CHAPTER 9

What Is the Difference between Evaluation and Research—and Why Do We Care?

SANDRA MATHISON

Offering a definition of evaluation as the process and product of making judgments about the value, merit, or worth of an evaluand does little to answer the perennial question: What is the difference between evaluation and research? Annually, this question is asked on EVALTALK, an international discussion listserv for evaluators. A quick search of the archives illustrates that each year since 1998 there has been a substantial thread in response to this question, posed often by novice evaluators. Indeed, in 1998 there were about 178 messages in response to the question, and the thread was sustained by the contributions of three prominent figures in evaluation: Michael Scriven, Michael Quinn Patton, and William Trochim. Essentially, the two Michaels asserted there is a difference between research and evaluation but that the two overlap, and Trochim argued that evaluation is no different than applied social science. In subsequent years, the threads have been shorter, but the question has not gone away.

This question about differences between evaluation and research is fueled by the fact that evaluation as a discipline draws on other disciplines for its foundations, and especially the social sciences for its methods. As evaluation has matured as a discipline and profession, this question is sometimes posed to clarify what is distinct about evaluation. This

delineation of a profession of evaluation is also tied to a discussion of who is and can be an evaluator. What knowledge and skills does evaluation require, and how do these differ from the knowledge and skills of social science researchers?

This chapter does two things. First, the chapter considers why this question is fundamental, including how the question relates to the development of evaluation as a discipline, a professional practice, and what knowledge and skills evaluators need. Second, I discuss the difference between research and evaluation. This section of the chapter draws on the evaluation literature, including the extensive discussions on EVALTALK. It is often the case that both evaluation and research are caricatured in order to make distinctions between the two. In the search for a simple way of distinguishing research and evaluation, similarities and differences are sometimes masked.

WHY DO WE CARE ABOUT THE DIFFERENCE?

Evaluation practice is ancient, but the discipline and profession of evaluation are quite contemporary. Indeed, evaluation as a profession and the concomitant elucidation of the discipline are only four or five decades old. This newness leads to self-consciousness about what it is we are doing when we say we are doing evaluation (and not research) and an explicit discourse about the fundamentals of evaluation (that differentiate it from research). Disciplines are fields of study having their own logic, often entailing various theories that enable study within the field. For the discipline of evaluation, there is a particular logic (Fournier, 1995; Scriven, 1999) and subtheories, including a theory of valuing, practice, prescription and use. (For a more complete discussion of these subtheories that constitute the discipline of evaluation, see Mathison 2004b.) One might imagine that such questions as "What is the difference between statistics (a newer discipline) and mathematics (a more established discipline)?" were also once asked as the discipline of statistics caught up with the practice of using probability in common-sense ways. The question about the difference between evaluation and research productively pushes evaluation theorists and practitioners to think about and describe the foundations of their discipline, first in comparison to social science research but increasingly in an analytic exploration of evaluation qua evaluation.

Evaluation, especially program evaluation in the United States, had its genesis in the provisions of the Elementary and Secondary Education Act (ESEA), passed in 1965. ESEA's requirement that the expenditure of public funds be accounted for thrust educators into a new and unfamil-

iar role, and researchers and educational psychologists stepped in to fill the need for evaluation created by ESEA. But the efforts of practitioners and researchers alike were only minimally successful at providing the kind of evaluative information envisioned. The compensatory programs supported by ESEA were complex and embedded in the complex organization of schooling. Research methods that focused on hypothesis testing were not well suited to the task at hand.

The late 1960s through the early 1980s were the gold rush days of evaluation. During this period, models of evaluation proliferated and truly exciting intellectual work was being done. A few traditionally educated educational psychologists experienced epiphanies that directed their thinking toward new ways to do evaluation. For example, Robert Stake, a psychometrician who began his career at the Educational Testing Service, wrote a paper called the "Countenance of Educational Evaluation" that reoriented thinking about the nature of educational interventions and what was important to pay attention to in determining their effectiveness (Stake, 1967). Egon Guba, a well-known education change researcher, abandoned the research, development, diffusion approach for naturalistic and qualitative approaches that examined educational interventions carefully and contextually. Lee Cronbach, a psychometric genius, focused not on the technical aspects of measurement in evaluation but on the policy-oriented nature of evaluation, an idea that led to a radical reconstruction of internal and external validity, including separating the two conceptually and conceptualizing external validity as related to the usability and plausibility of conclusions rather than as a technical feature of evaluation design (Cronbach, 1982).

The failure of social science research methods alone to answer questions about the value of programs spurred this tremendous growth in evaluation as a separate discipline. The matter is not, however, resolved. The debates about how best to determine what is working continue; an example is the current U.S. governmental preference for randomized clinical trials to determine what works. In the past, governmental funding supported the proliferation of evaluation and new ideas in evaluation practice; government agencies now presume to direct how evaluation ought ro be done.

For example, the U.S. Department of Education funds very little evaluation because of this narrowed definition of what the government now considers good evaluation. A quick search of the U.S. Department of Education website shows that it funded only 12 evaluation studies in 2003 and 6 in 2004. These are referred to as a new generation of rigorous evaluations. These evaluations must be randomized clinical trials, perhaps quasi-experimental or regression discontinuity designs. Few if any educational evaluations have heen of this sort; indeed, much of the

work since the 1960s has been directed to creating different evaluation methods and models of evaluative inquiry (not just borrowed research methods) that answer evaluative questions. These are questions about feasibility, practicability, needs, costs, intended and unintended outcomes, ethics, and justifiability.

The question about differences between evaluation and research is also about what knowledge and skills evaluators need, an especially critical matter if there are unique domains of knowledge and skills for evaluators. Evaluators are served well by having knowledge of and facility with most social science research methods, but as illustrated earlier, those alone are not adequate, either in terms of a methodological repertoire or the evaluation knowledge and skills domain. Scriven suggests evaluators must also know how to search for unintended and side effects, how to determine values within different points of view, how to deal with controversial issues and values, and how to synthesize facts and values (Coffman, 2003–2004).

While there are more graduate programs to educate future evaluators, it is still the case that many evaluators come to their work through a back door, often with knowledge of social science methods and statistics but relatively little knowledge about the knowledge domains suggested by Scriven. Given that this is the case, it is natural for this question to be a perennial one, as more and more novice evaluators are acculturated into a profession that requires them to acquire additional knowledge and skills.

In his 1998 American Evaluation Association presidential address, Will Shadish asked a series of questions about basic and theoretical issues in evaluation with which evaluators should be conversant (although they need not necessarily agree completely on the answers) in order to claim they know the knowledge base of their profession. For example, Shadish poses questions like "What are the four steps in the logic of evaluation?" and "What difference does it make whether the program being evaluated is new or has existed for many years?" and "What is metaevaluation and when should it be used?" The point here is not to answer the questions but to illustrate that evaluators are not simply social science researchers practicing in certain applied contexts, but rather they are engaged in a different practice and thus require special knowledge and skills.

The question about differences seems vexing at times, especially to those who have worked through it already. Novice evaluators who venture on to EVALTALK with this fundamental question will likely be met with a slightly impatient suggestion to check the archives, as the question has been thoroughly discussed. However, it is a good question, because it provides an impetus for evaluators to do the philosophical

and empirical work of saying what they are doing and how. And as a result, the potential for evaluation practice to be high-quality and to make significant positive contributions to understanding what is good and right is enhanced.

CHARACTERIZING THE ISSUE

It may be helpful to begin with an excerpt from the three evaluation scholars who provided much of the fodder for answering this question in their 1998 EVALTALK discussion. Below is a snippet from each that captures their perspectives.

According to Michael Quinn Patton (1998):

> The distinction between research and evaluation, like most of the distinctions we make, perhaps all of them, is arbitrary. One can make the case that the two are the same or different, along a continuum or on different continua completely.
>
> The purpose of making such distinctions, then, must guide the distinctions made. In my practice, most clients prefer and value the distinction. They want to be sure they're involved in evaluation, not research. The distinction is meaningful and helpful to them, and making the distinction helps engender a commitment from them to be actively involved—and deepens the expectation for use.

According to Bill Trochim (1998):

> Here's the question: Are there any differences between applied social research and *program* evaluation? Between applied social research and *process* evaluation? Between applied social research and *formative* evaluation? How about *summative* evaluation? Is there some combination of evaluation types that has so significant an overlap with the domain of applied social research that the two are hard or impossible to distinguish? My answer is YES—what we do in process–program–formative–summative evaluation (my new term for describing to Michael Scriven what "evaluation" is in my sense of the word!) is that it is virtually impossible to distinguish it (certainly in terms of methodology, less perhaps in terms of purpose) from what I understand as applied social research.

And, according to Michael Scriven (1998):

> Research, in the serious sense (by contrast with what tenth graders call looking something up in an encyclopedia) is extensive disciplined inquiry, whereas evaluation in the serious sense (by contrast with what wine columnists and art critics do) is disciplined determination of merit, worth, or

value. Generally speaking, this involves research, but not quite always since what judges do in assessing the credibility of witnesses or at a skating contest is entitled to be called serious evaluation and only involves the use of highly skilled judgment.

Since disciplined inquiry itself involves evaluation—what I call intradisciplinary evaluation (the evaluation of hypotheses, inferences, theories, etc.)—research is an applied field of evaluation (of course requiring vast amounts of further knowledge and skills besides evaluation skills); and for the reasons in the previous paragraph, much evaluation is a subset of research. So there's a double overlap.

Additionally, in an interview for an issue of the *Evaluation Exchange* reflecting on the past and future of evaluation, Scriven (Coffman, 2003–2004) addresses this question directly:

Evaluation determines the merit, worth, or value of things. The evaluation process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from the social sciences, and then integrates conclusions with the standards into an overall evaluation or set of evaluations.

Social science research, by contrast, does not aim for or achieve evaluative conclusions. It is restricted to empirical (rather than evaluative) research, and bases its conclusions only on factual results—that is, observed, measured, or calculated data. Social science research does not establish standards or values and then integrate them with factual results to reach evaluative conclusions. In fact, the dominant social science doctrine for many decades prided itself on being value free. So for the moment, social science research excludes evaluation.

However, in deference to social science research, it must be stressed again that without using social science methods, little evaluation can be done. One cannot say, however, that evaluation is the application of social science methods to solve social problems. It is much more than that.

WHAT IS THE DIFFERENCE BETWEEN EVALUATION AND RESEARCH?

Although there are arguments that evaluation and research, especially applied social science research, are no different, in general, evaluators do claim there is a difference but that the two are interconnected. Because evaluation requires the investigation of what is, doing evaluation requires doing research. In other words, determining the value, merit, or worth of an evaluand requires some factual knowledge about the evaluand and perhaps similar evaluands. But, of course, evaluation requires more than facts about evaluands, as suggested by Scriven in his

interview with *Evaluation Exchange*. Evaluation also requires the synthesis of facts and values in the determination of merit, worth, or value. Research, on the other hand, investigates factual knowledge but may not necessarily involve values and therefore need not include evaluation.

Even though research and evaluation are interconnected, there is often an attempt to find a parsimonious way to distinguish between the two. For example, research is considered by some to be a subset of evaluation; some consider evaluation to be a subset of research; some consider evaluation and research to be end points of a continuum; and some consider evaluation and research to be a Venn diagram with an overlap between the two. The similarities and differences between evaluation and research relate most often to the purpose of each (that is, the anticipated outcome of doing research or evaluation), the methods of inquiry used, and how one judges the quality of evaluation and research.

The Purpose of Evaluation and Research

The following list illustrates the plethora of pithy ways to capture the distinction between the purpose for and expected outcome of doing evaluation or research.

- Evaluation particularizes, research generalizes.
- Evaluation is designed to *improve* something, while research is designed to *prove* something.
- Evaluation provides the basis for decision making; research provides the basis for drawing conclusions.
- Evaluation—so what? Research—what's so?
- Evaluation—how well it works? Research—how it works?
- Evaluation is about what is valuable; research is about what is.

These attempts at a straightforward distinction between evaluation and research are problematic because they caricature both to seek clear differences. Simple distinctions between research and evaluation rely on limited, specific notions of both social science research and evaluation—social science research is far more diverse than is often suggested, including differing perspectives on ontology and epistemology, and evaluation as a professional practice is often what is portrayed rather than the foundations of evaluation, that is, as a discipline.

Take, for example, the popular distinction between generalization and particularization—neither of which is singularly true for either evaluation or research. While evaluation is profoundly particular in the sense that it focuses on *an* evaluand, evaluation findings may nonethe-

less and often do generalize. Cronbach's (1982) treatise on designing evaluations specifically addressed the issue of external validity, or the generalization of evaluation findings. The kind of generalization Cronbach described was not a claim about a population based on a sample but rather a knowledge claim based on similarities between UTOS (units, treatments, observations, settings). This suggests the possibility that such a claim might be made by the evaluator or someone completely separate from the evaluation but in a similar setting to the evaluand. (This external validity claim would be represented in Cronbach's notation as moving from utoS [a particular evaluation] to *UTOS [settings with some similarities in uto].) Formulated somewhat differently is Stake's notion of naturalistic generalizations, a more intuitive approach to generalization based on recognizing the similarities of objects and issues in like and unlike contexts (Stake, 1978).

Similarly, research may not be primarily about generalization. A historical analysis of the causes of the French Revolution, an ethnography of the Minangkabau, or an ecological study of the Galapagos Islands may not be conducted in order to generalize to all revolutions, all matriarchal cultures, or all self-contained ecosystems.

While evaluation is primarily about the particular, evaluations also provide the basis for generalization, especially in Cronbach's sense of generalizing from utoS to *UTOS and Stake's naturalistic generalizations. And, while much research is about generalizatioo, especially in the sense of making inferences from a sample to a population, much research is also about the particular.

Another common difference asserted is that evaluation is for decision making, while research is for affirming or establishing a conclusion. But, not all evaluation is about decision making or action. There are contexts in which evaluation is done for the sake of doing an evaluation, with no anticipation of a decision, change, or improvement. For example, evaluations of movies or books can be ends in and of themselves. Most practicing evaluators think first and foremost about the practice of evaluation rather than the discipline of evaluation and therefore focus, appropriately, on the expectation that evaluation will be useful, will lead to improvement, and will help in creating better living. But, of course, evaluation is a discipline, and when one thinks about the discipline of evaluation there is a clear distinction between making evaluative claims and making prescriptions. While these are logically connected, they are in fact distinct forms of reasoning.

In addition, some forms of social science research are closely aligned to social action or seeking solutions to social problems. For example, various forms of action research are about alleviating prob-

lems, taking action, determining what is and is not valued, and working toward a social state of affairs consistent with those values. Policy analysis is directly connected to decision making, especially with regard to the alleviation of all sorts of problems—social, environmental, health, and so on. The notion that research is about establishing facts, stating what is, drawing conclusions, or proving or disproving an assertion is true for some but not all social science research.

The link between social science research and action, including decision making, may be looser than for evaluation. Much social science research aspires to influence policies and practices but often does so indirectly and through an accumulation of research knowledge—what might be considered influence at a macro level. In fact, the results from an individual research study are typically considered insufficient to directly determine a decision or course of action since the results of a single study provide an incomplete picture. However, social scientists hope the studies they conduct will affect decision making by raising awareness of an issue, contributing to a body of research that will cumulatively inform decision making, identifying policy alternatives, informing policymakers, and establishing research evidence as the basis for adopting interventions (thus the preoccupation with evidence-based practices).

Evaluation may more directly affect decisions about particular evaluands, that is, affect micro-level decisions. But, the discipline's decades-long preoccupation with evaluation utilization, or lack of utilization, suggests that the connections between evaluative claims and decisions are also tenuous. An obvious example, because of the ubiquity of the program, is evaluations of the effectiveness of the DARE (Drug Abuse Resistance Education) program. Because the DARE program is everywhere, in virtually all schools in the United States as well as many other countries, many evaluations of the program have been conducted. But the results are mixed: early evaluations of DARE concluded it was effective, but more recent evaluations suggest the program does little to decrease the incidence of drug use (see Beck, 1998; and a special issue of *Reconsider Quarterly* [2001–2002] devoted to a critical analysis of drug education, including DARE). Just as an accumulation of research knowledge affects practice, so it is sometimes the case with evaluation, as illustrated by the DARE example. Only over time and with the investment of considerable programmatic and evaluative resources have evaluations of DARE actually resulted in any significant changes in the program.

While evaluation may be more likely to contribute to micro decision making, and research more likely to contribute to macro decision making, this distinction does not hold up so well when analyzed a little more carefully.

Differences in Social Science Research and Evaluation Methods

Evaluation, especially in the infancy of the discipline, borrowed generously from the social sciences for its methods of inquiry. In part, because evaluators were educated within social science traditions, especially psychology and sociology, and to a much lesser extent anthropology, ways of establishing empirical evidence were informed by these traditions. For some evaluators this has not changed, but for many the practice of evaluation has moved far beyond these early days. Because evaluation necessarily examines issues like needs, costs, ethicality, feasibility, and justifiability, evaluators employ a much broader spectrum of evidentiary strategies than do the social sciences. In addition to all the means for accessing or creating knowledge used by the social sciences, evaluators are likely to borrow from other disciplines such as jurisprudence, journalism, arts, philosophy, accounting, and ethics.

While there has been an epistemological crisis in the social sciences that has broadened the repertoire of acceptable strategies for data collection and inquiry, evaluation has not experienced this crisis in quite the same way. While the evaluation profession has explored the quantitative and qualitative debate and is attentive to the hegemony of randomized clinical trials, evaluation as a practice shamelessly borrows from all disciplines and ways of thinking to get at both facts and values. Elsewhere I use Feyerabend's notion of an anarchist epistemology to describe this tendency in evaluation (Mathison, 2004a). Anarchism is the rejection of all forms of domination. And so, an anarchist epistemology in evaluation is a rejection of domination of one method over any or all others; of any single ideology; of any single idea of progress; of scientific chauvinism; of the propriety of intellectuals; of evaluators over service recipients and providers; of academic text over oral and other written traditions; of certainty.

In evaluation practice one sees no singular commitment to social science ways of knowing but rather a consideration of how to do evaluation in ways that are meaningful in the context. There are many examples of this, but two will suffice to illustrate the point—consider the "*most significant change technique*" and "*photovoice.*" While both of these methods of inquiry could be used in social science, they have particular salience in evaluation because they focus explicitly on valuing and emphasize the importance of stakeholder perspectives (a perspective unique to evaluation and not shared by social science research).

The most significant change technique

> involves the collection of significant change (SC) stories emanating from the field, and the systematic selection of the most significant of these stories

> by panels of designated stakeholders or staff. The designated staff and stakeholders are initially involved by "searching" for project impact. Once changes have been captured, various people sit together, read the stories aloud and have regular and often in-depth discussions about the value of these reported changes. When the technique is implemented successfully, whole teams of people begin to focus their attention on program impact. (Davies & Dart, 2005)

The second example is photovoice, a grassroots strategy for engaging stakeholders in social change, that is, in asserting what is valued and good (Wang, Yuan, & Feng, 1996). Photovoice uses documentary photography techniques to enable those who are often the "service recipients" or "subjects" to take control of their own portrayal; it has been used with refugees, immigrants, the homeless, and people with disabilities. Photovoice is meant to tap personal knowledge and values and to foster evaluation capacity building—individuals learn a skill that allows them to continue to be a voice in their community. Hamilton Community Foundation in Canada uses photovoice to evaluate and transform neighborhoods challenged by poverty, unemployment, and low levels of education (see their website *www.photovoice.ca/index.php* for a description of the evaluation).

Particular approaches to evaluation are often connected to particular social science traditions, and so sometimes the unique and broader array of methods of inquiry in evaluation is overlooked. The two examples above (most significant change technique and photovoice) illustrate how evaluators have begun to develop evaluation-specific methods to adequately and appropriately judge the value, merit, or worth of evaluands. There are many other such examples: cluster evaluation, rural rapid appraisal, evaluability assessment, and the success case method, to name a few more.

Judging the Quality of Evaluation and Research

Another difference between evaluation and research—and one stressed especially by Michael Quinn Patton in the 1998 EVALTALK discussion of this question—is how we judge the quality of evaluation and research. In making this distinction, Patton suggested that the standards for judging both evaluation and research derive from their purpose. The primary purpose of research is to contribute to understanding of how the world works, and so research is judged by its accuracy, which is captured by its perceived validity, reliability, attention to causality, and generalizability. Evaluation is judged by its accuracy also, but additionally by its utility, feasibility, and propriety. These dimensions for judging the quality of

evaluation are clear in the Metaevaluation Checklist, which is based on the Program Evaluation Standards (Stufflebeam, 1999).

A distinguishing feature of evaluation is the universal focus on stakeholder perspectives, a feature not shared by social science research. And evaluations are judged on if and how stakeholder perspectives are included. While evaluation models are based on different epistemological foundations, all evaluation models attend to stakeholder perspectives—the assessments, values, and meanings of the evaluand's stakeholders are essential elements in any evaluation. Social science research may include the language of stakeholders, but their inclusion is not necessary. When research participants are referred to as stakeholders, it is often a reference to whom the data are collected from rather than a consideration of the stakeholders' vesred interests. The following excerpt from the Centers for Disease Control's (1999) *Framework for Program Evaluation* illustrates the essential inclusion of stakeholders in evaluation. No approach in social science research necessarily includes this concept.

> The evaluation cycle begins by engaging stakeholders (i.e., the persons or organizations having an investment in what will be learned from an evaluation and what will be done with the knowledge). Public health work involves partnerships; therefore, any assessment of a public health program requires considering the value systems of the partners. Stakeholders must be engaged in the inquiry to ensure that their perspectives are understood. When stakeholders are not engaged, evaluation findings might be ignored, criticized, or resisted because they do not address the stakeholders' questions or values. After becoming involved, stakeholders help to execute the other steps. Identifying and engaging the following three groups are critical:
>
> 1. those involved in program operations (e.g., sponsors, collaborators, coalition partners, funding officials, administrators, managers, and staff)
> 2. those served or affected by the program (e.g., clients, family members, neighborhood organizations, academic institutions, elected officials, advocacy groups, professional associations, skeptics, opponents, and staff of related or competing organizations)
> 3. primary users of the evaluation (e.g., the specific persons who are in a position to do or decide something regarding the program) In practice, primary users will be a subset of all stakeholders identified. A successful evaluation will designate primary users early in its development and maintain frequent interaction with them so that the evaluation addresses their values and satisfies their unique information needs.

CONCLUSION

Evaluation and research are different—different in degree along the dimensions of both particularization–generalization and decision-oriented–conclusion-oriented, the most common dimensions for making distinctions. But, evaluation and research are substantively different in terms of their methods: evaluation includes the methods of data collection and analysis of the social sciences but as a discipline has developed evaluation-specific methods. And, they are substantially different in how judgments of their quality are made. Accuracy is important in both cases, but the evaluation discipline uses the unique criteria of utility, feasibility, propriety, and inclusion of stakeholders.

As evaluation matures as a discipline with a clearer sense of its unique focus, the question of how evaluation is different from research may wane. However, as long as evaluation methodology continues to overlap substantially with that used in the social sciences and as long as evaluators come to the profession from more traditional social science backgrounds, this will remain a fundamental issue for evaluation. As suggested earlier, fundamental issues provide opportunities for greater clarity about what evaluation is as a practice, profession, and discipline.

REFERENCES

Beck, J. (1998). 100 years of "just say no" versus "just say know": Reevaluating drug education goals for the coming century." *Evaluation Review*, 22(1), 15–45.

Centers for Disease Control. (1999). *Framework for program evaluation in public health*. Atlanta: Centers for Disease Control. Retrieved December 15, 2005, from *www.cdc.gov/eval/framework.htm*.

Coffman, J. (2003–2004, Winter). Michael Scriven on the differences between evaluation and social science research. *The Evaluation Exchange*, 9(4). Retrieved February 26, 2006, from *www.gse.harvard.edu/hfrp/eval/issue24/expert.html*.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Davies, R., & Dart, J. (2005). *The most significant change technique: A guide to its use*. Retrieved February 26, 2006, from *www.mande.co.uk/docs/MSCGuide.pdf*.

Fournier, D. (Ed.) (1995). *Reasoning in evaluation: Inferential links and leaps* (New Directions for Evaluation no. 48). San Francisco: Jossey-Bass.

Mathison, S. (2004a). *An anarchist epistemology in evaluation*. Paper presented

at the annual meeting of the American Evaluation Association, Atlanta. Retrieved February 26, 2006, from *weblogs.elearning.ubc.ca/mathison/Anarchist%20Epistemology.pdf*.

Mathison, S. (2004b). Evaluation theory. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 142–143). Newbury Park, CA: Sage.

Patton, M. Q. (1998, January 16). Research vs. evaluation. Message posted to *bama.ua.edu/cgi-bin/wa?A2=ind9801C&L=evaltalk&P=R1418&I=1&X=15AF91488C1E74BD45&Y.*

Reconsider Quarterly, The Education Issue. (2001–2002, Winter), *1*(4; special issue).

Scriven, M. (1998, January 16). Research vs. evaluation. Message posted to *bama.ua.edu/cgi-bin/wa?A2=ind9801C&L=evaltalk&P=R2131&I=1&X=2F11357E5870213C59&Y.*

Scriven, M. (1999). The nature of evaluation: Part I. Relation to psychology. *Practical Assessment, Research and Evaluation*, *6*(11). Retrieved December 15, 2005, from *pareonline.net/getvn.asp?v=6&n=11*.

Shadish, W. (1998) Presidential address: Evaluation theory is who we are. *American Journal of Evaluation, 19*(1), 1–19.

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*(7), 523–540.

Stake, R. E. (1978). The case study in social inquiry. *Educational Researcher, 7*(2), 5–8.

Stufflebeam, D. (1999). *Metaevaluation checklist.* Retrieved December 15, 2005, from *www.wmich.edu/evalctr/checklists/program_metaeval.htm.*

Trochim, W. (1998, February 2). Research vs. evaluation. Message posted to bama.ua.edu/cgi-bin/wa?A2=ind9802A&L=evaltalk&P=R503&I=1&X=089CE94613356B8693&Y.

Wang, C., Yuan, Y. L., & Feng, M. L. (1996). Photovoice as a tool for participatory evaluation: The community's view of process and impact. *Journal of Contemporary Health*, *4*, 47–49.