Educational Evaluation and Policy Analysis
Summer, Vol. 6, No. 2, 1984, pp. 131–138

A Rating Scale for Assessing the Quality of Responsive/Illuminative Evaluations

Graham S. Maxwell
University of Queensland

Responsive evaluation recently has been identified as a major alternative to more traditional approaches to program evaluation (Guba & Lincoln, 1981). Lundgren (1977) explicitly linked responsive evaluation to illuminative evaluation (Parlett & Dearden, 1977; Parlett & Hamilton, 1976) in concert against the traditional approaches, which he terms "effectivity" approaches because of their tendency to focus on effects and to ignore processes. House (1978) characterized these latter approaches as involving utilitarian values and objectivist epistemology, as contrasted to the intuitionist-pluralist values and subjectivist epistemology of responsive and illuminative evaluations. Hamilton (1977) also stressed the common allegiance of responsive and illuminative evaluations to pluralist values in contrast to other approaches. Although these distinctions do not exhaustively partition the realm of evaluation models, they do identify two major categories which involve different assumptions and commitments and which suggest different intentions and methods for undertaking evaluations.

In view of the common assumptions and commitments of responsive and illuminative evaluations, each can draw strength from the other in a unified approach that will be referred to here as responsive/illuminative evaluation. Such a unified approach offers the possibility of a richer conceptual structure for guiding the evaluations of those espousing intuitionist-pluralist values and subjectivist epistemology than does each approach separately. It also might provide a stronger anchor against the traditional tide of utilitarian values and objectivist epistemology than each approach can separately. In practice, too, evaluators often draw on both when they decide that such approaches are appropriate, so there are practical reasons for emphasizing their compatibility and highlighting their common characteristics. It is hoped that the unified approach presented here can guide such practical considerations more systematically and wisely.

Of course, a unified approach is possible only if the essential characteristics of each approach are common to both or compatible with each other. In what follows, it will be seen that all essential characteristics of responsive and illuminative evaluations are either common or compatible and that any incompatible characteristics are nonessential. The essential characteristics can be formulated into a rating scale intended to assist in designing such evaluations, in monitoring their implementation, and in reviewing their quality.

Responsive/Illuminative Approaches to Evaluation

To emphasize the common and compatible characteristics of responsive and illuminative evaluations, it is necessary to summarize the main features of each. In the space available here, these summaries may oversimplify the richness of

the original accounts. However, those accounts still are not so easily available or so well disseminated that their content can be assumed to be well known, particularly when compared to traditional approaches. Furthermore, it is necessary here to highlight the most important aspects of each approach in order to establish the basis for the rating scale.

Responsive evaluation offers a transactional and participatory orientation toward establishing the meaning and value of a program. As described by Stake (1975a, 1975b, 1976) and by Guba and Lincoln (1981), it attends more to emergent issues than to preconceived expectations. As such, it attends to the program as it is rather than how it was intended to be. A starting point is discussion with people whose lives are touched by the program (the stakeholders) to discover how the program operates and what concerns people about it. Instead of adopting a particular value perspective, the evaluator seeks first to reveal the different, sometimes conflicting, values of these people. Typically, therefore, evaluators who adopt a responsive approach focus on observing the program itself, talking with the participants, identifying key issues, and portraying the salient characteristics, good and bad, successful and unsuccessful, as perceived and experienced by various stakeholders.

In general, formal measurement procedures are avoided in favor of more naturalistic observation and representation; interviews are adopted in preference to questionnaires; and anecdotes and work samples are used rather than performance tests. This emphasis on qualitative and discursive information requires that the evaluator check and recheck the accuracy of the observations, the pertinence of the portrayals, and the relevance of the findings by obtaining reactions from various audiences. These reactions help redirect and refocus the evaluation and allow further exploration of the implications of different points of view. In reporting, the aim is to reveal the varied components of the program and the different understandings of their worth.

This phenomonological orientation of responsive evaluation is also found in *illuminative evaluation* (Parlett & Dearden, 1977; Parlett & Hamilton, 1976), which takes the program as an entity to be explored and revealed in detail: its manner of operation, its history of development and change over time, the way it has been shaped by its context, its response to changing demands and circumstances, and its various effects. In exploring its effects, special attention is focused on the stakeholders' personal perceptions and experiences. This not only emphasizes the richness and diversity of these effects but their linkage to the events of the program. In other words, effects are seen as embedded in the context of the program. The intention is "to discover and document" what it is like to be involved in the program and "to discover and discuss [its] most significant features, recurrent concomitants and critical processes" (Parlett & Hamilton, 1976).

As with responsive evaluation, such an approach is oriented more to *ethnographic* procedures than to *measurement* procedures, more to naturalistic techniques than to experimental techniques, more to qualitative data than to quantitative data, and more to processes than to outcomes. It seeks to offer a representation of the program that can provide a vicarious experience of the program for outsiders, convey the program's important operational features, highlight its critical and problematic features, encompass its common and atypical experiences, expose its framework or interconnection of components, reveal the similarities and differences of various viewpoints and opinions, and untangle any puzzling dilemmas or confused thinking about it (Parlett & Dearden, 1977).

Neither approach offers strong methodological guidelines. This is because the evaluation design is seen as emerging responsively rather than being imposed preordinately (Stake 1975a, 1975b). However, Parlett and Dearden (1977) have proposed that illuminative evaluation involve three overlapping stages of development: *observation* (in which attention is given to program orientation and setting); *progressive focusing* (in which emergent issues are identified and explored gradually); and *explanation* (in

which reasons for effects and opinions are sought). These stages are also relevant for responsive evaluation.

Both approaches also point to the possibility of addressing several audiences, the need for tailoring reports to audiences, and the desirability of "thick description" (Geertz, 1973).

There are, of course, differences of emphasis and style. This is clear from their titles. Responsive evaluation is more concerned with responding to audience needs for information and attending to their different value perspectives; illuminative evaluation is more concerned with enhancing audience understandings through representations of perceptions and events. Therefore, responsive evaluation can be seen as more populistic and more deferential to stakeholders, whereas illuminative evaluation can be seen as more authoritative and expository. In turn, these suggest more concern in illuminative evaluation with synthesis and explanation, whereas responsive evaluation resists general explanation and focuses instead on assisting personal understanding. Responsive evaluation is therefore more amenable to nonpropositional forms of communication; its definitive product is understanding, rather than findings or reports. Understanding is certainly a goal of illuminative evaluation, but more as an effect than a product; that is, illuminative evaluation can dispense less easily with formal reports.

However, these distinctions are fairly subtle, more a matter of emphasis than of contrast. In practice, too, it is unusual for evaluators to adhere to a consistent interpretation of each approach and such distinctions become blurred. The justification for such blurring is embedded within the notion of responsive evaluation itself and for most evaluators the similarities between these two approaches must appear much more substantial than their differences. In fact, Guba (1976) has gone so far as to suggest that "the illumination [sic] model . . . might be thought of as a special form of the . . . responsive model" (p. 39).

Responsive evaluation and illuminative evaluation are clearly so similar in their paradigmatic assumptions and points of view (Smith, 1981b)—or, as Stake (1981) characterizes such underlying assumptions, in their "persuasions"—that they supplement and complement each other very nicely. The solutions they suggest to the problems of evaluation design (Sadler, 1982) are very similar and can be enhanced by joint consideration. That is, it seems beneficial to see these approaches as essentially the same approach, which we might term "responsive/illuminative evaluation."[1]

The Rating Scale

The absence of explicit guidelines or codified rules for undertaking responsive/illuminative evaluations makes it difficult for many evaluators to design such evaluations and to assess their efficacy. Although it is clear that responsive/illuminative evaluations cannot be designed or implemented prescriptively, some indications can be made of the sorts of characteristics that ought to be present in an evaluation if it is to be called responsive/illuminative.

The following rating scale offers a set of dimensions that characterize evaluations adopting a responsive/illuminative approach. These dimensions are not necessarily of equal importance—that would be too prescriptive; rather, their relative importance must be assessed within the context of the particular evaluation. Nor are they necessarily exhaustive—that would be too presumptuous; rather, these dimensions represent the most salient common and compatible characteristics of responsive and illuminative evaluations as proposed by their originators. In particular circumstances, some dimensions might be dropped and others added, although extreme modification might indicate that the stance is no longer responsive/illuminative. (See Table I.)

Glossary of Terms

The following glossary provides some brief explanations and typifying questions for the dimensions of the rating scale.

---

[1] Other closely related concepts are *portrayal* (Kemmis, 1977; Stake, 1979), *naturalistic enquiry* (Guba, 1978), and *case study* (Macdonald & Walker, 1975; Simons, 1980; Smith, 1978; Stake & Easley, 1978; Welch, 1981).

1. *Search for:*

(a) *Key issues:* An issue is a matter of contention among various stakeholder groups. Normally there are too many issues for the evaluation to address them all; hence the need to focus on the most important (see also 3a–3d).

(b) *Operational features:* How is the program organised and how does it operate? What events and sequences occur and how are they related? Who is involved, what are their roles, and how do they associate?

(c) *Contextual influences:* What is the history of the program? What changes have occurred over time and why? Why does it exist? How is it funded? What resources are available? What are the physical, social, and economic contexts?

(d) *Participant experiences:* What do observers see happening to participants? What do participants see happening to other participants? What do participants report about their own experiences?

(e) *Stakeholder judgments:* Stakeholders are people who are or may be affected (directly or indirectly) by the program. What do they have to say about the program? How do they judge its merit and worth? How do they react to other viewpoints?

2. *Shape study through:*

(a) *Observation:* Initially, instead of asking, What are the goals?, ask, What is happening? This requires on-site visits for field observation and discussions. The aim is thorough familiarization with all aspects of the program.

(b) *Progressive focusing:* The evaluation design evolves; it is flexible rather than static. Initially there is some oscillation between seeking more information and identifying patterns and trends. New information helps to confirm tentative conclusions and raise new questions. Gradually, the emphasis shifts from breadth to detail and from search to closure.

(c) *Explanation:* Look for reasons for opinions, relationships among events, and causes for effects (see also 3d).

3. *Emphasize:*

(a) *Emergent issues:* Issues (see 1a) serve to focus and direct the study itself and the reporting. They should emerge from the investigation (see 2b) through analysis of the concerns of various stakeholders (see 1e).

(b) *Naturalistic observation:* *Naturalistic* contrasts with *manipulative* and hence emphasizes observations in natural settings rather than in constrained settings such as experiments and tests. Observations include program events, background details, and participant opinions. These observations are organized and interpreted by considering their various relationships and meanings (see also 4b and 4c).

(c) *Participant discussion:* Interviews and discussions with program participants allow individual perspectives and values to be explored in detail and anchored in context.

(d) *Iterative checking:* Interim accounts can be continually referred back to informants for confirmation and elaboration.

4. *Report with:*

(a) *Simple and direct language:* Understanding is more likely if technical words and complex sentences are avoided in favor of vernacular words and simple expressions.

(b) *Naturalistic representations:* Be specific and detailed rather than general and abstract. Use examples from the naturalistic records (see 3b) such as quotations, photographs, work samples, and case studies.

(c) *Multiple perspectives:* Include all different viewpoints and opinions, not just the most typical or representative.

(d) *Explicit source of values:* Make it clear whose value perspective is being adopted, especially when it is the evaluator's.

5. *Aim to convey:*

(a) *Vicarious experience:* Reporting should be vivid and engaging. It should leave the audience with the feeling of knowing what it is like to be involved in or affected by the program.

(b) *Program portrayal:* Reporting should emphasize understanding rather than conclusions and recommendations. What is the program? How does it operate? How is it perceived by various stakeholders? What are its problems? What are its possibilities?

(c) *Untidy reality:* Irregular effects, alternative explanations, conflicting view-

points and unresolved issues should be highlighted as a basis for further discussion and investigation.

Uses of the Rating Scale

The rating scale defines a set of dimensions or characteristics that might assist in designing, monitoring, and reviewing evaluations whose orientation is responsive/illuminative. Most likely, in the design phase of such evaluations, the list of characteristics would be used as a checklist of features to incorporate in the design. However, ratings could be used to indicate the strength of importance attached to each characteristic and serve as a target to refer back to when monitoring the implementation of the design or assessing the quality of the completed evaluation. Such use of the set of characteristics as a checklist is fairly straightforward; use as a rating scale requires further elaboration.

As already indicated, the rating scale is intended to provide a basis for discussion about the quality of implementation of responsive/illuminative evaluations by allowing implicit standards to be made more explicit and amenable to examination and challenge. Over time, perhaps specific examples can be agreed upon to provide benchmarks for the different levels of quality on each characteristic. However, even if such consensus proves unattainable, the attempt may provide a stimulus and focus for the discussion so necessary to improving the quality of program evaluations (Cronbach et al., 1980).

More explicitly, the rating scale should prove useful in the following kinds of evaluations of evaluations: internal meta-evaluations, external meta-evaluations, descriptive-analytic studies, and critical reviews.[2] In each case, however, it is important to remember that attention is being directed at the complete evaluation study, not simply an evaluation report (which might not reveal the nature of the evaluation process itself).

1. The term *internal meta-evaluations* refers, here, to evaluations of evaluations conducted by the evaluator, the evaluation team, the evaluation management committee, or the evaluation advisory group, or even by an evaluation consultant if his or her role is the formative and participatory one described by Cook and Gruder (1978) as "consultant meta-evaluation." Such meta-evaluations are directed at either *monitoring* the progress of the evaluation while it is still in progress or *assessing* the study once it is completed. In both cases, the intention is to learn from past experience and to inform future decisions. In the first case (monitoring), modifications can be made to the evaluation to improve its quality while it is still in progress; in the second case (assessing), postmortem analysis can produce greater understanding of how the evaluation might have been improved and of how subsequent evaluations might be designed. In either case, the rating scale might be a tool for sharpening awareness, focusing discussion, and enhancing understanding.

2. The term *external meta-evaluations* here refers to evaluations of evaluations conducted by an independent person or team. Such meta-evaluations include *post-hoc analysis*, *multiple comparisons*, and *audience assessments*.

Post-hoc analysis is an external meta-evaluation that does not influence the conduct of the evaluation itself. Such a meta-evaluation might include information gathered (through observation, interview, and discussion) during the progress of the evaluation itself. In fact, many of the listed characteristics could be assessed only that way. But final judgment and reporting would be reserved until the evaluation was completed. This kind of meta-evaluation needs to include an explanation and defence of the various judgments of quality. Further, the meta-evaluation should be seen as a starting point for further discussion and clarification, both of the nature of the judgments and

[2] An adequate taxonomy of meta-evaluations currently does not exist. Cook and Gruder (1978) attempted one, but it is concerned only with the "technical quality" of "empirical summative evaluations" from the perspective of utilitarian values and objectivist epistemology to which responsive/illuminative evaluation does not subscribe. Stufflebeam (1981) offers a somewhat broader view of meta-evaluation but still tied to "effectivity" approaches and without any taxonomic categories. A broad and systematic taxonomy of meta-evaluations is urgently needed. What is offered here is merely a partial list of types of meta-evaluations—those that offer the strongest potential for using the rating scale.

the suitability of the standards implied. The aim should be not to criticize but to understand, and especially to suggest ways in which the quality of evaluations can be enhanced.

Multiple comparisons involve the analysis of several evaluations simultaneously so that their comparative strengths and weaknesses can be highlighted. The potential value of the rating scale for this activity is clear because it provides explicit dimensions on which to compare the studies. Clearly too, aggregation of separate post-hoc analyses into multiple comparisons may be facilitated by having a common vocabulary and analysis scheme.

Audience assessments are a natural extension of the responsive/illuminative approach. The rating scale provides a framework for discussions with the evaluation audiences (or stakeholders) about the aims of the evaluation and provides a basis for them to make judgments about its quality. It would, of course, be likely that the scale dimensions themselves be subject to question and discussion. This could help in extending understanding about evaluation to more people and in stimulating continual rethinking of our formulations of appropriate evaluation approaches and methods.

3. *Descriptive-analytic studies* (typically in the form of a technical report) are the formal counterpart of internal meta-evaluations. They offer a critical description and review of the evaluation, particularly its historical development, the thinking involving its design, the contextual influences shaping and modifying its progress, the reasons for various decisions taken, the alternative possibilities considered and the reasons for their rejection, the constraints of circumstances, resources and personnel and their influence on the evaluation, the procedures adopted for gathering and synthesizing information, and the considerations leading to the

TABLE I

Rating Scale for Use with Responsive/Illuminative Evaluations

Dimensions	Ratings			
	Very weak	Weak	Strong	Very strong
1. Search for:				
(a) Key issues				
(b) Operational features				
(c) Contextual influences				
(d) Participant experiences				
(e) Stakeholder judgments				
2. Shape study through:				
(a) Observation (first)				
(b) Progressive focusing (second)				
(c) Explanation (third)				
3. Emphasize:				
(a) Emergent issues				
(b) Naturalistic observation				
(c) Participant discussion				
(d) Iterative checking				
4. Report with:				
(a) Simple and direct language				
(b) Naturalistic representations				
(c) Multiple perspectives				
(d) Explicit source of values				
5. Aim to convey:				
(a) Vicarious experience				
(b) Program portrayal				
(c) Untidy reality				

final report(s). In other words, the emphasis is on critical explanation and defence of the reasons for the evaluation developing the way it did. Such studies differ from external meta-evaluations in that they are conducted by an insider and stress detailed explanation rather than global assessment of quality.

In such studies, the rating scale should be useful for prompting aspects of the evaluation to be discussed. Considerations of the quality of each of the characteristics of the scale should prompt discussion of alternatives that were or were not considered and which might or might not have improved the evaluation.

4. *Critical reviews* are not substantially different from post-hoc analyses, except that they are focused entirely on the final report, perhaps linked to interviews with the evaluator(s). Cook and Gruder (1978) call these *essay reviews* and Smith (1981a) calls them *formal criticisms*, but the intention is much the same. Here, the evaluation process is not observed and the report becomes paramount. The rating scale might be used to question its adequacy as a product of responsive/illuminative evaluation and to provide a "telling structure" for the critique (Eisner, 1979; Smith 1981a). Although some of the dimensions may be difficult to assess because they refer to evaluation processes rather than evaluation products, it can at least be asked whether the report (together with an accompanying technical report, perhaps) gives enough information about each dimension to allow the reader to judge how well the study has been implemented. As with all other uses, it is more important to create discussion than to be definitive.

Conclusion

A simple rating scale has been devised for assessing the implementation of responsive/illuminative evaluations. All dimensions of this scale need not necessarily be present in any particular evaluation; those that are present need not necessarily be of equal importance or warrant high quality representation. Part of the use of the rating scale must therefore be directed at a consideration of the relevance of the dimensions for the particular evaluation. Nevertheless, the rating scale provides a starting point for such discussions and should allow comparisons among several evaluations, and indeed of several evaluations of evaluations, to be carried out with greater clarity and understanding. It is hoped that much use will be made of the rating scale along the lines suggested. Perhaps, after evaluators have become more familiar with responsive/illuminative evaluations and various standards for good and poor exemplars established, such a scale can be dispensed with. In the meantime, however, it can be a vehicle for generating the necessary discussion to clarify, codify, and communicate those standards.

References

Cook, T. D., & Gruder, C. L. (1978). Meta-evaluation research. *Evaluation Quarterly, 2*, 5–51.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation: Aims, methods and institutional arrangements.* San Francisco: Jossey-Bass.

Eisner, E. W. (1979). *The educational imagination: On the design and evaluation of school programs.* New York: Macmillan.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. *CSE Monograph Series in Evaluation, 8.* Los Angeles: Center for the Study of Evaluation, University of California.

Guba, E. G. & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Hamilton, D. (1977). Making sense of curriculum evaluation: Continuities and discontinuities in an educational idea. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 318–347). Itasca, IL: F. E. Peacock.

House, E. R. (1978). Assumptions underlying evaluation models. *Educational Researcher, 7*(3), 4–12.

Kemmis, S. (1977). Telling it like it is: The problem of making a portrayal of an educational program. In L. Rubin (Ed.), *Curriculum handbook: Administration and theory.* Boston: Allyn and Bacon.

Lundgren, U. P. (1977). *Model analysis of pedagogical processes* (Studies in curriculum theory and cultural reproduction). Stockholm: CWK Gleerup.

Macdonald, B., & Walker, R. (1975). Case study and the social philosophy of educational research. *Cambridge Journal of Education, 5*(1), 2–11.

Parlett, M., & Dearden, G. (Eds.). (1977). *Introduction to illuminative evaluation: Studies in higher education.* Cardiff-by-the-Sea, CA: Pacific Soundings Press.

PARLETT, M., & HAMILTON, D. (1976). Evaluation as illumination: A new approach to the study of innovatory programs. *Evaluation Studies Review Annual, 1*, 140–157.

SADLER, D. R. (1982). Generalized evaluation design. *Evaluation News, 3*(3), 80–85.

SIMONS, H. (ED.). (1980). *Towards a science of the singular* (CARE Occasional Paper, No. 10). Norwich, England: Centre for Applied Research in Education, University of East Anglia.

SMITH, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. *Review of research in education* (Vol. 6, pp. 316–377).

SMITH, N. L. (1981a). Criticism and meta-evaluation. In N. L. Smith (Ed.), *New perspectives in evaluation*: Vol. 2. *New techniques for evaluation* (pp. 266–274). Beverly Hills: Sage.

SMITH, N. L. (1981b). Developing evaluation methods. In N. L. Smith (Ed.), *New perspectives in evaluation: Vol. 1. Metaphors for evaluation: Sources for new methods* (pp. 17–50). Beverly Hills: Sage.

STAKE, R. E. (1975a). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Merrill.

STAKE, R. E. (1975b). *Program evaluation: Particularly responsive evaluation* (Occasional Paper No. 5). Kalamazoo: Evaluation Centre, Western Michigan University.

STAKE, R. E. (1976). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation, 2*, 19–22.

STAKE, R. E. (1979, April). *Portrayal of programs through evaluation studies*. Paper presented at the CIRCE-IOE Spring Evaluation Conference, University of Illinois, Urbana.

STAKE, R. E. (1981). Persuasions, not models. *Educational Evaluation and Policy Analysis, 3*(1), 83.

STAKE, R. E., & EASLEY, J. A., JR. (1978). *Case studies in science education*. Urbana: Center for Instructional Research and Curriculum Evaluation, University of Illinois.

STUFFLEBEAM, D. L. (1981). Meta-evaluation: Concepts, standards and uses. In R. A. Berk (Ed.), *Educational evaluation methodology: The state of the art*. Baltimore, MD: Johns Hopkins University Press.

WELCH, W. (ED.). (1981). *Case study methodology in educational evaluation: Proceedings of the 1981 Minnesota Evaluation Conference*. Minneapolis: Minnesota Research and Evaluation Center, University of Minnesota.

Author

GRAHAM S. MAXWELL, Lecturer, Department of Education, University of Queensland, St. Lucia, Q. 4067, Australia. *Specialization:* Educational evaluation and research.