

THE LANGUAGE OF EVALUATION THEORY: INSIGHTS GAINED FROM AN EMPIRICAL STUDY OF EVALUATION THEORY AND PRACTICE

Christina A. Christie
School of Behavioral and Organizational Sciences
Claremont Graduate University
Claremont, CA, USA

Mike Rose
Graduate School of Education and Information Studies
University of California, Los Angeles

Abstract: Broad concern for language issues in evaluation has been limited in comparison to other social science disciplines. In this article, some occasions of definitional or conceptual confusion with evaluation theory language are identified that emerged during a study conducted by Christie. We suggest that much of the language we use to describe evaluation practice is steeped in theoretical terminology, which may limit the utility of the language. We also argue that theoretical language ought to be used with great care, with attention to the subtleties and nuances of terms, for there may be unexpected confusion or ambiguity in the field about the terms we routinely use. A research agenda is offered, suggesting that it would be both an informative as well as a useful task for us to learn more about the everyday “folk theories” of the field and the vernacular used to describe them.

Résumé: Une préoccupation généralisée quant aux questions de langage dans le domaine de l'évaluation a été limitée comparativement à d'autres disciplines des sciences sociales. Dans cet article, nous présentons certaines possibilités de confusion dans les définitions ou les concepts, occasionnées par le langage de la théorie de l'évaluation et repérées dans le cadre d'une étude menée par Christie. Nous suggérons qu'une bonne partie du langage que nous utilisons pour décrire la pratique de l'évaluation est profondément ancrée dans la terminologie théorique, ce qui peut en limiter l'utilité. Nous avançons aussi que le langage théorique devrait être utilisé avec beaucoup de précaution, en fai-

Corresponding author: Christina A. Christie, Claremont Graduate University,
School of Behavioral and Organizational Sciences, 123 East Eighth Street,
Claremont, CA 91711, USA; <tina.christie@cgu.edu>

sant attention aux subtilités et aux nuances des termes, parce qu'il peut y avoir de la confusion ou de l'ambiguïté dans le domaine en ce qui a trait aux termes que nous utilisons de façon systématique. Nous offrons un plan de recherche et nous suggérons qu'il serait utile et avantageux pour nous d'en apprendre davantage sur les «théories populaires» courantes du domaine et le vernaculaire utilisé pour les décrire.

INTRODUCTION

While language issues are well considered in social science research, broad concern for these issues in evaluation has been more limited (Hopson, 2000). The most common discussions of language in evaluation focus primarily on issues related to the difficulties clients have with the specific terminology in which evaluation questions, data, and results are cast. This is best illustrated by the work of senior scholars (e.g., Fetterman, 1988; Patton, 1997) whose writing has made evaluators keenly aware of the link between accessible language in evaluation reports and the utilization of those reports.

More recent discussions have focused on the manner by which language can shape meaning in social policy and programs (Hopson, Lucas, & Peterson, 2000). For example, Madison (2000) considers the sociopolitical nature of language in evaluation and the role of translator played by evaluators when considering evaluation outcomes. The translation of program findings into policy language and subsequent decision-making was explored by Cabatoff (2000). And Kaminsky (2000) discussed how examining metaphors in the evaluation language of individuals and groups offers insight into the ways people frame and resolve problems.

The complexity of the implications of definitional problems in academic theory is suggested in the limited literature on evaluation language. In this article we explore, via close analysis of evaluation texts, the issue of language in evaluation theories and how that language may impact the utility of those theories. We believe that our field could benefit from a more explicit consideration of the language we use. To illustrate our thesis, we identify particular instances of definitional or conceptual confusion that occurred during an empirical study of evaluation practice. We argue that further research specifically examining evaluation language is needed to refine our evaluation theories and untangle the interplay of the complex and

dynamic set of factors associated with the utility of evaluation language and its impact on evaluation practice.

THE ROLE OF LANGUAGE IN AN EMPIRICAL STUDY OF EVALUATION

A Brief Description of the Empirical Study

A recent study by Christie (2003), which empirically examined the translation of evaluation theory into practice, led us to wonder if there were linguistic issues in our field that are worthy of reflection. The primary objective of the study was to develop an empirically derived comparative framework of the reported practices of evaluation theorists and to assess and describe within this framework the “real world” practices of people in the field. The main data collection instrument used in this study was constructed in concert with eight distinguished theorists,¹ representing a broad range of perspectives. Theorists were asked to submit items corresponding to three dimensions of evaluation practice (methods, values, and use) that they would use to help determine whether or not an evaluator was utilizing their theories to guide his or her work. The instrument was then developed from these items. Finally, each of the eight theorists themselves completed the instrument, reporting on their own practices. Data were analyzed, using multidimensional scaling procedures, to ascertain and describe these practices.²

Two main constructs emerged from the analysis of distinguishing theorists' practice: *scope of stakeholder involvement* and *method proclivity*. *Scope of stakeholder involvement* depicts the extent to which stakeholders direct the process of an evaluation. Thus, for the theorists in the study, one extreme represents the involvement of stakeholders in all aspects of the evaluation, while the other extreme characterizes practices that involve stakeholders in limited, yet important, aspects of the evaluation. The *method proclivity* dimension depicts the extent to which an evaluation is guided by a technical, method-driven approach. Therefore, one extreme reflects the use of a design/methodology that has as a central feature predetermined research steps (e.g., an experimental design), while those at the other extreme are partial to framing evaluations by something other than a method-driven approach. An example of someone at this extreme is Patton who asserts, “utilization-focused evaluation is a problem-solving approach that calls for creative adaptation to changed and changing conditions, as opposed to a tech-

nical approach, which attempts to mold and define conditions to fit preconceived models of how things should be done” (1997, p. 131).

Using the same instrument, data were collected from 138 practicing evaluators in the field. These data were compared to theorists’ data, in relationship to the *scope of stakeholder involvement* and *method proclivity* dimensions. This analysis offers an empirically derived understanding of evaluators’ practices.

How and Where Issues of Language Emerged

Text analysis (see, for example, Lanham, 1983) provided the linguistic lens through which we examined the issues of language that emerged in Christie’s research. Such analysis encourages a close consideration not only of the conceptual content of a text, but also of the text’s diction, grammar, and rhetorical structure. We were interested in the style of the message as well as the message itself. Using this framework for analysis, language issues emerged repeatedly, in several ways, related to the terms, concepts, and associated meanings that help set the frame for evaluation. First, we found that much of the language we use to describe evaluation practice is steeped in theoretical terminology, which may limit the practical utility of the language. Second, we found that theoretical language ought to be used with great care, with attention to the subtleties and nuances of terms, for there may be unexpected confusion or ambiguity in the field about the terms we routinely use.

The issue of theoretical language first emerged when designing the study’s main data collection instrument. As mentioned, theorists were asked to submit items from which an instrument was developed. Specifically, they were asked to submit up to nine items related to three dimensions of evaluation theory, items that one could respond to using an 11-point scale ranging from “this is very dissimilar to the way I conduct evaluation” to “this is very similar to the way I conduct evaluation.” Some theorists submitted lengthy sentences, even paragraphs, to demonstrate the practical application associated with their theoretical standpoint — and many of the statements were rich with abstract language. The charge then became one of revising the statements so that the language was more accessible and the intent of the statement clear to practitioners, while always retaining the essence of the theorist’s original statement. Extreme caution was taken when revising the statements, and an interesting thing happened during the revisions. This close attention to semantics, to the specifics of

theorists' meaning, led us back to the full texts of their theoretical publications, reading them again, but in a new and revealing light — with a sharper eye toward application.

To illustrate this process, we briefly describe how one easily revisable statement was reworked. The submitted statement read, "Evaluations should be used for instrumental and enlightenment purposes." The first step in revising this statement was to examine this particular theorist's publications in order to determine his precise definition of "instrumental" and "enlightenment" use. Once the meanings of these terms were established within the context of the theorist's work, the statement was worded so that a person who might not have a working evaluation vocabulary could understand what was meant by instrumental and enlightenment use. Then the statement was further amended to fit the response scale. The final item read, "I believe that evaluations are to be used not only to inform program decision-making but also to build upon the current generalized knowledge base of the particular program/subject being studied."

Other statements submitted were written using more colloquial evaluation theory language, that is, they were written using terms or phrases that are familiar to those who regularly discuss evaluation but may be less familiar to those who are not part of the academic theory/research world. Take, for example, the following statement submitted by one theorist: "I work hard to understand what the empirical question is that needs to be answered and then take off the shelf whatever tools I need to answer it as best I can." This statement was revised to make certain that practitioners could understand the meaning of phrases such as "take off the shelf whatever tools" as well as the intent behind the focus on an "empirical question." The revision process was identical to the one previously described: examining the theorist's written texts to ensure proper translation of specific words and expressions and then reworking the statement into a questionnaire item. For this example, the final item read, "When conducting an evaluation, I believe it is critical to understand the main evaluation question and then use all scientifically tested instruments available to answer it."

The process of using the published work of a theorist to clarify his own theoretical statements caused us to speculate about the language of theory itself. While it is true that there is nothing more useful than a good theory, the theorist is under no real professional or ethical obligation to write theory with a direct eye toward prac-

tice. This process of reading and summarizing/paraphrasing and subsequent rereading revealed to us, in a way we had not quite seen before, some of the conceptual density and abstraction of the language of theory. Of course, all professional languages exhibit such attributes, but one might ask whether evaluation theory should be written differently, since by its very definition, evaluation is a practical enterprise, intended for application.

The issue of theoretical language emerged again in the Christie study when comparing different theorists' practice. The main analytic procedure used to analyze theorists' data was multidimensional scaling, a class of methods for spatially representing observed similarities (or dissimilarities) among objects (variables or cases). The main output is a spatial representation, consisting of a geometric configuration of points on a map. The interpretation of dimensions is a very important part of the analysis because the axes yielded are, in themselves, meaningless. One analytical method for interpreting dimensions (described in Kruskal & Wish, 1978) is to use multiple regression techniques to regress some meaningful variables on the coordinates for the different dimensions. To help derive meaning about the dimension being interpreted, Christie considered the map stimulus coordinates (the position on each axis of the two-dimensional map) in addition to the statistical analyses for each theorist.

This qualitative analysis of the quantitative positions held by the theorists in relation to one another directed us again to reexamine each theorist's writing. It was the careful reading of theorists' text that ultimately led to the discernment and interpretation of each dimension. Text analysis was particularly important, for example, when interpreting the dimension *scope of stakeholder involvement*. This dimension represents the extent to which theorists involve stakeholders in the evaluation process. This would seem to be a fairly straightforward construct. Yet, in examining the positions of two theorists — David Fetterman and Ernest House — we saw complication.

Examining the Term "Social Justice" as an Example of Linguistic Distinctions

Both Fetterman and House advocate for social justice in evaluation, that is, conducting evaluations that are directed toward making a difference in society (Stufflebeam, 2001). Since they share this common goal, conventional thought leads one to hypothesize that, in

practice, each would involve stakeholders in the evaluation process similarly. Yet, the analysis revealed House and Fetterman establishing either end point of the derived stakeholder involvement dimension, suggesting that they are most different in relationship to their views on stakeholder involvement. So back we went to the theorists' publications, in search of a meaningful interpretation of the data. After careful study of their work, House and Fetterman's disparate views were clarified. The clarification offered by Christie was corroborated by both Fetterman (2003) and House (2003). This new understanding made it possible to identify issues related to some of our conventional understandings of theoretical language.

Stufflebeam (2001) suggests — and conventional evaluation thought supports the notion — that one of the main orientations 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). How then could it be that House anchored one end of the *scope of stakeholder involvement* dimension, representing limited stakeholder participation, and Fetterman the other end, characterized by stakeholders having a role in all aspects of the evaluation, from inception to conclusion? In this case, once again, the data brought us back to the texts of these two theorists, with particular attention being given to the use and definition of the term “social justice.”

This examination revealed that Fetterman and House define social justice differently. In Fetterman's *empowerment evaluation*, the program community, which includes participants at all levels of the organization, determines 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 the program community, providing participants with relevant information and thereby improving practice. Thus, Fetterman regards social justice as empowering those in the program to help themselves, and he pursues this agenda by developing evaluations for them that foster self-determination, enlightenment, and liberation for pre-existing roles and formats (Fetterman, 2000). For Fetterman, knowledge equals power. House, on the other hand, focuses his attention on groups that are typically underrepresented in the evaluation process. Most often this group includes the program's participants (clients, consumers, recipients). He argues that it is morally correct for evaluators to represent, within their evaluations,

the interests and needs of those “unjustly ignored” (House, 1991) and has dedicated most of his career to discussing, philosophically, the need for evaluations to represent those who do not have a voice in society. Historically, House has expressed concern about stakeholder involvement, suggesting that in particular stakeholder processes, the agenda of those in power prevails over those of minority groups, the poor, and the powerless (House, 1993). House has also been critical of approaches that tailor the evaluation to the interests of smaller stakeholder groups (usually called primary intended users) because of their failure to serve underrepresented populations (Alkin, 1990).

House’s and Fetterman’s distinct definitions of social justice in evaluation have always existed. What we are suggesting, however, is that our conventional understanding of commonplace terms such as *social justice* can lead to erroneous assumptions about a theoretical position. House’s and Fetterman’s individual definitions of social justice, when translated into practice, prescribe different procedures. Without the initial identification of their disparate positions with respect to practice (i.e., *stakeholder involvement*), the necessity to distinguish their differing definitions of social justice may not have surfaced.

To understand House’s and Fetterman’s positions related to *stakeholder involvement*, their differing definitions of social justice had to first be recognized. The process of unearthing each theorist’s notions of social justice involved an initial move away from the findings of the Christie study to the theorists’ writings. Careful analysis of the texts led us back to the question items and the raw data, to determine which items distinguished Fetterman’s and House’s positions. Completing the cycle, the analysis of the raw data brought us back, once again, to their written theoretical work. In returning to the text after the intermediate steps, a more precise interpretation of the study’s findings was possible. This process forced us to reconsider some of commonplace terms of evaluation theory and revealed the degree to which social researchers and evaluators use a set of common terms and assume that we all apply roughly the same meaning to them.

To be sure, such blending and confusing of terms is part of language use, professional or otherwise. But, again, given the applied nature of evaluation, given that our terms make things happen in the world, the situation that presented itself in the Christie study does lead us to wonder if there shouldn’t be more explicit discussion in our field

about the language we use when we write evaluation theory. Do we need to think about our language through the lens of the philosophy of language or of communication theory? Theorists could guide such efforts by providing some thought as to how one might construct theoretical language that is more readily and precisely translated into practice.

Can language provide insights into practice?

A final and related issue, certainly of immense practical importance, is the large number of people in the field who conduct evaluations without having been properly or sufficiently trained in evaluation theory at all. This was one of the findings of Christie's study: only 10% of the sample of practicing evaluators ($n = 138$) indicated the use of a particular theory to guide their work. When asked if there was a specific text used to direct their evaluation practice (another indicator of theoretical preference), less than 6% endorsed such a text. The implication is that a very small proportion of evaluators use an explicit theoretical framework to guide their practice. These findings certainly suggest that the profession needs to provide more and enhanced theoretical training for practicing evaluators. Yet the direction of this article leads us to wonder: even if all practitioners suddenly, miraculously, received this improved training, would we still face the problem of having people misinterpret or misunderstand our theories? We talk about utilization of evaluation findings, but ought we also be concerned with the way our language impedes or enhances the use of theory?

Human beings do not act without some kind of belief or assumption about what they are doing, which in essence is an implicit theory. Thus, it is interesting to speculate as to what kinds of implicit or folk theories of evaluation exist in the field, and how everyday practitioners form the notions about evaluation that they use to guide their work. We wonder what "on the ground" evaluation theories people use, and we are curious about the vernacular language in the field. Findings from Christie, indicating that the statistical fit between theorists' and everyday evaluators' practice is weak, suggest further that the approaches used by everyday evaluators are not easily aligned with the constructs that we assume to be our conceptual terrain.

It seems likely that everyday practitioners acquire their theories from one or more of the following sources: from reading or study of

evaluation theory, either in or outside of a formal academic setting; from conferences, seminars, workshops; through informal professional networks; from ideas “in the air”; and from the explicit or implicit notions about evaluation contained in the mandates or documents received from superiors. These various sources lead to what we suspect to be a range of folk theories of evaluation that drive considerable real-world practice. We know very little about these folk theories and the language that is adopted to express them. We argue that it would be both an informative as well as a useful task for us to learn more about these everyday theories and languages. In-depth interviews with evaluators, asking them to describe their evaluation practices, could begin to ascertain and understand some of evaluation’s folk theories. We suspect they inform practice — perhaps, in some settings, more than our own published/established theoretical constructs and languages. Furthermore, such inquiry would help us to think more profoundly about the definitional questions that we raised earlier and would help us to think as well about the construction of theoretical languages.

CONCLUSION

In this article we have tried to raise questions about the possible utility of the theoretical language we use and pinpointed some occasions of definitional or conceptual confusion. Such issues might well play into the investigation of folk theoretic language of practice in enlightening ways. For example, is the language used in folk theory more accessible than the language of our academic theory? Furthermore, understanding that there are definitional problems in academic theory, we wonder what kind of definitional confusion or overlap one might find in a diverse field of practice. A focused study of the language of evaluation might help us to answer all these lines of inquiry. And that would have both theoretical and practical benefit.

NOTES

1. The eight theorists that participated in Christie’s study were Robert Boruch, Huey-tysh Chen, J. Bradley Cousins, Elliot Eisner, David Fetterman, Ernest House, Michael Patton, and Daniel Stufflebeam.
2. Classical multidimensional unfolding (CMDU) was the specific MDS procedure used in this study. CMDU is an individual differences analysis that portrays difference in preference, perception, think-

ing, or behaviour and is useful when either examining the differences between subjects in relationship to one another or when investigating the differences between subjects in relationship to stimuli. The main output is a spatial representation, consisting of a geometric configuration of points on a map. 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. The map of evaluation theorists' practice produced two dimensions ($R^2 = .928$). Using ALSCAL (Alternating Least-square SCALing Algorithm), the coordinates derived from a previous MDS analysis on one set of subjects can be used to determine the dimension coordinates for a new group of subjects. The dimension coordinates from the first sample are "fixed" while solving for the dimension coordinates for the second sample, allowing for the coordinates from the second sample to be interpreted based on the solution derived from the first sample. This procedure was used to compare theorists' practice (the fixed sample) and the practice of Healthy Start evaluators (the second sample). Multiple regression techniques are used to interpret dimensions (for a complete description of multidimensional scaling and related analyses, such as using multiple regression for dimension interpretation, see Kruskal & Wish, 1978).

REFERENCES

- Alkin, M.C. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.
- Cabatoff, K. (2000). Sociolinguistic dynamics of gender in focus groups. In R. Hopson (Ed.), *How and why language matters in evaluation*. (New Directions for Program Evaluation, No. 86, pp. 43–54). San Francisco, CA: Jossey-Bass.,
- Christie, C.A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. In C.A. Christie (Ed.), *The practice-theory relationship*. (New Directions for Program Evaluation, No. 97) (pp. 7–36). San Francisco, CA: Jossey-Bass.
- Fetterman, D.M. (2003). Fetterman-House: A process use distinction and a theory. In C.A. Christie (Ed.), *The practice-theory relationship*. (New Directions for Program Evaluation, No. 97) (pp. 47–52). San Francisco, CA: Jossey-Bass.
- Fetterman, D.M. (2000). *Foundations of empowerment evaluation*. Newbury Park, CA: Sage.

- Fetterman, D.M. (1988). Ethnographic educational evaluation. In D.M. Fetterman (Ed.), *Qualitative approaches to evaluation in education* (pp. 116–137). New York: Praeger.
- Hopson, R.K. (Ed.). (2000). *How and why language matters in evaluation*. (New Directions for Program Evaluation, No. 86). San Francisco, CA: Jossey-Bass.
- Hopson, R.K., Lucas, K.J., & Peterson, A.J. (2000). Translating evaluation findings into “policy language.” In R. Hopson (Ed.), *How and why language matters in evaluation*. (New Directions for Program Evaluation, No. 86, pp. 29–42). San Francisco, CA: Jossey-Bass.
- House, E. (1991). Evaluation and social justice. Where are we? In M. McLaughlin & D. Phillips (Eds.), *Evaluation and education at quarter century* (pp. 233–247). Chicago: University of Chicago Press.
- House, E. (1993). *Professional evaluation*. Newbury Park, CA: Sage.
- House, E. (2003). Stakeholder bias. In C.A. Christie (Ed.), *The practice-theory relationship*. (New Directions for Program Evaluation, No. 97) (pp. 53–56). San Francisco, CA: Jossey-Bass.
- Kaminsky, A. (2000). Beyond the literal: Metaphors and why they matter. In R. Hopson (Ed.), *How and why language matters in evaluation*. (New Directions for Program Evaluation, No. 86, pp. 69–80). San Francisco, CA: Jossey-Bass.
- Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.
- Lanham, R. (1983). *Analyzing prose*. New York: Scribners.
- Madison, A. (2000). Language in defining social problems and in evaluating social programs. In R. Hopson (Ed.), *How and why language matters in evaluation*. (New Directions for Program Evaluation, No. 86, pp. 17–28). San Francisco, CA: Jossey-Bass.
- Patton, M.Q. (1997). *Utilization-focused evaluation: A new century text*. Thousand Oaks, CA: Sage.
- Stufflebeam, D. (2001). *Evaluation models*. (New Directions for Program Evaluation, No. 89). San Francisco, CA: Jossey-Bass.

Christina A. Christie is an assistant professor and the associate director of the Institute of Organizational and Program Evaluation Research at Claremont Graduate University. She is chair of the Theories of Evaluation Division of the American Evaluation Association and a founder of the Southern California Evaluation Association.

Mike Rose is a professor in the Graduate School of Education and Information Studies at the University of California, Los Angeles, specializing in Social Research Methodology.

