

7

The editors give each of the theorists a chance to respond to questions posed about the context of the situation in relation to their own experience in the field, exploring how the exercise had an impact on their evaluation designs.

The Reality of Unreal Situations: Caveats and Insights

*Marvin C. Alkin, Christina A. Christie, Jennifer C. Greene,
Gary T. Henry, Stewart I. Donaldson, Jean A. King*

The specifics of the case (the context) are highly important to how different theorists will react in conducting an evaluation. By this, we think of contexts in a very broad sense. A context is not only a disciplinary field (say, education) or a specific kind of program within that field (Head Start, for example), but also all the particularities that relate to the organization, the individuals within that organization, and those who are affected by the program. The context is also influenced by how broadly the evaluator chooses to view the program.

In the instance of Bunche–Da Vinci, we have an elementary school in a low-income, heavily bilingual area. We also have a program that is a unique partnership of the school and the Da Vinci Learning Corporation. Each of the theorists has done evaluations related to education, though not necessarily elementary schools. The four theorists admittedly have differential levels of experience in dealing with instances like Bunche–Da Vinci. King and Greene have had the most firsthand experience with situations like that presented to them. Henry has dealt with the evaluation of educational programs, but within the context of public policy. Thus, his experience with evaluating education is at a more macro level. Donaldson has performed evaluations in a variety of fields, with his primary exposure to education outside the context of K–12. What impact does this have on the way in which the theorists deal with Bunche–Da Vinci?

The way this exercise was constructed also has implications for how the theorists might approach the evaluation. Some theoretical approaches

are simply less compatible with this exercise. Some evaluators might feel uncomfortable with being commissioned by a school principal to perform an evaluation of the program. Some might desire to work with higher-level administration to ensure that the evaluation is directed at someone with the authority to make decisions. Others might find it inimical to conduct an evaluation in which there is not initial broad stakeholder involvement. Thus, it is clear that the evaluation theorists' responses in the previous chapters represent but a single instance of how they might implement an evaluation, and the parameters of the scenario may have constrained their ability to explicate their approaches fully.

To examine the extent to which the context of this exercise and the exercise itself had an impact on the products produced by the four theorists, we posed the following four questions for their response:

1. What information did you see as missing from the case, and what major assumptions did you make in the absence of that information?
2. How similar (or dissimilar) is the setting described in the case to those in which you typically conduct evaluations?
3. To what extent can the evaluation proposed in your chapter be viewed as fully explicative of your approach?
4. What is your assessment of the likelihood that you would be able to implement and complete the evaluation as you have described it?

Responses

What information did you see as missing from the case, and what major assumptions did you make in the absence of that information?

JENNIFER C. GREENE: There was insufficient information on the educational, organizational, and political context of the school district and its urban setting. Urban school districts are enormously complex settings, with their own internal political challenges and the external demands of accountability adding considerable pressure to those in charge. I made few assumptions about these dimensions of the Bunche–Da Vinci evaluation context. Instead, I concentrated on the smaller context of the school, assuming that Mary García as the principal had the authority and responsibility to act without extensive consultation with district authorities.

Also missing was information about the Da Vinci Learning Corporation and about the theory and logic of its program for “low-performing schools.” In this regard, I assumed the corporation was to be a minor player in the evaluation. And I held the theory of the program up for evaluative scrutiny and critique.

It was also very difficult to get a sense of the school community—the life of the school’s neighborhoods and their residents. Particularly perplexing was the apparent mobility of the neighborhoods and the mobility of the students in the school, juxtaposed with reported parent and student satisfaction with the school. Especially, who are the Spanish-speaking families coming into the community and the school? Are they recent immigrants or people moving from elsewhere? From what Spanish-speaking countries do they originate? How homogeneous or heterogeneous is this particular community? And so forth. My evaluation design did not assume these were central issues in the school but did include and attend to them.

Furthermore, there was little information on the legal parameters of the public-private partnership that was the Bunche–Da Vinci Learning Academy. I assumed this was worthy of further investigation and critique in the evaluation.

GARY T. HENRY: At first, I was stymied by the single school focus of the description. After thinking more deeply about the circumstances, I realized that many of the questions that García and Chase would find compelling were most likely to be comparative and that the philanthropies were likely to be interested in the learning model as well as Bunche–Da Vinci Academy. Since the evaluations that I have done are statewide and comparative, and generally include counterfactuals, elevating the evaluation’s focus made the proposal more similar to evaluations that I have done.

Rather than make assumptions, I developed a comprehensive approach to the evaluation. The proposal contains almost all of the elements that I would frequently include in an evaluation, although rarely am I able to include all four in a single evaluation proposal. They are likely to be sequential, which is consistent with the proposal, but often they are funded separately.

STEWART I. DONALDSON: The most notable challenge I encountered was not being able to ascertain and verify information from dynamic interactions with the stakeholders. This forced me to use a new and untested evaluation method to make decisions: my imagination. I highlighted many of the major assumptions I needed to make in order to illustrate more fully my approach. I think the more subtle central assumptions I made included the assumption that García and Chase would allow us the time and access to meet with the diverse stakeholders to discuss the program and evaluation. I also assumed they were truly interested in objective external evaluation, and not just planning to use my evaluation to support decisions they had already made, personal agendas, or as a way to find data that could be used to justify terminating the contract with their corporate partner.

JEAN A. KING: There are at least two categories of critical unknowns in the setting that would raise my interpersonal antennas and make me ever vigilant to signs of trouble. First are the uncertainties of this evaluation's client and the role that district politics may unavoidably play at Bunche–Da Vinci, regardless of what we accomplish within the school. Numerous questions spring to mind: How long will Superintendent Chase give free rein to Mary García and her reform efforts? How actively are school board members involved with the school, and how do they perceive its situation? To what extent will the nationally conspicuous collaboration with Da Vinci affect changes the school might propose? How viable is the Da Vinci collaboration to begin with? What is the district's budget situation in the coming years?

Second are the multiple uncertainties related to the school itself and to the people who work there: the instructional program as implemented, the details of the principal's proposed and enacted changes, the depth of discontent among the faculty and staff, García's leadership capabilities, the realities of the community (for example, historical tensions, leadership, and other assets available), and so on. The project would likely demand a great deal of presence on-site to resolve these uncertainties and address the conflicts or dilemmas arising from them. My chapter assumed that the principal agreed to evaluation capacity building using a participatory process over time and that none of these issues would interfere with the evaluation process. In all likelihood, unknown elements will unavoidably affect the process in ways that can be understood only as they develop.

2. How similar (or dissimilar) is the setting described in the case to those in which you typically conduct evaluations?

JENNIFER C. GREENE: I am very familiar with public school K–12 settings, including schools in underserved and underresourced settings. I have conducted many evaluation projects in our nation's public schools, and my graduate education is in the field of education. Thus, I bring considerable experience and understanding to a public K–12 evaluation context. I am deeply concerned about the well-being of marginalized people in our society and about the inequities and injustices they continue to suffer. And I have devoted time to thinking about how evaluation can be best positioned to serve the interests of such people. I am most committed to *public* programs and policies that concern our national priorities and our democratic commitments. Thus, in multiple ways, the Bunche–Da Vinci scenario is a good match to my experience and sensibilities as an evaluator.

GARY T. HENRY: After shifting from the single school to a comparison of schools using the learning model, the setting was reasonably similar to those in which I frequently work. In most cases, I am interested in evaluating interventions that are intended to improve educational performance and I

am always focused on outcomes. So with a sleight of hand to elevate the focus and make it comparative, the setting became comfortable for me and the type of evaluation I usually conduct.

STEWART I. DONALDSON: My career has involved evaluation adventures into many diverse settings and contexts. For example, I have worked on evaluations in a variety of education, corporate, university, public health, health care, community, and nonprofit settings. While I have never encountered a case exactly like Bunche–Da Vinci and do not specialize in K–12 evaluation, many features of this case seemed familiar. Namely, I have recently evaluated technology-enriched curricula and measured student performance, evaluated educational programs for ethnically and culturally diverse bilingual youth living in low-income and impoverished conditions, and worked on evaluations that engaged school and corporate administrators. I would certainly agree this is not a typical case for me, but I did not feel I was a complete stranger to the context or at a loss for predicting issues and challenges that would likely emerge.

JEAN A. KING: I have worked in urban school settings for many years although more commonly at the secondary level than the elementary, so the Bunche–Da Vinci setting would be fairly typical for me. One small consequence of the No Child Left Behind legislation, with its emphasis on standardized testing, which often trumps program evaluation in districts, has been to move my evaluation practice beyond schools. The general process of capacity building applies to these differing contexts well; the contexts affect how it develops over time.

3. To what extent can the evaluation proposed in your chapter be viewed as fully explicative of your approach?

JENNIFER C. GREENE: The proposed evaluation reflects my views on evaluation to a significant extent. It captures much of what I believe is important in evaluation practice. This includes active and explicit engagement with critical values in evaluation (contextuality, inclusion, learning and education, diversity and difference, and the public good); an emphasis on the meaningfulness of lived experience, or on the program as experienced in context by various stakeholders; attention to the relational dimensions of being an evaluator and especially to the character of respect and reciprocity therein; and an understanding of evaluation as a moral and ethical practice, in and of itself.

Some of my other evaluation commitments and predispositions did not get as much airtime as I would have ideally liked. These include the following. First, the evaluation design I proposed does include a mix of methods and methodological perspectives. But this mix is not as thoughtful and therefore not as powerful as it potentially could be. Particularly in politi-

cally challenging contexts, a thoughtful mix of methods can provide an important avenue for respectful and meaningful engagement with varied perspectives and interests. This could be very useful in the Bunche–Da Vinci context. Second, the evaluation design I proposed pays insufficient attention to these internal politics, particularly to the interests and concerns of important players in district and corporate politics. In the value-engaged approach I am promoting in this volume, the value of inclusion means inclusion of not only the marginalized but also those currently in power, as their concerns are also important and legitimate. The design overlooked, even neglected, the interests of the powerful. Third, I believe that theory-oriented perspectives on evaluation are also useful and potentially powerful. These were underused in this design.

GARY T. HENRY: The main assumption that was necessary was that the principal and the superintendent would be interested in comparative data about the implementation and effectiveness of the Da Vinci Learning Model. I would like to have better understood why parents preferred this school but complained about the focus of the instruction and emphasis on technology. I also wondered about the absence of district personnel and district-initiated activities.

STEWART I. DONALDSON: My approach has evolved, and continues to evolve, in response to developments in the discipline and practice of evaluation and my cumulative experiences practicing evaluation science. The approach I described and used in this volume reflects my current thinking about a rather specific but substantial slice of evaluation practice. I refer to this domain as program effectiveness evaluation (Fitzpatrick, 2002), which involves evaluating a program under real versus ideal conditions. I distinguish this from program efficacy evaluation, or evaluating a program under ideal or research-like conditions. In my view, many of the contemporary debates and even texts in evaluation do not make this distinction clear enough. It is very common to observe program effectiveness evaluations being designed with, or judged by, program efficacy evaluation methods and standards.

I have developed my approach to program evaluation so that it can be widely used and adapted to both worlds: program efficacy and effectiveness evaluation. I have used program theory-driven science employing randomized control trials to evaluate program efficacy (Donaldson, Graham, and Hansen, 1994) and to isolate harmful side effects of programs implemented in public and private elementary and intermediate school settings (Donaldson, Graham, Piccinin, and Hansen, 1995). I have also used program theory-driven evaluation science in a wide range of program effectiveness evaluation projects (for example, see Donaldson and Gooler, 2003). I believe one of the most desirable features of this program evaluation approach is that it promises to bridge the divide between social and evaluation scientists

evaluating program efficacy (see Lipsey and Wilson, 1993), and the work of evaluation practitioners typically conducting program effectiveness evaluations. The proposed Bunche–Da Vinci evaluation is intended to be explicative of how I would approach a program effectiveness evaluation based on the information I was provided and the assumptions I made.

JEAN A. KING: The chapter I wrote in response to the case accurately details my current evaluation approach. I believe so strongly in collaboration that I will not typically take a job in which clients request what I would call an evaluator-driven evaluation, that is, where the evaluator (or external evaluation team) is solely responsible for decisions related to the study (for example, the questions, the methods, the data collection and analysis, and report development). In the past few years, my approach has moved directly into the clients' setting, and I now choose projects where individuals are interested in building their personal and organizational capacity to conduct evaluations on their own after I leave.

4. *What is your assessment of the likelihood that you would be able to implement and complete the evaluation as you have described it?*

JENNIFER C. GREENE: The likelihood that I could implement and complete this evaluation as proposed is moderate to high, unless I have radically misread the situation or there is vitally important information currently missing from the scenario—for example, a threatened civil rights lawsuit by concerned parents or an educational budget shortfall of millions of dollars in the district. If there are no critical factors missing from the scenario, I feel the likelihood of successful implementation is moderate or high because the design is a relatively modest one, which is intended to fit within the existing rhythms of the school as much as possible and has the potential for generating valuable information for stakeholders in the context.

Challenges to successful implementation are headed by the evaluation design's commitment to including all legitimate stakeholder groups in key evaluation decisions. Evaluators usually believe their work is of considerably more importance than program stakeholders in the context at hand. It remains very difficult and challenging to involve already busy people and people with little history of being invited to participate in the work of evaluation. While this is not a fully participatory evaluation process, it does require meaningful stakeholder involvement in order to be successful. Without such participation, the evaluation risks being an activity disconnected from the lifeblood of teaching and learning in the Bunche–Da Vinci Learning Academy. Yet only with such participation, at least by a handful of influential stakeholders, can the evaluation fulfill its promise.

GARY T. HENRY: Of course, an obstacle to conducting the evaluation would be the amount of funding required, but education evaluation sponsors and policymakers are asking for better evidence before propagating programs of questionable effectiveness. Solid evidence about effectiveness doesn't come cheap, and this evaluation would be expensive. Ideally, an efficacy study, which would have been less costly than the evaluation that I proposed, would have been required before the Da Vinci Learning Model was accepted at Bunche School. It is more common to have policy out in front of evidence than the reverse, unfortunately.

STEWART I. DONALDSON: I have come to appreciate over time that clients, stakeholders, and people in general are highly unpredictable in evaluation practice, especially when you are required to base predictions on superficial information. Programs and organizations like the Bunche Academy and Da Vinci Corporation, are also typically dynamic and unstable. However, if García and Chase are truly interested in professional evaluation and my assumptions are accepted, I think it is very likely that steps 1 and 2 would be implemented with great success. They could be so successful that step 3 is not necessary. If step 3 is needed, I believe it is likely it could be implemented and completed as I have described. But please be cautioned: the devil is almost always in the program or evaluation plan implementation.

JEAN A. KING: Despite the uncertainty, my practice in numerous settings over many years supports two probable results. I am fairly confident that even in the worst-case scenario where the unknowns described above become insurmountable constraints, we could both implement and complete an evaluation process and that it would make at least some difference. First, it would generate new data and compile existing data, enabling us both to attain the proposed evaluation outcomes related to the school's instructional program (that is, knowledge about the implementation of the principal's new system and understanding of the school's parallel test scores) and to make collective sense of what people might do to improve it. Second, the collaborative evaluation process—the active involvement of as many Bunche–Da Vinci faculty and staff as are willing to participate—would likely have its own outcomes. People would pick up the evaluation process by engaging in it, they would unavoidably learn the accountability mandates facing the school, they would develop trust as they reflected together, and, in small ways at first, they would begin to systematize their individual, intuitive evaluation mind-sets.

While I am hopeful that “evaluation present” would generate results for the school, I am less sanguine about “evaluation future”: the creation of a viable infrastructure for evaluation at Bunche–Da Vinci. Ideally, at the end of the first year, several components would be in place: the Evaluation Advisory Committee, a subcommittee for studying test scores, a model for highly visible group studies, and perhaps a process for individual action

research. But I know how incredibly difficult it is to sustain evaluation activities in schools and other nonprofit organizations. Those who know and support evaluation internally, the key actors to sustain the infrastructure, may leave or become otherwise engaged professionally or personally. External mandates may overwhelm a fledging infrastructure with limited resources. Resources, both those explicit to evaluation (such as data entry and analysis) and those that support it (such as the availability of substitute teachers), may dry up. All of those possibilities are likely at Bunche–Da Vinci. In my experience, though, the school is likely to benefit in the short run from the evaluation activities: people will study the principal's changes, make collective sense of the test scores, and perhaps engage in individual improvements in their classrooms, for example. In addition, as I have recast this evaluation situation, there is the added benefit of building an evaluation infrastructure. The likely beneficiaries of this are some of the individuals who participate—those who increase their evaluation knowledge and skills or the people who gain positive attitudes toward evaluation. Regardless of the outcomes of this study, such individuals may pursue additional training or become leaders of later evaluation efforts.

Editors' Thoughts

Any case scenario, no matter how thoughtfully done, cannot possibly provide all of the information that an evaluator might want to know. It certainly cannot capture all of the nuances of the situation. Theorists were urged to make appropriate assumptions in order to fill in missing information necessary for them to be able to complete the exercise. Explicit statements of understandings and actions within the chapters themselves fail to convey all of the thoughts about the situation that must have run through the minds of each of the chapter writers.

By asking theorists to indicate information that was missing and the major assumptions that they made, we hoped to understand further the mind-set under which they operated. They did not disappoint us. With great clarity, the theorists pointed out a panoply of ideas. Greene expressed dismay about the insufficiency of information about the “educational, organizational, and political context of the school district and its urban setting.” Other theorists elaborated on this concern by pointing to such elements as district politics, school attitudes, and “the realities of the community.”

The case was complicated by the presence of the Da Vinci Learning Corporation and the need to better understand the organization and its program, as well as further details of the contractual relationship between Da Vinci and the school district.

Assumptions can be only tenuously stated, but even with additional information cannot be affirmed without face-to-face interactions with stakeholders. This characteristic of the exercise, of course, was the basis for much

of the difference to be found in presentations. Theorists were allowed (indeed required, because of the lack of information) to make assumptions about the results of interactions. This is perhaps unavoidable in a simulation exercise.

Theorists acknowledged a diversity of theoretical views and prior experience with contexts similar to Bunche–Da Vinci. Greene and King, in particular, were very comfortable with the typicality of the setting in relationship to their experience. Donaldson and Henry were less familiar with this specific kind of setting, but nonetheless felt able to deal with this situation. We do not know whether the theorists would actually behave in practice in the ways described in these chapters. Would Henry, for example, redirect the evaluation question to provide a firm foundation for his approach? Would King either refuse the evaluation or redefine it into a process of long-term, collaborative capacity building, regardless of the client's urgent questions? We only have the stories that each of the authors told and must assume that their evaluation descriptions are a reasonable starting point to what they would do.

Notwithstanding differences in familiarity with setting and necessary assumptions, theorists generally felt comfortable with their ability to propose an evaluation for Bunche–Da Vinci. All of them felt that the product they had produced represented their views on evaluation at an appropriate level. Greene, for one, expressed regrets about elements of the design that she wished that she had had more time to address (mixed methods, internal politics, theory-oriented perspectives). However, even with these minor regrets, she felt that her views had been adequately represented. So too might the other theorists have wished to include additional items in their design.

Evaluators are optimistic about the potential success of their practice. All of the theorists are at least moderately optimistic that they could implement and complete the evaluation as described. The editors of this volume are less sanguine that all would go as proposed. In our view, Greene's evaluation proposal is more typical (in our view) of evaluation practice and moderate in its expectations. We believe that it could be implemented quite adequately, but perhaps without as much stakeholder involvement as she might have wished for. Henry has proposed a very broad-based study that constitutes a multiyear evaluation of the Da Vinci program. It would require a substantial amount of funding and cooperation and participation from many governmental agencies. There could be great difficulties in implementing such an evaluation.

Donaldson proposes an evaluation that focuses heavily on understanding the program theory of the Bunche–Da Vinci model. That aspect of it requires a great deal of participation that might be viewed as helpful but time-consuming. We believe that the step 1 and 2 tasks that he describes are quite doable, and perhaps the rest of the evaluation is as well. King is confident that “even in the worst-case scenario . . . we could both implement and complete the evaluation.” We hope that she is right. However, we believe that the amount of time necessary for participating in all of the activities that

she describes could be a major impediment. As King notes, she is very selective about the evaluation projects that she is willing to undertake and engages in a process of reconnoitering and doing substantial upfront work in order to determine whether to accept the assignment. By engaging in the process, she has been forced to make the assumption that Bunche–Da Vinci passes the test. But her response to the first question casts some doubt on whether that assumption is correct.

References

- Donaldson, S. I., and Gooler, L. E. "Theory-Driven Evaluation in Action: Lessons from a \$20 Million Statewide Work and Health Initiative." *Evaluation and Program Planning*, 2003, 26, 355–366.
- Donaldson, S. I., Graham, J. W., and Hansen, W. B. "Testing the Generalizability of Intervening Mechanism Theories: Understanding the Effects of School-Based Substance Use Prevention Interventions." *Journal of Behavioral Medicine*, 1994, 17, 195–216.
- Donaldson, S. I., Graham, J. W., Piccinin, A. M., and Hansen, W. B. "Resistance-Skills Training and Onset of Alcohol Use: Evidence for Beneficial and Potentially Harmful Effects in Public Schools and in Private Catholic Schools." *Health Psychology*, 1995, 14, 291–300.
- Fitzpatrick, J. "Dialog with Stewart Donaldson." *American Journal of Evaluation*, 2002, 23(3), 347–365.
- Lipsey, M. W., and Wilson, D. B. "The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from Meta-Analysis." *American Psychologist*, 1993, 48, 1181–1209.

MARVIN C. ALKIN is an emeritus professor in the Social Research Methodology Division in the Graduate School of Education and Information Studies at University of California, Los Angeles.

CHRISTINA A. CHRISTIE is an assistant professor and associate director of the Institute of Organizational and Program Evaluation Research at Claremont Graduate University.

JENNIFER C. GREENE is a professor of educational psychology at the University of Illinois, Urbana-Champaign.

GARY T. HENRY is a professor in the Andrew Young School of Policy Studies and Department of Political Science at Georgia State University.

STEWART I. DONALDSON is dean and professor of psychology at the School of Behavioral and Organizational Sciences, Claremont Graduate University.

JEAN A. KING is a professor at the University of Minnesota, where she coordinates the Evaluation Studies Program in the Department of Educational Policy and Administration.