

I An empirically derived comparative framework offers an understanding of the similarities and differences between eight theories of evaluation when they are applied.

What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory

Christina A. Christie

Over the past thirty years eminent evaluation theorists have appealed for increased empirical knowledge of evaluation, based on the notion that such knowledge is necessary to explain the nature of evaluation practice (for example, Cousins and Earl, 1999; Stufflebeam and Shinkfield, 1985; Worthen and Sanders, 1973). By *empirical knowledge*, I mean experientially based knowledge acquired through formal study (Smith, 1983). Although an appeal for the empirical study of evaluation practice has been made repeatedly, it has met with little response.

Many evaluation scholars agree that the link between evaluation practice and theory is an area of much-needed inquiry (Cousins and Earl, 1999; Scriven, 1991; Shadish, Cook, and Leviton, 1991; Smith, 1993; Worthen, 1990). The benefits of the empirical study of current evaluation practice include, but are not limited to, generation of information necessary to refine current practice and develop alternative approaches to evaluation (Smith, 1993), advancement of conceptions about the connection between theory and practice, and increased understanding of the influence of context on the nature of evaluation practice.

I am indebted to the eight theorists for their participation in this study and to Marvin C. Alkin for his invaluable insight and thoughtful discussion that served as the inspiration for this study. This chapter is based on my doctoral dissertation: "What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory" (University of California, Los Angeles, 2001).

The benefits of empirical knowledge for the advancement of theory and practice are evident when examining the seminal studies on utilization conducted during the 1970s and early 1980s (for example, Alkin, Daillak, and White, 1979; Alkin and Stecher, 1983; Patton and others, 1977). Knowledge gained from these studies helped reshape notions about use (Patton, 1997). In fact, empirical studies provided the foundation from which a class of evaluation theories focusing on use were developed and advanced. These theories, built in part on the empirically derived knowledge of the factors that influence and contribute to use, aim to ensure that evaluation results have a direct impact on program decision making.

Although the benefits of empirical study have been established, very few comprehensive studies investigating the influence of evaluation theory on practice have been published. The scarcity of empirical research on evaluation can be partially explained, perhaps, by evaluation's inherently pragmatic, rather than empirical, aim. Yet one of the most significant if least common modes for enhancing evaluation theory and practice has been empirical study (Shadish, Cook, and Leviton, 1991). Cousins and Earl (1999) advise: "We need to move beyond the relatively weak connected theoretical musings and anecdotal reports of practice. . . . We need to add to the empirical knowledge base through carefully developed and executed studies that have the potential to extend our theories and guide our practices, studies that manifestly strengthen the link between theory and practice" (p. 316).

The current study provides initial research on a range of evaluation theories and how these theories can be translated into practice. Presented is a descriptive comparative framework depicting how theories are used in practice by theorists and evaluators. The primary objectives of this study are summarized by the following three research questions: *How do the reported practices of evaluation theorists configure empirically? How can these reported practices best be described and compared? And how does the reported practice of evaluation practitioners compare with the practice of evaluation theorists?*

Connecting Evaluation Theory to Practice Empirically

Existing studies have investigated the link between specific theoretical issues and practice. They have considered issues such as how practitioners examine their own practice (Williams, 1988), program implementation (Pressman and Wildavsky, 1973), federal evaluation practices (Wholey and others, 1970), and evaluation utilization (Alkin, Daillak, and White, 1979; Alkin, Koscecoff, Fitzgibbon, and Seligman, 1974; Patton and others, 1977; King and Pechman, 1982). There is consensus, however, that the most heavily studied issue has been evaluation utilization (Shadish, Cook, and Leviton, 1991; Smith, 1993).

Shadish and Epstein (1987) are among the few who have attempted to go beyond utilization issues to study a broad array of evaluation practices.

They explored the frequency and covariates of evaluation practices related to methods, timing of evaluation, the role an evaluator should play, sources and types of questions to ask, how to construct dependent variables, and how to facilitate use. The results of this study revealed four discriminate patterns of evaluation practice: *academic*, similar to basic science, and quantitative studies; *stakeholder service*, which formulates evaluation questions based on stakeholder needs; *decision-driven*, which formulates questions based on pending decisions and legislation; and *outcome*, which judges the effectiveness and worth of the program. Smith (1993) asserted that the information yielded from this study “greatly facilitates the interpretation of subtle points of theory and practice” (p. 239).

The Shadish and Epstein study (1987) employed instruments designed by the study researchers. And the inherent difficulty in designing a survey instrument that ascertains the theoretical orientations of practicing evaluators presents a shortcoming of the Shadish and Epstein study. Theoretical orientation often cannot be accurately assessed through direct questioning because evaluation practitioners usually are not proficient in theory (Smith, 1993), and so are unable to identify their particular theoretical orientation. Identifying the particular theory guiding an evaluator’s work would require the careful delimitation of practices related to particular theories. This is problematic, because delimitation assumes that a theorist has communicated his or her intent clearly enough to be delimited. Given the clearly stated intent of a theorist, delimitation is still contingent on a person’s accurate understanding and interpretation of theories.

The alternatives to survey designs also have limitations. Observational studies would likely require intensive shadowing of the evaluator. Case studies—the most common method used when studying evaluation practice—are limited to depicting the practices of a small, select group of evaluators. This approach provides depth but not breadth of understanding of how evaluators use theory to guide their work.

The current study is unique and different from previous empirical studies of evaluation in several important ways. First, it is one of the few survey studies of evaluation practice. Thus, it offers relatively general empirical knowledge about evaluation practice. Second, this study attempts to overcome the limitations placed on previous empirical survey studies of evaluation practice by the survey instruments used. The instrument for the current study was developed in concert with eight distinguished evaluation theorists representing a broad array of theoretical evaluation perspectives: Robert Boruch, Huey-tsyh Chen, J. Bradley Cousins, Elliot Eisner, David Fetterman, Ernest House, Michael Patton, and Daniel Stufflebeam. The theorists were asked to submit survey items corresponding to three dimensions of evaluation practice—methods, values, and uses—that they would employ to determine whether an evaluator was using their theory to guide his or her work. Thus, the theorists’ understanding of their own theories is represented in the items, rather than the ideas or agenda of the independent researcher.

Also unique to the present study is the sample, which consisted of 138 evaluators in the Healthy Start program. Healthy Start is a statewide California Department of Education program designed to link schools to a variety of academic, social, and health services with the goals of assisting families to meet basic needs, such as food and shelter, and improving children's physical and mental health. Every Healthy Start site is required to conduct an evaluation. But not all of the people conducting these evaluations are evaluators per se. Instead, some are program and school administrators—suggesting that the characteristics of the sample resemble those of the people who typically conduct today's evaluations.

The present study also answers the direct call that Shadish made, in his 1998 presidential address to the American Evaluation Association, for the development of a comparative framework of evaluation practice. During this address, Shadish argued that evaluation theory is at the core of evaluation's identity as a profession. He appealed specifically for the development of a comparative framework of current evaluation practice in order to help identify "real theoretical innovation" (p. 16).

Conceptual Framework of Current Study

Alkin and Ellett (1985) maintained that all evaluation prescriptive theories must consider: "(a) issues related to the methodology being used; (b) the manner in which data are to be judged or valued; and (c) the user focus of the evaluation effort" (p. 1279). Building on this notion, Alkin and House (1992) developed a classification framework for evaluation theories. The Alkin and House taxonomy has three dimensions: *methods*, *values*, and *use*. In each of the dimensions lies a continuum that further defines it. The continuum in the methods dimension ranges from quantitative to qualitative. The continuum in the values dimension extends from unitary to plural—that is, using a single criterion (unitary) or multiple (pluralist) criteria when making evaluative judgments about programs. The continuum in the use dimension extends from enlightenment to instrumental: enlightenment use is academically oriented, contributing to the understanding of an issue, whereas instrumental use is service oriented and concerned with directly informing decision- or policy-making processes.

Theories can be categorized into one of the three dimensions based on their predominant focus. That is, theorists are first placed in the one dimension they consider to be their greatest priority. Their ideologies are then further distinguished by their position along the continuum inside that dimension.

Method

The current study was conducted in several stages. First, statements were solicited from eight distinguished evaluation theorists whose views represent varying theoretical perspectives, in order to construct a *theory-to-practice*

instrument, which was the main source for data collection in this study. Next, the statements generated were used to construct the core items of the theory-to-practice assessment instrument. Finally, the eight theorists and 138 practicing evaluators completed the instrument.

Theorist Selection. The eight theorists were purposefully selected to represent the widest range of theoretical positions possible, within reason, in the Alkin and House taxonomy. The method dimension includes Robert Boruch and Elliot Eisner; Boruch (see Boruch, Snyder, and DeMoya, 2000) is on the quantitative and Eisner (see Eisner, 1985) on the qualitative end of this continuum. The values dimension includes Ernest House, David Fetterman, and Huey Chen. The similarity between House (see House, 1991) and Fetterman (see Fetterman, 1996) is notable; both are concerned with empowering the underrepresented, and their evaluation approaches can be categorized as social agenda–advocacy oriented. Chen lies on the opposite (unitary) end of the continuum. His theory-driven evaluation approach (see Chen, 1990), is concerned mostly with using program theory to frame an evaluation. The use dimension includes Daniel Stufflebeam (see Stufflebeam, 1983), Michael Patton (see Patton, 1997), and J. Bradley Cousins (see Cousins and Earl, 1995). Each of these theorists views use as a crucial goal of evaluation, yet there are notable differences in their approaches. One important difference is that Patton and Cousins are concerned with defined, finite potential user groups.

Development of Study Instrument and Procedure. The eight theorists were asked to assist in developing the theory-to-practice instrument. Each was asked to submit at least one statement demonstrating the practical application associated with his theoretical standpoint related to each of the three dimensions of evaluation (methods, values, and use). In addition to the one statement for each dimension, they were free to submit up to six additional statements corresponding to any of the three dimensions as they saw appropriate or desirable. In order to assess the degree to which an activity is or is not carried out when using a particular theory, the theorists were asked that they write their statements to accommodate an eleven-point response scale (0 to 10), with responses ranging from “This statement is very similar to how I conduct evaluation” to “This statement is very dissimilar to how I conduct evaluation.” The final instrument contained thirty-eight items related to evaluation practice (sixteen related to methods, twelve related to values, and ten related to use).

The instrument was piloted with five practicing evaluators for readability and clarity. Each evaluator was asked to complete the instrument, including the background items. Immediately after the evaluator completed the instrument, each was interviewed for up to forty-five minutes to solicit feedback. Feedback led to minor revisions on only a few items, which did not entail altering content in any way.

The eight theorists and 138 Healthy Start evaluators then completed the instrument. They were instructed to respond according to how they *actually* conduct evaluations rather than how they would *ideally* conduct evaluations.

Study Participants. Evaluation practitioners surveyed in this study were recruited through California State's Healthy Start program. Through Healthy Start program funds, schools develop collaborative partnerships with local health and service agencies. There is no single model for the services; rather, local collaboratives design a specific program to fit the needs of children and families in their community. Available services range from basic needs (food, clothing, shelter), to family support (parenting education and child care), to health and medical care, tutoring, dropout prevention, and job training. Included in the Healthy Start award is the stipulation that all grantees conduct a yearly evaluation of their program.

Healthy Start program evaluators were selected as study participants for several reasons. First, the program is unique in that though its focus is on social service coordination it is housed on public school campuses. This allowed us to obtain a sample of evaluators practicing in at least two different areas: education and social service. Second, Healthy Start is a statewide program, creating a sample in which the contexts where evaluators practice vary naturally. Last, because each program is unique, the state does not determine the evaluation questions; however, it does require that specific elements of the Healthy Start program be evaluated. These requirements help to standardize, at least to some extent, the possible range of evaluation questions the participants investigate. This standardization helps control for differences in evaluators' practice that might occur as a result of the type of evaluation questions investigated.

Table 1.1 presents the demographic characteristics of the Healthy Start evaluator sample. In this table, internal evaluators (64.5 percent of the sample) and external evaluators (35.5 percent of the sample) are separated in order to highlight variations between the groups.

The sample was largely female (78.3 percent) and white (76.8 percent), with 64 percent being both. As indicated in Table 1.1, the internal evaluators were more racially diverse than the external group ($\chi^2_{3,.01} = 10.971$). The greater part of the sample was over age forty, and there were no apparent meaningful differences in age between the external and internal evaluators. With respect to education, statistically significant ($\chi^2_{3,.001} = 19.108$) differences were found between external and internal evaluators, although most of the sample (71.7 percent) reported having advanced degrees. A large portion reported having a master's degree (57.2 percent), whereas far fewer (14.5 percent) reported having a doctoral degree. However, 75 percent of those with doctoral degrees were external evaluators.

When asked to rate their evaluation knowledge and skills, internal evaluators had a mean score of 2.6 (possible range from 1 to 4), whereas external evaluators had a mean score of 3.5. This observed difference was statistically significant ($\chi^2_{3,.001} = 35.032$). The lower mean score could be attributed to experience level because, overall, the external evaluators reported having significantly ($\chi^2_{4,.05} = 11.498$) more years of experience conducting evaluations than the internal evaluators.

Table 1.1. Demographic Description of Healthy Start Evaluators

	Percent of External Evaluators (<i>n</i> = 49)	Percent of Internal Evaluators (<i>n</i> = 89)	Percent of Total Sample (<i>n</i> = 138)
Gender:			
Male	26.5 (13)	17.2 (15)	20.3
Female	73.5 (36)	82.8 (72)	78.3
Age:			
Under 30	6.1 (3)	8.2 (7)	7.2
31–39	28.6 (14)	18.8 (16)	21.7
40–49	32.7 (16)	30.6 (26)	30.4
50+	32.7 (16)	42.4 (36)	37.7
Race:			
Asian	0	2.3 (2)	1.4
Black	2.0 (1)	4.6 (4)	3.6
Latino-Latina	2.0 (1)	17.2 (15)	11.6
White	89.8 (44)	71.3 (62)	76.8
Mixed race-ethnicity	6.1 (3)	4.6 (4)	5.1
Education:			
High school-some college	0	9.0 (8)	5.8
Bachelor's degree	18.4 (9)	24.7 (22)	22.5
Master's degree	51.0 (25)	60.7 (54)	57.2
Doctorate degree	30.6 (15)	5.6 (5)	14.5
Self-rating of evaluation knowledge and skills:			
Minimal	2.1 (1)	9.1 (8)	—
Average	10.4 (5)	37.5 (33)	—
Good	20.8 (10)	34.4 (32)	—
Excellent	66.7 (32)	17.0 (15)	—
Number years conducting evaluation:			
Less than 1	6.3 (3)	16.1 (14)	—
1–3	16.7 (8)	34.5 (30)	—
3–5	35.4 (17)	24.1 (21)	—
5–10	20.8 (10)	17.2 (15)	—
More than 10	20.8 (10)	8.0 (7)	—

Note: Due to rounding, columns may not add to 100 percent.

Very few evaluators in this study reported subscribing to a particular theoretical framework. Only 10 percent of the sample (*n* = 15) indicated using a particular theory to guide their work. When asked if there was a specific text used to guide evaluation practice—another indicator of theoretical preference—only eight evaluators identified one. The responses to these two items suggest that a very small proportion of evaluators in this study were using an explicit theoretical framework to guide their practice.

Analytic Procedure. Multidimensional scaling (MDS) procedures were used to compare the practice of different theorists and to compare the practice of evaluators to the practice of theorists. The term *multidimensional scaling* refers to a class of methods for spatially representing observed similarities

(or dissimilarities) among stimuli. These methods use proximities among any kind of objects—variables or cases—as input. A *proximity* is a number that indicates the perceived similarity or dissimilarity between any two objects. The main output is a spatial representation, consisting of a geometric configuration of points on a map.

MDS was born from the desire to create distance models as opposed to the linear or vector models implied by methods such as factor analysis (Cox and Cox, 2001). MDS attempts to arrange objects in a space with a particular number of dimensions in order to reproduce the observed distances (distance is a conceptual representation of the empirical association). MDS, then, is a way to “rearrange” objects in an efficient manner to arrive at a configuration that best approximates the observed distances.

Classical multidimensional unfolding (CMDU) is the specific MDS procedure used in the current study. CMDU is an individual-differences analysis that portrays differences in preference, perception, thinking, or behavior and is useful either when examining the differences between subjects in relationship to one another or when investigating the differences between subjects in relationship to stimuli.

When examining individual differences (differences between subjects), the amount of variance explained by the CMDS solution is the key determinant of dimensionality, rather than the stress measure. The procedure yields an R^2 value (between 0 and 1) that indicates the amount of distance between subjects that is accounted for by the mapped dimensions. Thus, the higher the R^2 value, the more variance between subjects is being explained by the map dimensions. The selection of the appropriate number of dimensions is left to the judgment of the researcher based on these values. In the present study, a minimum R^2 value of .85 was established as the requirement for adequate scalability.

The interpretation of dimensions is usually the final step of the analysis. The actual orientations of the axes from the MDS analysis are arbitrary, and can be rotated in any direction. An analytical way of interpreting dimensions (described in Kruskal and Wish, 1978) is to use multiple regression techniques to regress some meaningful variables on the coordinates for the different dimensions. The variable that is thought to have a systematic relationship to a position in the MDS configuration is used as the dependent variable, and the coordinates of the configuration are the independent variables. For a variable to provide an acceptable interpretation of a dimension, the multiple correlation for the variable must be high and the variable must have a high regression weight on that dimension. As a minimum requirement, the multiple correlation should be no smaller than .70 and statistically significant at the .05 level or better. Regression weights should also *ideally* be larger than .70. It should be noted that the sign (negative or positive) of the beta provides an indication of item contrast (but perhaps no other additional information).

Using ALSCAL (Alternating Least-Square SCALing Algorithm), it is possible to use the coordinates derived from a previous MDS analysis from

one set of subjects to determine the dimension coordinates for a new group of subjects (described in Kruskal and Wish, 1978). Kruskal and Wish suggest using this procedure when “the stimulus dimensions are known from prior investigations, and the only interest is to determine the dimension weights for a new group of subjects” (p. 69). This procedure requires that the configuration derived from the first sample be used as the starting configuration for the second analysis and that the iterations be programmed to run for zero iterations. This keeps the stimulus dimension fixed while solving for the dimension weights of the second sample. In other words, the dimension coordinates from the first sample are fixed while solving for the dimension coordinates for the second sample. Thus, the coordinates from the second sample can be interpreted based on the solution derived from the first sample. In this study, the theorists were the fixed sample and the evaluators were the second sample.

Theorists’ Results: Theorists’ MDS Practice Map

Dissimilarity data from the survey were used in an ALSCAL, rectangular, matrix conditional, Euclidean distance model. Each element in the matrix was an integer between zero and ten. The map derived from this multidimensional unfolding analysis served to depict theorists’ practice as dimensions, addressing the study’s first research question: How do the reported practices of evaluation theorists configure empirically?

CMDU Results. The map of evaluation practice produced two dimensions ($R^2 = .928$). A three-dimensional solution was initially hypothesized to be most desirable, because the instrument was constructed to represent Alkin’s three evaluation practice domains. However, the objective of this procedure is to obtain the solution that explains the most variance in the fewest number of dimensions. In the current study, the three-dimension solution ($R^2 = .957$) accounted for only slightly more variance (the amount of distance between theorists that is accounted for by the map dimensions) than the two-dimension solution. Because two dimensions accounted for most of the variance, this solution appeared to represent the primary issues distinguishing theorists most accurately.

CMDU Interpretation. Linear regression analysis was used, in conjunction with a qualitative analysis of the theorists’ stimulus coordinates from the two-dimensional ALSCAL analyses, to derive provisional names for the two ALSCAL dimensions. Thus, the study’s second research question—How can the reported practices of evaluation theorists best be described and compared?—was addressed.

All thirty-eight survey items were regressed on the ALSCAL dimension coordinates. That is, thirty-eight separate regression analyses were conducted, each with one of the thirty-eight survey items as the dependent variable and ALSCAL solution coordinates serving as the two independent variables.

In addition to the statistical analysis, map stimulus coordinates for each theorist for a particular dimension were considered in the dimension interpretation. That is, theorists' coordinate positions on a single dimension, irrespective of their position on the other dimension, were used to help derive meaning about the dimension being interpreted. Table 1.2 provides the map stimulus coordinates for the two dimensions. Theorists were ranked according to the size of the coordinates, from low to high. The magnitude of the coordinates identified those theorists whose views were most dissimilar on each dimension. Figure 1.1 shows a plot of the two dimensions. In addition, raw survey data from the theorists were examined in some cases to ensure accurate dimension interpretation.

This study presents a single interpretation of each dimension based on careful, systematic analysis of study data. Yet it should be noted that the interpretations offered are not necessarily definitive and alternative interpretations are plausible. Nevertheless, for the purposes of the present study, the dimensions have been understood to represent two significant facets of evaluation theory and practice: *scope of stakeholder involvement* and *method proclivity*.

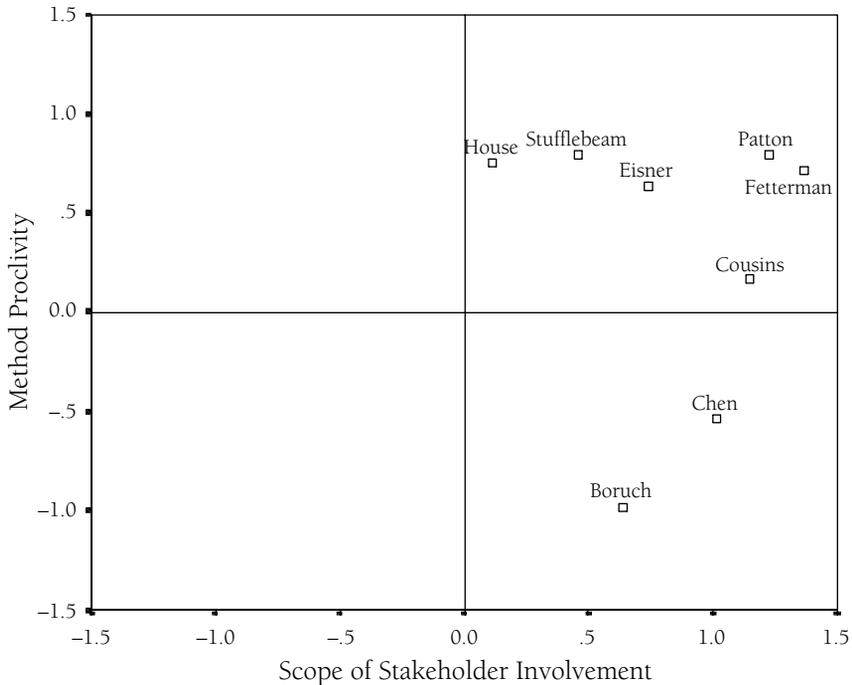
Dimension 1. The interpretation “scope of stakeholder involvement”—that is, the extent to which the evaluation process is directed by stakeholders—was given to the first dimension. This interpretation was based on the linear regression analyses and the positioning of theorists according to their ALSCAL stimulus coordinates. The survey data were examined to help understand better the ALSCAL stimulus coordinates for this dimension.

The name for this dimension was borrowed from Patton (1997), who discusses the “scope of intended user involvement” as one of the six dimensions affecting evaluator involvement and user engagement in the evaluation process. The dimension ranges from users being simply an audience for the evaluation findings to intended users being involved in all aspects of the evaluation from start to finish. Users in the middle of the dimension might, for example, be involved in defining the evaluation's focus but not in the selection of methods or the analysis of data.

Table 1.2. Theorists' Practice Map: ALSCAL Dimensional Stimulus Coordinates

<i>Dimension 1:</i> <i>Scope of Stakeholder Involvement</i>		<i>Dimension 2:</i> <i>Method Proclivity</i>	
House	0.1133	Boruch	-0.9829
Stufflebeam	0.4596	Chen	-0.5358
Boruch	0.6373	Cousins	0.1650
Eisner	0.7406	Eisner	0.6346
Chen	1.0135	Fetterman	0.7144
Cousins	1.1444	House	0.7542
Patton	1.2259	Stufflebeam	0.7934
Fetterman	1.3678	Patton	0.7946

Figure 1.1. Two-Dimensional Map of Theorists' Practice



In the current study, Dimension 1 was interpreted in relationship to stakeholders rather than intended users. According to Patton's theory, *intended users* are a small subset of stakeholders who have a role in program decision making and are in the position to use evaluation findings. This definition of stakeholders is too limited for the purposes of this study because it restricts stakeholders to a very defined group, and it is specific to utilization-focused evaluation theory and therefore is not an inclusive term. Instead, stakeholders were defined in this study simply as "individuals with a vested interest in the outcome of evaluations" (Gold, 1983, p. 64) and could include, but were not limited to, funders, board executives, staff, consumers, and community members.

Results from the regression analyses yielded nine survey items that were used in the interpretation of Dimension 1. These items are shown in Table 1.3. A qualitative examination of the ALSCAL stimulus coordinates also contributed to the interpretation of Dimension 1. As shown in Table 1.2, House, Stufflebeam, Boruch, and Eisner all had stimulus map coordinates between 0 and .8, whereas Chen, Cousins, Patton, and Fetterman had coordinate values greater than 1.0. House and Fetterman established the end points of Dimension 1, suggesting that they are most different in relationship to their views on stakeholder involvement in the evaluation process. It was House's and Fetterman's disparate views that assisted in the

Table 1.3. Items Used in Dimension 1 Interpretation

Item	Beta	R ²
3. When I conduct an evaluation, the primary users (the main people using the evaluation results) help conceptualize and decide upon the evaluation questions.	.799	.749
4. When I conduct an evaluation, if while the evaluation is under way the plan is not working, corrections can be made and new strategies can be identified and implemented.	.666	.804
17. When I conduct an evaluation, primary users (the main people using the evaluation results) judge the appropriateness and quality of the evaluation's methodology.	.837	.865
23. I believe that the evaluator's interpretation of the findings and conclusions can be unbiased.	-.797	.739
25. When I conduct an evaluation, the primary users (the main people using the evaluation results) help interpret the meaning of the evaluation data.	.806	.765
27. My approach as an evaluator is to assess objectively the evaluation's quality against validated technical standards.	-.708	.710
31. I believe that the ultimate purpose of program evaluation is to enhance knowledge and strategies for designing and implementing programs to improve human conditions, not just to assess the effectiveness of programs.	.719	.753
32. I believe that during the final stage of the evaluation, an important role of the evaluator is to work with the primary users (the main people using the evaluation results) to help determine the next steps.	.703	.801
33. My evaluation approach intends to create changes in the culture of the organization (agency) where the evaluation is being conducted.	.755	.707

naming of this dimension. Because Fetterman's position on this dimension is more readily interpretable, I begin by discussing it.

In Fetterman's *empowerment evaluation*, program participants determine the type of evaluation they prefer and their own goals and strategies for accomplishing it. They also determine the type of information they need to document progress toward their goals. Moreover, evaluation is incorporated into the everyday lives of program participants, providing programs with relevant information and thereby improving practice. For this reason, Fetterman's position at the extreme of the scope of stakeholder involvement dimension was best characterized as *stakeholders having a role in all aspects of the evaluation*, from inception to conclusion.

The interpretation of House's position was less obvious. However, after careful thought and analysis of his writings, study data, and regression results, it became clear that he leans toward less stakeholder involvement. I am not suggesting, however, that this precludes stakeholders from his evaluation process, which has as its aim to promote social justice.

House is most concerned with social justice in relationship to program participants (consumers, recipients). He maintains that it is morally correct for evaluators to represent, with their evaluations, the interests and needs of those "unjustly ignored" (1991, p. 234) and he has dedicated most of his career to discussing philosophically the need for evaluations to represent those who do not have a voice in society. Only recently has he offered a framework for evaluation (with Howe), referred to as *deliberative*

democratic evaluation (House and Howe, 2000). This framework rests on the principles of inclusion, deliberation, and dialogue. Stakeholders are involved in a House-type evaluation, but in a particular, democratically deliberative manner. In other words, it would be remiss to suggest that stakeholders do not play an important role in a House evaluation. In the confines of deliberative democratic evaluation, House describes the role of the stakeholder as dialoguing with other stakeholders about genuine interests and deliberating findings and evidence in an effort to ensure that all perspectives are democratically and fairly given voice. He is much less concerned about involving stakeholders in the selection of methods or other evaluation activities. Rather, he focuses his energy on ensuring that social justice, as he defines it, is being served through stakeholder participation. For House, then, stakeholders are involved in circumscribed, but nevertheless important, parts of the evaluation process. Thus, this extreme end of the Dimension 1 continuum can be best characterized as *stakeholders having a role in limited aspects of the evaluation*.

Following, I discuss the position of each theorist on Dimension 1 rather briefly, with the exception of Stufflebeam; I discuss him in more detail because his position could be misconstrued without proper explanation. Stufflebeam refers to his evaluation approach as *decision-oriented*. The primary characteristic of decision-oriented models is that the evaluation is structured by the actual decisions to be made. Using Stufflebeam's approach, the evaluator engages a "representative stakeholder panel to help define the evaluation questions, shape evaluation plans, review draft reports, and disseminate the findings" (2001, p. 57). Stufflebeam suggests engaging the stakeholder panel regularly. He maintains that the success of an evaluation depends on regular interactions with this panel. The evaluator keeps the panel abreast of the formative information produced from the evaluation so that decisions about both the program and the evaluation can be made on a sound basis. Thus, stakeholders are engaged in important evaluation processes, yet they do not actually *participate* in the selection of methods, data collection, or the interpretation of results—these activities are left to the evaluator. With this in mind, Stufflebeam's position on Dimension 1 near the limited stakeholder involvement end is understandable.

Boruch and Eisner, who are positioned rather close to each other on Dimension 1, discuss the importance of stakeholder participation in the evaluation design and in the interpretation of results for the purpose of increasing use. Chen, in his theory-driven evaluation, cautions against using the values of key stakeholders to construct program theory because their views can be biased and subjective (Chen and Rossi, 1980). However, he does support more extensive stakeholder involvement in other aspects of the evaluation process in the interest of decision making and utilization. Next on Dimension 1 are Cousins and Patton, whose views are quite comparable with respect to extent of stakeholder involvement. Both support a breadth and depth of stakeholder involvement, with the goal being

increased use. Although Patton (1997) is very committed to user involvement, based on the premise that evaluations should be judged by their utility and actual use, his position relative to Cousins may seem slightly uncharacteristic because Cousins' approach is actually identified as a participatory model. This finding suggests that the description of an individual's theoretical orientation does not always portray the reality of the situation when examined in relationship to another's practice. In this case, although Cousins' theoretical approach is called participatory, when examined in relationship to Patton, it was found to be less participatory.

Dimension 2. The interpretation method proclivity—that is, the extent to which the evaluation is guided by a prescribed technical approach—was given to the second dimension. The survey items used in the interpretation of Dimension 2 are listed in Table 1.4. As noted there, the items that were found to contribute significantly to the interpretation of the dimension had comparable R^2 values and regression weights. Thus, all items were considered equally when naming the dimension.

As shown in Table 1.4, four items (2, 5, 13, and 14) contributed to one end of the continuum, whereas two items (19 and 21) helped to interpret the other end of this dimension. All related to the methods dimension of the Alkin and House (1992) taxonomy. The items contributed heavily to the overall interpretation and subsequent naming of Dimension 2, not just the anchoring end.

The regression analysis examination of individual items and an examination of the ALSCAL stimulus coordinates assisted in the interpretation of Dimension 2. As noted in Table 1.2, Boruch's and Chen's coordinate values were less than zero whereas those of Cousins, Eisner, Fetterman, House, Stufflebeam, and Patton were between zero and one. Boruch and Patton established the end points of Dimension 2, indicating that they were most different with respect to the extent that their evaluations are guided by a

Table 1.4. Items Used in Dimension 2 Interpretation

<i>Item</i>	<i>Beta</i>	<i>R²</i>
2. When conducting an evaluation, I believe it is critical to understand the main question and then use all scientifically tested instruments available to answer it.	-.839	.725
5. I believe that in order for evaluations to be most effective stakeholders must be included—but do not necessarily have to participate—in the evaluation process.	-.784	.779
13. Although qualitative methods can be useful, my primary method of choice when conducting an evaluation is quantitative.	-.877	.773
14. Using my evaluation approach, the application of research method(s) in an evaluation is most effective when it is guided by a program's conceptual framework, model, or theory.	-.641	.712
19. I believe that evaluation conclusions are mixtures of facts and values.	.874	.842
21. Using my evaluation approach, stakeholders' assumptions about a program are integrated into the evaluation process in order to ensure its relevancy and usefulness.	.669	.810

specified technical approach. It was Boruch's and Patton's disparate views that assisted in the naming of this dimension. Because Boruch's position on this dimension is more readily interpretable, I begin by discussing it.

The experimental research design is at the center of Boruch's evaluation approach. It involves the systematic application of rigorous randomized research designs with the goal of obtaining the least equivocal estimate of a social program's effects (Boruch, Snyder, and DeMoya, 2000). Randomized controlled experimental evaluations have essential, predetermined research steps. One such prerequisite is randomized assignment of participants to one of two or more statistically equivalent treatment groups, with the goal of maximizing the internal validity of the evaluation. This kind of approach is characteristic of method-driven evaluation. Thus, it is evident why Boruch is positioned at one extreme of the method proclivity dimension. This extreme end of Dimension 2 was best characterized as *partiality to a particular methodology that has as a feature predetermined research steps*.

The interpretation of Patton's opposing position was also quite evident and can be summarized by a quote from his text *Utilization-Focused Evaluation*. Patton (1997) asserts: "Utilization-focused evaluation is a problem-solving approach that calls for creative adaptation to changed and changing conditions, as opposed to a technical approach, which attempts to mold and define conditions to fit preconceived models of how things should be done" (p. 131). This stance suggested that the other end of Dimension 2 represents *partiality to framing evaluations by something other than a particular methodology with predetermined research steps*.

A brief discussion of the other theorists' positions is necessary. Chen falls on the negative side of the dimension with Boruch. Evaluations conducted by Chen are driven by the program theory. This evaluation model features predetermined steps, specifically around developing the program theory. And although Chen is a strong proponent of using both quantitative and qualitative methods, an examination of his survey data suggested that he is more inclined to use scientifically tested quantitative methods in his own practice.

Crossing over to the positive side of the dimension, positioned closest to zero was Cousins, a self-described empiricist (personal communication with the author, November 2000). This position on Dimension 2 is consistent with his own perception of his work. Eisner has a strong proclivity for qualitative methods, which, in the context of this dimension, seems to position him toward the center. His position toward the middle, as opposed to the end, of the dimension could be because the items that most definitively defined Dimension 2 were quantitative. This suggests that, in the context of this study, those with a proclivity for quantitative methods held more statistical weight on Dimension 2 than those with a proclivity for qualitative methods. Further, Eisner's position near the center of this dimension indicates that method—in his case qualitative and judgmental—plays at least some important role in his ideas about evaluation practice.

Evaluators' Results: Patterns of Practice

The main section of the instrument asked participants to rate on an eleven-point measurement scale (0 = very dissimilar to 10 = very similar), how similar or dissimilar each statement was to the manner in which they conduct evaluations. These data were used to address the third research question: How does the reported practice of evaluation practitioners compare with the practice of evaluation theorists?

Evaluators' Practice Maps. The decision was made to divide the sample and examine the practice of internal and external evaluators separately. This decision was based on differences in professional characteristics (previously noted) as well as technical issues related to obtaining more meaningful representations in multidimensional scaling.

Internal Evaluators' CMDU Results. Using the technique described by Kruskal and Wish (1978), coordinates were derived for the internal evaluators in relationship to the theorists' two-dimensional ALSCAL solution. The purpose of the analysis was to see how the data for internal evaluators looked when forced into a map configured by the theorists' data. This was accomplished by using the configuration derived from the theorists' analysis as the starting configuration for the evaluators' analysis while programming SPSS to run for zero iterations. Figure 1.2 shows a plot of the internal evaluators in relationship to the theorists' based on the theorists' two-dimensional ALSCAL solution.

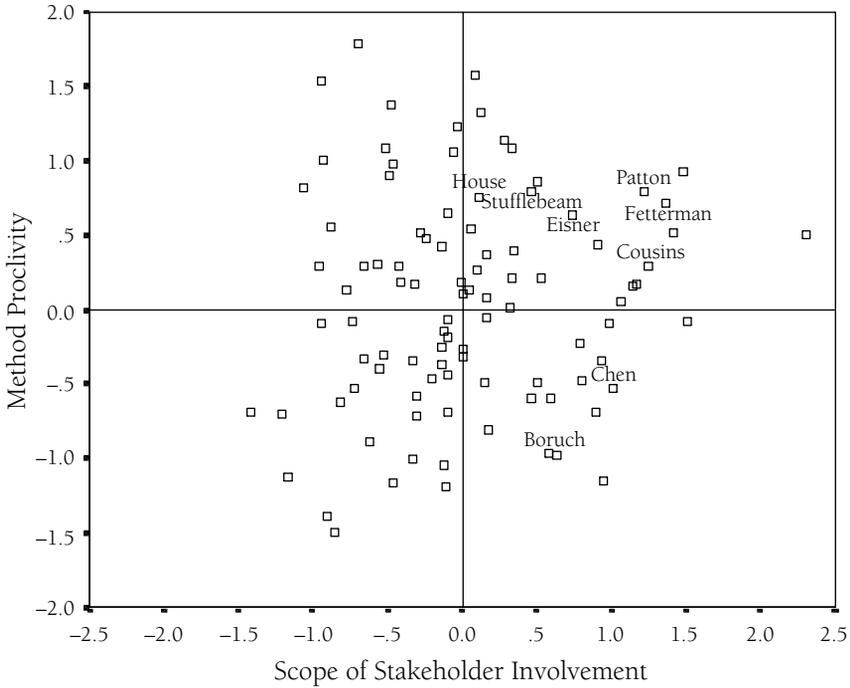
External Evaluators' CMDU Results. As with the internal evaluators, using the technique described by Kruskal and Wish (1978) coordinates were derived for the external evaluators in relationship to the theorists' two-dimensional ALSCAL solution. It is important to note that the "true" dimensional solution for the external evaluators was more like the theorists' than was the internal evaluators' solution. Figure 1.3 shows a plot of the external evaluators in relationship to the theorists, based on the theorists' two-dimensional ALSCAL solution.

Describing the Practice of Evaluators in Relationship to the Practice of Theorists. The practice of Healthy Start evaluators is described in two ways. First, evaluators' placement in relationship to the ALSCAL dimensions derived from the theorists' data is discussed. Then, based on their proximity to one another on the ALSCAL map, the similarities between the evaluation practice of evaluators and theorists is discussed.

Location of Evaluators in Relationship to Map Dimensions. Evaluators were forced into the theorists' two-dimensional ALSCAL solution using the procedures described earlier. This made it possible to examine evaluators in relationship to the theorists' ALSCAL map and the two dimensions: scope of stakeholder involvement, and method proclivity.

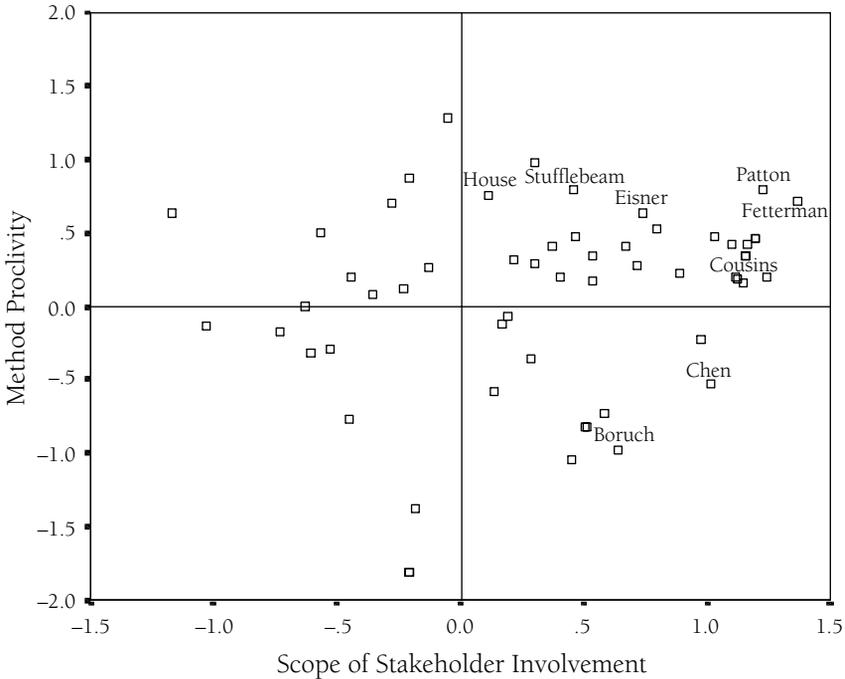
For the purposes of discussing evaluators' practice, the interpretation of Dimension 1 needed expanding. House, who anchored the end of

Figure 1.2. Two-Dimensional Map of Internal Evaluators' Practice in Relationship to Theorists



Dimension 1 closest to the negative side of the x axis, fell slightly off-center to the positive side of the axis. Thus, the interpretation of stakeholders having a role in limited aspects of the evaluation actually corresponds with the zero point, rather than the negative points, of the x axis. This was problematic because, as noted in Figures 1.2 and 1.3, practicing evaluators fell along the negative side of the x axis. To describe these points, the interpretation of Dimension 1 was expanded. Using the expanded interpretation, stakeholder involvement becomes progressively more limited as points move away from zero into the negative end of the x axis. That is, the further to the left of the y axis one was, the less stakeholders were involved in the evaluation process. Those at the extreme negative end of the x axis (around -1.5) perceived the stakeholders as being little more than the audience for the evaluation findings. The interpretation for Dimension 2, method proclivity, did not need expanding. In the context of theorists' data, this dimension was understood to range from partiality to a particular methodology that has as a feature predetermined research steps, to partiality to framing evaluations by something other than a particular methodology with predetermined research steps. Boruch and Patton anchored each end of this dimension.

Figure 1.3. Two-Dimensional Map of External Evaluators' Practice in Relationship to Theorists



Figures 1.2 and 1.3 show the placement of internal and external evaluators in each quadrant of the theorists' derived ALSICAL map. Quadrants I, II, III, and IV were designated left to right, starting in the top row, respectively. The interpretation of each dimension (scope of stakeholder involvement and method proclivity) was used as the basis for understanding the practice of evaluators, and discussed in relationship to each quadrant. This discussion begins with a description of the evaluators' practice in relationship to the x and y axes.

ALSICAL Dimension 1, scope of stakeholder involvement, is represented along the x axis. Therefore, practices related to stakeholder involvement are discussed using the y axis as a reference point. As observed in Figure 1.3, 63.3 percent of the external evaluators were to the right of the y axis (quadrants II and IV). This signifies that, at the very least, they involve stakeholders in limited but important parts of the evaluation process. When examining the internal evaluators with respect to stakeholder involvement (Figure 1.2), less than half (42.7 percent) were to the right of the y axis (quadrants II and IV). This indicates that, generally, stakeholders have a more limited role in evaluations conducted by internal evaluators than in those conducted by external evaluators.

ALSICAL Dimension 2, method proclivity, is represented along the y axis. Therefore, practices related to this dimension are discussed using the

x axis as a reference point. Evaluators below the *x* axis (in quadrants III and IV) are partial to using a particular methodology that features predetermined research steps. Almost half (49.5 percent) of the internal evaluators were below the *x* axis compared with 36.8 percent of the external evaluators, suggesting that internal evaluators are more partial to methodologies with predetermined research steps than are external evaluators. As one moves above the *x* axis (quadrants I and II), evaluators are partial to some other, non-methodological, aspect of evaluation.

This analysis of evaluators in relationship to each dimension provides only a broad depiction of their practice. A more comprehensive understanding of the practice of Healthy Start evaluators is offered when considering the dimensions together. Thus, the evaluators' practice is described next, by way of the four map quadrants. This discussion centers on placement in the four quadrants.

When examining the internal and external evaluators separately, the largest proportion of external evaluators (44.9 percent) were in quadrant II. The remaining external evaluators were distributed evenly among the other three quadrants. Over half of the internal evaluators were in quadrants I and III, with the largest proportion in quadrant III (31.5 percent). Approximately the same proportions of internal (18.0 percent) and external (18.4 percent) evaluators were in quadrant IV.

Six of the eight theorists (House, Stufflebeam, Eisner, Cousins, Fetterman, and Patton), 44.9 percent of the external evaluators, and 24.7 percent of the internal evaluators were in quadrant II. Thus, estimating roughly, almost twice as many external evaluators resembled the practices of the theorists. Generally speaking, evaluators located in quadrant II involve stakeholders in their evaluations to some degree, ranging from limited aspects to all aspects of the evaluation process. In addition, they are partial to some non-methodological element of evaluation.

Boruch and Chen were located in quadrant IV, indicating they are partial to a particular methodology featuring predetermined research steps while also concerned with involving stakeholders in at least some important aspects of the evaluation. Almost the same proportion of external (18.4 percent) and internal (18.0 percent) evaluators were in quadrant IV, denoting, once again very generally, that approximately 18 percent of the total sample reported evaluation practices similar to those of Boruch and Chen. Worth mentioning is the fact that the smallest proportion of internal evaluators fell into this quadrant.

Quadrants I and III were map areas that did not include theorists. Quadrant III had the largest proportion of internal evaluators (31.5 percent). Evaluators in this quadrant are positioned to the left of the *y* axis, meaning that, generally, stakeholders have limited or little involvement in the evaluation process, and below the *x* axis, indicating that they are partial to a particular methodology that has predetermined research steps.

Approximately one-quarter—25.8 percent—of the internal evaluators and 18.4 percent of the external evaluators were in quadrant I. Evaluators in

this quadrant are partial to some non-methodological element of evaluation. Involvement of stakeholders in the evaluation process is limited, and in some cases, stakeholders are likely to be seen as the audience for the evaluation findings rather than hands-on participants.

Proximity of Evaluators to Theorists. Evaluators' positions on the MDS map were compared directly with theorists according to proximity. The proximity of an evaluator to a theorist was determined using a distance formula that calculated the sum of squared deviations of each evaluator's ALSCAL coordinates from each of the eight theorist's ALSCAL coordinates. Eight scores, one corresponding to each of the eight theorists, were calculated for each evaluator. The smaller the score, the more similar the evaluator was to a theorist. An evaluator's proximity to a theorist was considered meaningful if the derived distance score was .2 or less.

As noted earlier, a very small percentage of Healthy Start evaluators reported using an explicit theoretical approach to guide their work. Identifying the practice of evaluators in relationship to that of theorists makes it possible to draw inferences about the types of theories being used. In the current study, theorists' practice is understood in relationship to the two ALSCAL dimensions. Following is a description of evaluators' practice, as it is understood in the context of the theorists' derived ALSCAL dimensions. Evaluators who are within a meaningful proximity of a theorist are considered to have practices similar to that theorist.

Table 1.5 presents frequencies and percentages of internal and external evaluators who were in meaningful proximity of each theorist. A total of fifty evaluators—twenty-five internal and twenty-five external—were within a meaningful range of theorists, configuring seven "theory groups." A theory group is defined by a theorist's coordinates and contains evaluators whose distance scores were .2 or less from a particular theorist. Each theorist, with the exception of Patton, who had no evaluators within a meaningful range of his coordinates, had a theory group. Cousins' group was the largest ($n = 16$), then Boruch ($n = 8$), House ($n = 7$), Stufflebeam ($n = 7$), Chen and Eisner had the same size theory group ($n = 5$), and Fetterman's was the smallest ($n = 2$).

As evidenced in Table 1.5, this study found that a greater proportion of external evaluators (36.2 percent) were in a theory group than internal evaluators (28.1 percent), indicating that the external evaluators' practice was more like that of the theorists. Overall, evaluators were most like Cousins (32.0 percent of the external evaluators in a theory group), particularly the external evaluators (44.0 percent of those in a theory group). After Cousins, external evaluators were most similar to Boruch and Eisner, equally.

These results indicate that most external evaluators (who were within meaningful proximity of a theorist) involve stakeholders in the evaluation process in more than just a limited capacity. However, there were no external evaluators within meaningful proximity of the theorists who involve

Table 1.5. Percent of Evaluators in Meaningful Proximity to Theorists

	Internal Evaluators			External Evaluators			Total Sample		
	n	Percent of n (n = 89)	Percent of Theory Group n (n = 25)	n	Percent of n (n = 49)	Percent of Theory Group n (n = 25)	n	Percent of n (n = 138)	Percent of Theory Group n (n = 50)
Boruch	4	4.49	16	4	8.16	16	8	5.79	16
Chen	4	4.49	16	1	2.04	4	5	3.62	10
Cousins	5	5.61	20	11	22.44	44	16	11.59	32
Eisner	1	1.12	4	4	8.16	16	5	3.62	10
Fetterman	2	2.24	8	0	0	0	2	1.44	4
House	5	5.61	20	2	4.08	8	7	7.24	14
Patton	0	0	0	0	0	0	0	0	0
Stufflebeam	4	4.49	16	3	6.12	12	7	4.34	14
Not in range	64	71.91	—	24	48.09	—	88	63.76	—

Note: Due to rounding, percentages may not add to 100.

stakeholders in almost all aspects of the evaluation (for example, Fetterman and Patton). This suggests that external evaluators in this study involve stakeholders in evaluation processes such as defining the evaluation questions and interpreting the meaning of data and results, but they do not involve stakeholders in a more participatory manner (for example, designing data collection instruments). There is little probability of this group involving stakeholders in the selection of methods because results show that most external evaluators (within meaningful distance of a theorist) are at least moderately partial to using a more methods-driven evaluation approach. This suggests that, although concerned with stakeholder involvement, external evaluators are partial to their methods and conduct evaluations accordingly.

Internal evaluators were distributed rather evenly between the theory groups. More internal evaluators were within proximity to Cousins and House than to any other theorist. Boruch, Chen, and Stufflebeam each had 16 percent of the internal evaluators (within meaningful proximity of an evaluator) in their theory group.

These results reflect the diversity in practices among the internal evaluators who were within meaningful proximity of a theorist. There was no single theorist to whom the internal evaluators were closest. Rather, the twenty-five internal evaluators who were within meaningful proximity of a theorist were divided almost evenly among five of the eight theorists. This implies that the practices of internal evaluators cannot be characterized generally into any one genre of theoretical approaches.

The ALSICAL maps presented earlier displaying the proximities between theorists show that an evaluator could be in very close proximity to more than one theorist; this was the case for eight evaluators. Evaluators were considered very similar to two theorists if the difference in their

distance scores between the two theorists was .05 or less. Although very close to two theory groups, for the purposes of this study these eight evaluators were considered in only one theory group, the group to which they were closest.

Discussion of Theorists' Results

This study's results show that the practices associated with empowerment evaluation (Fetterman) are very similar to those with a utilization focus (Patton and Cousins). This in itself is not surprising, but the finding challenges the conventional belief that Fetterman and House (social justice–deliberative democratic evaluation) are most similar in their approach to evaluation in that both aim to empower through program evaluation. This is not to suggest that either Fetterman or House lacks a social justice agenda. The two theories share the same “big-picture” goal of promoting social justice through evaluation; however, the specific goals of each theory differ. This became evident with the examination of the theorists' practice, which revealed that the extent to which, and more importantly, the ways in which Fetterman and House involve stakeholders in the evaluation process differ.

Nuances Among Theories. The literature on evaluation theory often discusses both Fetterman's and House's theories as social justice or social agenda approaches (Stufflebeam, 2001). Stufflebeam says about social justice approaches that “if—as many persons have stated—information is power, then these approaches employ program evaluation to empower the disenfranchised” (p. 62). One of the main aspects of a social justice evaluation model is to “employ the perspectives of stakeholders as well as experts in characterizing, investigating, and judging programs” (p. 62). Yet the current study found that, with respect to stakeholder involvement, House's and Fetterman's practices are not similar. In fact, House and Fetterman anchor the opposing ends of the stakeholder involvement dimension: House anchored the one end—stakeholders having a role in limited aspects of the evaluation—and Fetterman anchored the other end—stakeholders having a role in all aspects of the evaluation, from inception to conclusion. To understand this finding better, each theory was carefully scrutinized. This examination revealed that Fetterman and House define social justice differently.

House maintains that evaluation should advance the interests of the least privileged in society (1991), whereas Fetterman states that the goal of empowerment evaluation is to foster self-determination (1996). To achieve this goal, Fetterman enlists program participants—including clients—essentially to conduct their own evaluations (1997). House is not as concerned with empowering anyone specifically, but rather with ensuring that those who are often unjustly ignored are given a voice. Thus, House is most interested in individuals who are likely to be program recipients—for example,

homeless youth. For House to pursue a social justice agenda, stakeholder involvement is not required. In fact, House has been critical of stakeholder involvement, arguing that theories that give preferential treatment to stakeholders fail to serve the underrepresented—and hence do not meet his social justice criteria (Alkin, 1990).

In contrast, Fetterman is concerned with empowering the program community, which he defines as program staff and clients. His philosophy of social justice involves teaching people to define program goals, gathering information to help understand if the goals are being met, and using that information in ways that improve the program. To this end, stakeholder involvement is mandatory in almost all aspects of the evaluation. When conducting an empowerment evaluation, the evaluator defines his or her role as the facilitator of the evaluation process, which is determined and conducted by the program community.

Thus, it has become evident that two theories sharing similar big-picture goals do not necessarily share the same theoretical tenets or employ similar practical frameworks. The differences between Fetterman's and House's theories would have been less discernible, however, if we did not have an empirical understanding of how each theorist conducts evaluation with respect to stakeholder involvement.

Results from this study also helped clarify the theoretical positions of Cousins and Patton. Based on the names (Cousins—participatory evaluation—and Patton—utilization-focused evaluation) and technical descriptions of each theory, one might expect Cousins to be higher on user involvement than Patton. But this study found the opposite, revealing an important point: the prescribed practices of a theory are not necessarily best reflected in its name or description.

Cousins' participatory evaluation and Patton's utilization-focused evaluation theories are based on empirical knowledge about the conditions that best facilitate the use of evaluation results—both theories emanate from a concern for fostering utilization. Studies have shown (see Cousins and Leithwood, 1986) that when evaluations are appropriately designed and users are involved and committed to the process, ultimately the results are deemed relevant and utilization of the results is evident.

In the spirit of utilization, the principal tenet of Cousins' theory is stakeholder involvement. Cousins defines participatory evaluation as “applied social research that involves trained evaluation personnel *and* practice-based decision makers working in partnership” (Cousins and Earl, 1995, p. 8). Cousins' aim is to increase user “buy-in” and ownership of the evaluation, which in turn increases the potential that the results will be utilized. Patton also maintains (1997) that utilization increases when users are involved in the evaluation process.

Results from this study confirmed that both Cousins and Patton prioritize user (stakeholder) involvement in practice. However, it was found that Cousins involves stakeholders in the evaluation process slightly less than

Patton does (refer to Figure 1.1). To broaden our understanding of why Cousins was lower than Patton on the stakeholder involvement dimension, we must also consider the interaction between the scope of stakeholder involvement and method proclivity dimensions. Results indicated that Cousins is higher on the method proclivity dimension than Patton (refer to Figure 1.1). This suggests that Cousins uses a more restricted methodological approach, thus limiting the flexibility of the evaluation design, which in turn confines the extent of Cousins' stakeholder involvement. This illustrates how the two dimensions can affect one another and suggests further that titles and descriptions are not as telling when theories are examined more closely and compared with one another. In short, you cannot always judge a book by its title.

Stakeholder Involvement. Regardless of the extent to which theorists discuss stakeholder involvement in their writing, results from this study show that all theorists involve stakeholders in the evaluation process (refer to Figure 1.1). This finding is notable, because stakeholder involvement is formally stated as central to only three of the theorists in this study (Cousins, Fetterman, and Patton). One plausible explanation for this is the process of theory development. By theory development, I am referring to the processes through which one develops and refines theoretical notions.

Alkin (1991) describes six ways in which theorists come to modify their views on evaluation, as well as their written theories. Two of the change influences noted by Alkin help explain why all eight theorists in this study involve stakeholders in the evaluation process: the second, *confrontation with other theories*, and the sixth, *influences of personal interactions*. These change influences are particularly relevant to this finding because the debate both for and against stakeholder involvement has been readily discussed in the evaluation theory literature. Through these change processes, theorists' notions about stakeholder involvement can change, providing a possible explanation for why theorists, no matter their theoretical orientation, involve stakeholders at some level in the evaluation process.

This revelation is interesting, because not all theorists have traditionally been proponents of stakeholder involvement. Alkin, Hofstetter, and Ai (1998) detailed three common concerns about stakeholder involvement: determination of which stakeholders' interests should guide the evaluation, the extent to which the evaluator can adequately represent the interests of stakeholders, and advocacy. The concerns discussed by Alkin and his colleagues have been noted widely in evaluation articles and have met with little or no rejection by the theory community. Yet despite the theoretical concerns related to stakeholder involvement, all of the theorists in this study involve stakeholders, at least minimally, in the evaluation process. I offer as a plausible explanation of this finding that, in practice, the trend has turned toward increased stakeholder involvement, even across a broad theoretical spectrum. Perhaps over time, after reading published material and participating in formal and informal discussions at professional conferences

and meetings with respected colleagues, theorists have compared their own thoughts about stakeholder involvement and begun to practice strategies that they might not have considered previously or that they even formally advocate. Consequently, such actions could lead to changes in their theoretical ideology. Conceivably, however, many individual theorists have chosen not to incorporate changing notions related to stakeholder involvement in their writing because of a common perception that, at this juncture, it is understood to be a part of the evaluation process, no matter one's theoretical approach.

Discussion of Evaluators' Practices

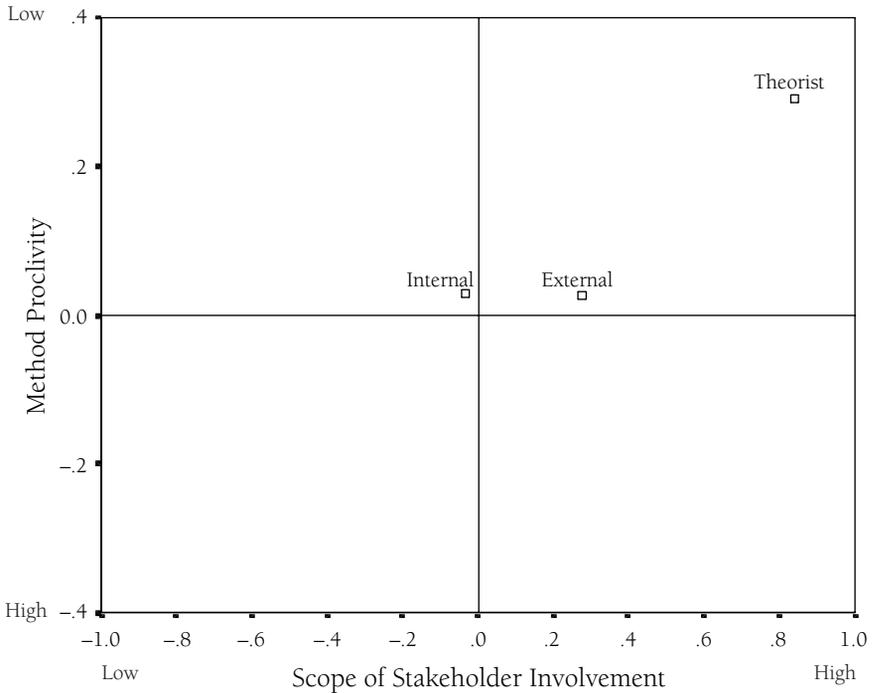
Through the analysis of Healthy Start evaluators, this study offers an understanding of the evaluation practices of the "common evaluator" in the context of evaluation theory. This section will consider two areas: the role of the internal evaluator as a stakeholder in the evaluation and the theoretical frameworks used by the Healthy Start evaluators, as they are understood in the context of theorists' practice, and the implications of these findings.

Evaluators as Stakeholders. Figure 1.4 presents the mean scores for the theorists and practicing evaluators—both internal and external—on the two ALSCAL dimensions, thus reflecting the average positions of each group with respect to scope of stakeholder involvement and partiality to method-driven evaluation frameworks. Results from this study, as noted in the figure, indicate that, overall, practicing evaluators are less focused on stakeholder involvement than theorists. In the group of practicing evaluators, internal evaluators were found to be even less concerned with stakeholders than external evaluators.

Most of the internal evaluators in this sample were program administrators who conduct evaluation as one aspect of their job. Program administrators are themselves key program stakeholders. Because of their intimate relationship to the program, internal evaluators may assume that they understand how other stakeholders think and feel about the program, and hence, do not find it necessary to involve other program associates and participants. For this reason, I speculate that these program administrators—internal evaluators may not consider involving other stakeholders in the evaluation process.

Evaluations are inherently political (Shadish, Cook and Leviton, 1991; Weiss, 1972). That is, the evaluation is set against a political landscape. An evaluator must be sensitive to these politics and consider them when making evaluative choices. The differences between internal and external evaluators offer some insight into how the politics related to being an internal evaluator affect evaluations, and specifically, the tendency of internal evaluators not to involve other stakeholders in the evaluation process. Furthermore, politics may also influence internal evaluators' emphasis on prescribed methods.

Figure 1.4. Mean Scores for Theorists and Healthy Start Evaluators on ALSICAL Dimensions



Note: Scales contracted from Figures 1.2 and 1.3.

Chelimsky (1998) provides an argument, centered around the credibility of the evaluation, that helps explain why the internal evaluators in this study preferred framing evaluations using methods with predetermined research steps and were less focused on stakeholders. She purports that, if evaluation is not impartial, it will not have prestige in public debates. Serving as both a program administrator and evaluator, as the internal evaluators in this study did, a person may have a priori judgments about a program, thus introducing bias into the evaluation study. As a consequence of the internal evaluators' inherently subjective position as employees of the program, they may feel compelled to use more objective methods. Thus, in an effort to maintain objectivity internal evaluators are likely to use more quantitative approaches, because they believe them to increase the credibility of the findings, while diminishing the need for program advocacy on the part of the evaluators—which also increases their credibility.

Strict adherence to somewhat prescribed Healthy Start evaluation guidelines may also explain why the internal evaluators were, on average, found to employ evaluation approaches with predetermined research steps. These guidelines require that each site gather particular information (for

example, number of clients served, health-related outcomes). The less experienced evaluators—in the context of this study, the internal evaluators—may be inclined to adhere strictly to the state evaluation guidelines. Some internal evaluators may view the state guidelines as a formula for practice rather than a guide for understanding the state’s information needs. It is important to note, however, that only one person reported using the state guidelines as a text for guiding evaluation practice. Thus, the propensity of internal evaluators to conduct method-driven evaluations is more likely to be motivated by reducing bias than by following the state guidelines.

Although a greater proportion of internal evaluators showed a proclivity for employing a method-driven framework, the average positions for external and internal evaluators on Dimension 2 (method proclivity) was virtually the same (noted in Figure 1.4). Alkin (1990) argues that evaluators want to avoid professional embarrassment by having their evaluations “held in disrepute by colleagues”; as a result, they conduct evaluations based on “justifiable” or “legitimate” models (p. 165). Conducting evaluations is a full-time job for many of the external evaluators in this study. Thus, to avoid criticism from colleagues that could jeopardize the potential for future work, external evaluators may consider it “safer” to conduct evaluations using models that are perceived to be legitimate. Hence, it can be said that the propensity for both internal and external evaluators to employ method-driven frameworks is influenced by the perception that the results of such studies are more defensible.

Evaluators’ Theoretical Frameworks. An additional point worth mentioning concerns the evaluators’ use of theoretical frameworks. Some theorists (Scriven, 1991; Chelimsky, 1998) argue that theory is “something of a luxury” to an evaluator (Scriven, 1991, p. 360). Results from the current study substantiate this assertion, or perhaps more accurately, show that theory is not requisite to evaluation practice. Specifically, this study found that only 36 percent of the evaluators were within meaningful proximity of a theorist, indicating that most do not use frameworks aligned with a specific theoretical model.

In contrast, Patton claims that after his *Utilization-Focused Evaluation* text was published in 1978 school and human service evaluators (much like the sample of the current study) wrote to him and said about his book . . . “not that they had learned something new, but that they finally had something to call what they did” (Alkin, 1990, p. 143). In other words, Patton’s text helped evaluators legitimize what they were already doing. However, this empirical examination of practice found that evaluators did not necessarily follow a theory in its entirety. Instead, this study found that evaluators adopted select portions of a given theory. Therefore, it is more accurate to say that evaluators use some of the theoretical notions put forth by particular theories rather than fully implement a particular theory. Furthermore, the practices of those who did claim to be using a specific theory did not correspond empirically with the practices of the identified

theorist. Thus, I conclude that the gap between the “common” evaluator and the notions of evaluation practice put forth by academic theorists has yet to be bridged.

References

- Alkin, M. C. *Debates on Evaluation*. Thousand Oaks, Calif.: Sage, 1990.
- Alkin, M. C. “Evaluation Theory Development II.” In M. McLaughlin and D. Phillips (eds.), *Evaluation and Education at Quarter Century*. Chicago: University of Chicago Press, 1991.
- Alkin, M. C., Daillak, R., and White, B. *Using Evaluations: Does Evaluation Make a Difference?* Thousand Oaks, Calif.: Sage, 1979.
- Alkin, M. C., and Ellett, F. S., Jr. “Evaluation Models and Their Development.” In H. J. Walberg and G. D. Haertel (eds.), *International Encyclopedia of Education: Research and Studies*. New York: Pergamon, 1985.
- Alkin, M. C., Hofstetter, C. H., and Ai, X. “Stakeholder Concepts.” In A. J. Reynolds and H. J. Walberg (eds.), *Advances in Educational Productivity*. Greenwich, Conn.: JAI Press, 1998.
- Alkin, M. C., and House, E. “Evaluation of Programs.” In M. C. Alkin (ed.), *Encyclopedia of Educational Research*. New York: Macmillan, 1992.
- Alkin, M. C., Koscecoff, J., Fitzgibbon, C., and Seligman, R. *Evaluation and Decision Making: The Title VII Experience*. Los Angeles: University of California, Center for the Study of Evaluation, 1974.
- Alkin, M. C., and Stecher, B. “Evaluation in Context: Information Use in Elementary School Decision Making.” *Studies in Educational Evaluation*, 1983, 9, 23–32.
- Boruch, R., Snyder, B., and DeMoya, D. “The Importance of Randomized Field Trials.” *Crime and Delinquency*, 2000, 46(2), 156–180.
- Chelimsky, E. “The Role of Experience in Formulating Theories of Evaluation Practice.” *American Journal of Evaluation*, 1998, 19(1), 35–55.
- Chen, H. *Theory-Driven Evaluations*. Thousand Oaks, Calif.: Sage, 1990.
- Chen, H., and Rossi, P. H. “The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science.” *Social Forces*, 1980, 59, 106–122.
- Cousins, J. B., and Earl, L. M. “Participatory Evaluation in Education: What Do We Know? Where Do We Go?” In J. B. Cousins and L. M. Earl (eds.), *Participatory Evaluation in Education*. London: Burgess Science Press, 1995.
- Cousins, J. B., and Earl, L. M. “When the Boat Gets Missed: Response to M. F. Smith.” *American Journal of Evaluation*, 1999, 20(2), 309–317.
- Cousins, J. B., and Leithwood, K. A. “Current Empirical Research on Evaluation Utilization.” *Review of Educational Research*, 1986, 56(3), 331–364.
- Cox, T. F., and Cox M. A. *Multidimensional Scaling*. (2nd ed.) New York: Chapman & Hall, 2001.
- Eisner, E. W. *The Art of Educational Evaluation: A Personal View*. Bristol, Pa.: Falmer Press, 1985.
- Fetterman, D. M. “Empowerment Evaluation: An Introduction to Theory and Practice.” In D. M. Fetterman, S. J. Kaffarian, and A. Wandersman (eds.), *Empowerment Evaluation*. Thousand Oaks, Calif.: Sage, 1996.
- Fetterman, D. M. “Empowerment Evaluation and Accreditation in Higher Education.” In E. Chelimsky and W. Shadish (eds.), *Evaluation for the 21st Century*. Thousand Oaks, Calif.: Sage, 1997.
- Gold, N. “Stakeholders and Program Evaluation.” *New Directions for Program Evaluation*, 1983, 17, 63–72.
- House, E. “Evaluation and Social Justice. Where Are We?” In M. McLaughlin and D. Phillips (eds.), *Evaluation and Education at Quarter Century*. Chicago: University of Chicago Press, 1991.

- House, E., and Howe, K. R. (eds.). *Deliberative Democratic Evaluation*. New Directions for Program Evaluation, no. 85. San Francisco: Jossey-Bass, 2000.
- King, J., and Pechman, E. *Improving Evaluation Use in Local Schools*. Washington, D.C.: National Institute of Education, 1982.
- Kruskal, J. B., and Wish, M. *Multidimensional Scaling*. Thousand Oaks, Calif.: Sage, 1978.
- Patton, M. Q. *Utilization-Focused Evaluation: A New Century Text*. (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.
- Patton, M. Q., and others. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research." In C. H. Weiss (ed.), *Using Social Research in Public Policy Making*. Lexington, Mass.: Heath, 1977.
- Pressman, J. L., and Wildavsky, A. *Implementation: How Great Expectations in Washington Are Dashed in Oakland*. Berkeley: University of California Press, 1973.
- Scriven, M. *Evaluation Thesaurus*. (4th ed.) Thousand Oaks, Calif.: Sage, 1991.
- Shadish, W. R. "Evaluation Theory Is Who We Are." *American Journal of Evaluation*, 1998, 19(1), 1–91.
- Shadish, W. R., Jr., Cook, T. D., and Leviton, L.C.L. *Foundations of Program Evaluation*. Thousand Oaks, Calif.: Sage, 1991.
- Shadish, W. R., and Epstein, R. "Patterns of Program Evaluation Practice Among Members of the Evaluation Research Society and Evaluation Network." *Evaluation Review*, 1987, 11(5), 555–590.
- Smith, N. L. "Citizen Involvement in Evaluation: Empirical Studies." *Studies in Educational Evaluation*, 1983, 9, 105–117.
- Smith, N. L. "Improving Evaluation Theory Through the Empirical Study of Evaluation Practice." *Evaluation Practice*, 1993, 14(3), 237–242.
- Stufflebeam, D. L. "The CIPP Model for Program Evaluation." In G. F. Madaus, M. S. Scriven, and D. L. Stufflebeam (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Norwell, Mass.: Kluwer-Nijhoff, 1983.
- Stufflebeam, D. L. (ed.). *Evaluation Models*. New Directions for Program Evaluation, no. 89. San Francisco: Jossey-Bass, 2001.
- Stufflebeam, D. L., and Shinkfield, A. J. *Systematic Evaluation*. Norwell, Mass.: Kluwer-Nijhoff, 1985.
- Weiss, C. "Utilization of Evaluation: Toward Comparative Study." In C. Weiss (ed.), *Evaluating Action Programs*. Needham Heights, Mass.: Allyn & Bacon, 1972.
- Wholey, J. S., and others. *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Washington, D.C.: Urban Institute, 1970.
- Williams, J. E. "Mapping Evaluation Theory: A Numerically Developed Approach." *Dissertation Abstracts International*, 49(9), AAT 8825488, 1988.
- Worthen, B. R. "Program Evaluation." In H. J. Walberg and G. D. Haertel (eds.), *International Encyclopedia of Educational Evaluation*. New York: Pergamon, 1990.
- Worthen, B. R., and Sanders, J. R. "Evaluation as Disciplined Inquiry." In B. Worthen and J. Sanders (eds.), *Educational Evaluation: Theory and Practice*. Worthington, Ohio: Charles A. Jones, 1973.

CHRISTINA A. CHRISTIE is an assistant professor and the associate director of the Institute of Organizational and Program Evaluation Research at Claremont Graduate University, School of Behavioral and Organizational Sciences. She is chair of the Theories of Evaluation Division of the American Evaluation Association.

Evaluators who wish to integrate theory-based approaches into their practices are encouraged to pursue additional readings (including those referenced in this document) and, as appropriate, to seek additional support in undertaking a theory-based evaluation.

Acknowledgements. Theory-based approaches to evaluation attempt to understand an intervention's contribution to observed results through a mechanistic or process interpretation of causation, rather than determining causation through comparison to a counterfactual.

Step 2: The evaluator maps out a series of potential mini-theories that relate the various contexts of a program to the multiple mechanisms by which it might operate to produce different outcomes. What is project evaluation, how to evaluate projects, what indicators to consider, and how to design a project evaluation plan.

General guidelines. As a process, project evaluation takes a series of steps to identify and measure the outcomes and impacts resulted from project completion. In this article, let's find out how to evaluate projects, what indicators to consider, and how to design a project evaluation plan.

Indicators. A Project Evaluation Plan is a detailed document that defines and sets forth practices and sequence of activities for analyzing and examining the project by certain evaluation criteria. This document aims to determine project effectiveness and efficiency through tracking progress on each objective, completion of activities, and dates of completion. An impact evaluation approach that iteratively maps available evidence against a theory of change, then identifies and addresses challenges to causal inference. Click an approach on the left to navigate to it.

Critical System Heuristics. Horizontal Evaluation. A particular type of case study used to jointly develop an agreed narrative of how an innovation was developed, including key contributors and processes, to inform future innovation efforts. Click an approach on the left to navigate to it.

Innovation History. A particular type of case study used to create a narrative of how institutional arrangements have evolved over time and have created and contributed to more effective ways to achieve project or program goals. Click an approach on the left to navigate to it.

Most Significant Change. Oftentimes, evaluators focus on the technical details of the evaluation rather than the overall purpose. Focusing on theory at the onset of a project ensures the process (i.e. stakeholder involvement, methodology, data collection, analysis, reporting) is intentional, purposeful, and more useful for the client. This article seeks to bridge the gap between evaluation theory and practice by outlining four theories and contexts in which they are relevant. I also provide resources for implementing each approach.

Utilization-Focused Evaluation Theory. Michael Quinn Patton, PhD, developed Utilization