howell, the situation appeared to settle. We heard no further threats, and some revisions softened the offensive sections. Our quarterly report described the situation (Table 9.1).

### Table 9.1 Quarterly Report to Funding Agency

<table>
<thead>
<tr>
<th>Dates/Timeframe</th>
<th>Tasks/progress this quarter</th>
<th>Problems encountered?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul—Aug</td>
<td>Research Team work: reading and compiling joint group-association suggested edits on book chapters; phone meetings and frequent e-mail exchanges with lead researcher and her team; attempting to help negotiate different perspectives of research leader and association team</td>
<td>Monitoring researcher’s adherence to project goals has consumed an extraordinary amount of effort and time. Lead researcher has held to an inappropriate conception of the researcher’s role that is at odds with the goals of the project. The research approach has led several project participants to threaten to pull out of the project. We are concerned that this may threaten the usability of a key project deliverable—the book—but we don’t know for certain as we have not seen the “final” draft to date. Further comments attached below.</td>
</tr>
</tbody>
</table>
BOX 9.4. Questions We Used to Guide Our Qualitative Inquiry and Actions (Challenge 4)

- What effects do our actions and words as evaluators have on project activities?
- What is the nature of our interactions with the various players? What moral principles guided our actions?
- How can our actions and words as evaluators facilitate dialogue across different groups? How can my practice support justice, beneficence, and respect?
- Are we using knowledge and expertise appropriately? At what points, or when, should we intervene—or should we simply report activity?

The situation was further complicated because the publisher wanted a book that raised “controversial issues,” so few further changes were made. As I wrote in an October e-mail to Norman and Alice:

As Andy put it, about 90% is good (maybe only 80%), but the front and the back (what people read when deciding whether to read more) is certainly not what you contracted for. As it appears that we are not likely to get any additional revisions from Howell, the best approach might be to ensure that potential audiences/readers hear the innovative school perspective.

The book went to press with what the schools and association felt to be unsupported judgments. The association chose to back off but to publish their own review of the book.

Challenge 5: Supporting the Filmmaking and Professional Development: Reciprocal Relationships with Other Players

Meanwhile, other components of the BTG project were moving along. The filmmaker was busy taping and editing. The professional development agency was planning and conducting
the study tours. These activities differed both procedurally and substantively from the research component in that those contracted built interactive and positive relationships with the school people so that the schools felt ownership of the content in both these products (film and tours).

For both components, our role was to review and facilitate review by others and to collect data to provide feedback on participants’ reactions. The filmmaker, who had no training in qualitative research methods, demonstrated a near ideal qualitative research process—filming segments and editing was an interactive data collection and analysis process using collaborative question-framing, coding, theme identification, and member checking. Working with Alice, we scheduled and facilitated viewing and review sessions to provide opportunity for representatives from various stakeholder groups to offer immediate and face-to-face feedback to the filmmaker. Through open-ended surveys, we collected brief answers, which we grouped by theme and shared with the filmmaker. Even deciding on the title became a collaborative and iterative process; judging by the e-mail exchanges and time to decision, everyone had an opinion, and all opinions were heard. Thus, no ethical challenges surfaced, because the process was truly transparent and participatory, and any concern was addressed immediately and collaboratively. The process was considered to be mutually beneficial, and the resulting film was seen as a joint effort. At the premier of the final film, leaders from the school remarked on how much they had learned about themselves because of their involvement in the filming and reviewing process. One critique remained: The length of the film was too long.

Design and implementation of the study tours followed a similar process. Again, the professional development experts who had been contracted created multiple avenues during planning and piloting for participants from the five study schools to give feedback. The process resulted in various modifications to the design that contributed to what were ultimately perceived by all as successful study tours. We collected evaluation data through various channels: surveys with rating scales; interviews with open-ended questions; web forums; focus groups; feedback loops for member checks; debriefs and conference calls with the coordinator and tour planners. The following quote was
in answer to the question: What worked well?: “The up-front involvement of Host Schools, based on info and research—Rose and Linda asked them: ‘What are you interested in presenting? They were invested” (study-tour debrief conference call). Our evaluation summary to the association for this phase of the project was positive, emphasizing the evidence that all participants were involved and felt their voices were heard.

CONCLUSION

What makes this case both challenging and interesting is that the evaluators needed to consider ethical practice at two levels—our own and the researchers—and the interactions between both. What moral principles were practiced? Were principles of respect, beneficence, and justice followed? Could relationships be defined as caring and reciprocal? Can the results of both the research and the evaluation be considered trustworthy? How were both care and benefits balanced with burden?

Regarding the research, the questions at first appear straightforward: In what ways did the research harm or care for participants? How did the research benefit the participants? Were individual or group rights respected and protected? What consequences resulted? Evidence indicated that the research process did indeed burden the school participants in their daily work and threatened harm. Moreover, the research team was not qualified to conduct a qualitative study; they were inexperienced
and untrained; for example, as we noted in communications, descriptions were thin, and they referred to participants as "subjects". Not understanding the principles underlying qualitative research practices, the team reverted to and applied what they did know, that is, externally-generated or objectivist procedures and reasoning. Any researcher (but especially one using qualitative methods, because of the close, face-to-face interactions with participants) will encounter ethical issues; thus, competence in conducting research depends on adhering to agreed-on moral principles. This team's lack of skill in handling technical and procedural methods and in failing to articulate their own moral principles limited their ability to deal with ensuing ethical encounters. Put simply, technical or procedural incompetence can easily lead to ethical incompetence. The study and its results were compromised in that the product did not faithfully represent participants' actions, experiences, and intents. Although participants remained open to benefits from the research, they did express feeling both harm and burden.

Given the questionable benefit and care from researchers' interactions with the participants, we needed to question our actions along with the actions of the researchers. Because as evaluators we assumed a more etic perspective (that of a self-conscious outsider and supported by data through one of the five senses), we saw that participants carried a burden that was not seen as balanced with benefit, and we witnessed instances of harm and disrespect. In what ways did our actions facilitate or mitigate the situation? What actions did we, as evaluators, take to support or inhibit the harm-care balance or to facilitate respectful and ethical interactions? Also, the project was funded with federal money, so we questioned whether we met our responsibility to public interests and values; the AEA Guiding Principles for Evaluators remind us to go beyond analysis of particular stakeholder interests to consider the welfare of society as a whole.

Regarding our practices, we asked: Did our actions serve to prevent harm to participants? Did our relationships enact care and respect? Did our efforts support fairness and reciprocity, consistently ensuring that those who bore the burden of our evaluation activities also benefited? Did our decisions serve to protect or enhance the public good? Given the politics of the evaluation setting and the federal demands, could we have acted differently?
While our answers to each of these questions are not all we had hoped for, we do believe we were respectful and fair in all of our interactions.

Throughout the evaluation, we sought to act reflexively; that is, we continually examined how we were making sense of how others were making sense of their worlds (see Rosman & Rallis, 2003), recognizing that all of us—researchers, participants, and evaluators—were morally interdependent, engaged in ongoing and changing interactions that were guided by our respective ethical principles. Caring about the participants and the relationships drove many of our decisions and actions. We believe we practiced caring reflexivity.

We found ourselves taking on several roles that emerged over the course of the BTG project; at various times, we acted as: critical friend; advocate and protector; facilitator of co-construction/negotiator; coach, educator, and technical assistance provider; and ethical interpreter. We believe that acting in each of these roles served us in conducting an ethical evaluation using predominately qualitative methods—and contributed, to some extent, to safeguarding respect for persons, beneficence, and justice, the “de facto Kantian” principles of justice and fairness that are meant to guide ethical deliberations of IRBs (Hemmings, 2006). That these roles were recognized and accepted was confirmed when at the final yearly gathering of the DOE grantees in Washington, DC, Alice (project coordinator) introduced us as her “critical friends. They have helped us define and re-define our work.”

Andy’s and my actions as evaluators can be judged according to AEA guiding principles. Certainly, our data collection and analyses adhered to high technical standards for qualitative methods, exploring with the client (the association) as well as the other participants the best evaluation approaches and methods to offer sufficient detail for them to understand, interpret, and critique their work. Our performance was competent, drawing on our long experience and training in evaluation. We continually reflected on and modified our approaches and methods. We held the integrity/honesty principle: we negotiated; we disclosed; we recorded and reported; we revealed our interests; we represented accurately our procedures; and we worked to resolve the multiple conflicts (both related to any potential
misunderstandings of our work and related to the research. Our overriding concern was to respect the security, dignity, and self-worth of all participants—including the researchers; we advocated for and provided a forum for the less powerful voices. Finally, we were guided by the project's purpose of replication and dissemination for the overall improvement of public education—another AEA guiding principle (responsibility for general and public welfare).

In our work with the researchers, we were, I believe, directed by non-consequentialist moral principles: our concerns focused on the experience of the participants—were the means respectful, caring, and beneficial? We gathered data on the process from multiple sources and continuously checked to ensure that our reading of the data was (and was perceived to be) grounded in thick descriptions. Our feedback, summaries, and questions aimed to encourage dialogic exchanges across perspectives. Our emphasis on the process provided multiple moral challenges.

In the end, the final products—book, film, study-tours—were what mattered most, so the consequentialist principles provide a more useful analysis. The book was published, and nearly all stakeholders agreed that the case studies told important stories of successful innovative schools. We had negotiated and brokered an acceptable understanding between researchers and program: data were revisited and some changes were made, but the final cross-case analyses chapters still represented interpretations from the particular perspective and values of the researchers and were based on what we regarded as "thin" data. Many still believed that the book exceeded the scope of the researchers' contract with the association and their agreements with the schools. Drawing on evaluation data, the association, in agreement with the schools, wrote a widely disseminated review that highlighted what they viewed as positive aspects of the book and raised questions and offered counterarguments to perceived negative aspects. The review ended with the following paragraphs:

It is, however, the final chapter with which we, and the five schools studied, take most issue. Again, it begins well, with useful summary material. But the focus changes to a consideration and critique of standardized testing, which we feel is outside the scope of the book and might have more to say about the researcher's perspective than about the schools themselves. In addition, the author asks whether
government should support schools that focus on college success, a question we find startling, given the realities of the global economy and the high level of skills and education it requires. Finally, the last few pages of the text move into a philosophical consideration of "what should matter" in education. The implication here is that the schools are too narrowly focused on tests and standards, rather than the more "profound" goals of education. We'd argue that preparing low-income, urban students for college is profound, in fact critical, and each school's unwavering commitment to continuous improvement helps them determine and regularly evaluate "what matters."

With these caveats, [the book] is a very useful piece of research and writing. It gives us a window into five high-performing urban innovative schools, describes how they operate, and demystifies the process of closing the achievement gap. The results these schools have achieved—first generation college attendance, entrance into highly competitive high schools, laudable performance on assessments, safe and productive school environments—are impressive, and they offer hope to anyone concerned about improving education for our neediest students. (ASSOCIATION)

Ultimately, the schools were satisfied but harbored a sour feeling for academic researchers, as evidenced by their failure to invite Howell to the final gala, which was in all other ways a joyous celebration. More importantly, the other components, the film and the study tours, were seen as representing balanced views of the schools and their work as well as fulfilling the goals and scope of the project. Borrowing from the consequentialist perspective, I would be tempted to suggest that all's well that ends well. However, a fairness or caring perspective presents a different picture. As one school leader questioned after reading the final version of the book: "Who's story is this anyway? That is NOT what's happening at Horizon!" The question of whose story is told is one any ethical qualitative evaluator should ask.

ACKNOWLEDGMENTS

I am grateful to Andy Churchill, my evaluation teammate, for the hours of conversation and caring reflexivity we shared during the evaluation and for the insights he offered as I wrote. I also appreciate the association project coordinator, who accepted us, the evaluators, as her critical friends. Also helpful were the
comments that Gary Ciarczak, Soria E. Colomer, and Edith Stevens offered during a roundtable discussion on evaluator as critical friend at the 2009 American Evaluation Association Annual Conference.

KEY CONCEPTS

Beneficence
Caring reflexivity
Coding
Consequentialist ethical theory
Critical friend
Culture
Ethic of care
Ethic of individual rights and responsibilities
Ethic of justice
Etic perspective
Feedback loops
Guiding principles for evaluators
Member checking
Non-consequentialist ethical theory
Objectivist
Quasi-ethnographic study
Reflexive reasoning

DISCUSSION QUESTIONS

1. This chapter offers an example of how evaluators encounter surprises. Evaluators can choose the evaluations they agree to conduct; however, once they have agreed to conduct the evaluation, evaluators are obligated to work with all of the stakeholders associated with the program. It is not uncommon to encounter one or more people involved in the evaluation who do not live up to the expectations or standards the evaluators hold. Discuss some strategies for ensuring that the evaluation is of the highest quality, even when one or more players fall short of the evaluator’s expectations or standards.

2. The author indicates that she had needed to negotiate—and renegotiate—the scope and details of the evaluation. The
renegotiation of the evaluation resulted in a "structural" overlap between the research and the evaluation.

a. What might be an advantage of this overlap?
b. Given the problems that resulted, in hindsight, how might the overlap have been structured to prevent these issues?

3. One could assume that the conflict between the school leaders and the researchers would have occurred regardless of the specific role the evaluators played. Would adhering to the original evaluation strategy have helped or hurt the final outcome of the project? Please consider all three elements of the project (research-based book, film, and professional development/study tours), as well as the overall project.

4. Was it ethical to provide a brief training to the research team? Please justify your answer. What other alternatives might the evaluator have considered?

NOTES

1. Names of persons, institutions, and locations have been changed; pseudonyms are used in all cases except for my own name and that of my co-evaluator.

2. All e-mails reproduced here, whether directed to a specific individual or a group, were copied to at least one other person. None was a private communication.

3. Although we assert that in many aspects the research did not honor these hallmarks, we ensured that the issues were raised openly, and ultimately the offending sections were removed from the book.

REFERENCES


