Review of Educational Research
Spring 1983, Vol. 53, No. 1, Pp. 117–128

The Conceptualization of Educational Evaluation: An Analytical Review of the Literature

David Nevo
Tel-Aviv University

Recent decades have been productive in the conceptualization of educational evaluation, trying to clarify its meaning and exposing the distinction between evaluation and other related concepts. This article reviews the evaluation literature through an analytical framework representing issues addressed by major evaluation approaches in education. The analytical framework is comprised of 10 dimensions referring to: (1) the definition of evaluation, (2) its functions, (3) the objects of evaluation, (4) the variables that should be investigated, (5) criteria that should be used, (6) the audiences that should be served, (7) the process of doing an evaluation, (8) its methods of inquiry, (9) the characteristics of the evaluator, and (10) the standards that should be used to judge the worth and merit of an evaluation. Some implications for the advancement of evaluation theory and practice conclude this review of the literature.

Many attempts have been made in recent years to clarify the meaning of evaluation and expose the distinction between evaluation and other related concepts such as measurement or research. The literature contains many approaches regarding the conceptualization of evaluation and the determination of its countenance in education. Many of those approaches have been unduly referred to as "models" (e.g., the CIPP Model, the Discrepancy Model, the Responsive Model, or the Goal-Free Model) in spite of the fact that none of them includes a sufficient degree of complexity and completeness that might be suggested by the term "model." Stake (1981) rightly suggested that they be referred to as persuasions rather than models.

For the benefit of those of us who lost their way between the various evaluation models, approaches, and persuasions, several attempts have been made to put some order into the growing evaluation literature through classifications of evaluation approaches. Such classifications (e.g., Guba & Lincoln, 1981; House, 1980; Popham, 1975; Stake, 1976; Stufflebeam & Webster, 1980; Worthen & Sanders, 1973) made a significant contribution through their critical reviews of the evaluation literature denoting similarities and differences among the various approaches. Those classifications were based on a somewhat holistic approach by placing each evaluation model as a whole in one of the labeled categories with some other models. Trying to do justice to each evaluation model as a whole they sometimes ignored the major issues underlying the agreements and disagreements among the various evaluation approaches.

Stufflebeam (1974) suggested eight questions to be addressed in any attempt to conceptualize evaluation. Nevo (1981) revised Stufflebeam's list of questions and extended it to 10 major dimensions in a conceptualization of evaluation. These 10

dimensions represent the major issues addressed by the most prominent evaluation approaches in education. They will be used here as an organizer for an analytical review of the literature on educational evaluation.

The 10 dimensions for our analysis are expressed by the following questions:

1. How is evaluation defined?
2. What are the functions of evaluation?
3. What are the objects of evaluation?
4. What kinds of information should be collected regarding each object?
5. What criteria should be used to judge the merit and worth of an evaluated object?
6. Who should be served by an evaluation?
7. What is the process of doing an evaluation?
8. What methods of inquiry should be used in evaluation?
9. Who should do evaluation?
10. By what standards should evaluation be judged?

We shall review the literature seeking the various answers to those questions provided by the various evaluation models, approaches, and persuasions. The significance of such a review for evaluation practitioners as well as evaluation theoreticians and researchers will be pointed out at the conclusion of the article.

1. *How is evaluation defined*? Many definitions of evaluation can be found in the literature. The well-known definition originated by Ralph Tyler perceives evaluation as "The process of determining to what extent the educational objectives are actually being realized" (Tyler, 1950, p. 69). Another widely accepted definition of evaluation has been that of providing information for decisionmaking suggested by various leading evaluators such as Cronbach (1963), Stufflebeam (Stufflebeam et al., 1971), and Alkin (1969). In recent years considerable consensus has been reached among evaluators regarding the definition of evaluation as the assessment of merit or worth (Eisner, 1979; Glass, 1969; House, 1980; Scriven, 1967; Stufflebeam, 1974), or as an activity comprised of both description and judgment (Guba & Lincoln, 1981; Stake, 1967). A joint committee on standards for evaluation, comprised of 17 members representing 12 organizations associated with educational evaluation, recently published their definition of evaluation as "the systematic investigation of the worth or merit of some object" (Joint Committee, 1981, p. 12).

A major exception to that consensus regarding the judgmental definition of evaluation is represented by the Stanford Evaluation Consortium group who defined evaluation as "[a] systematic examination of events occurring in and consequent of a contemporary program—an examination conducted to assist in improving this program and other programs having the same general purpose" (Cronbach et al., 1980, p. 14). Cronbach and his associates (1980) clearly reject the judgmental nature of evaluation advocating an approach that perceives the evaluator as "an educator [whose] success is to be judged by what others learn" (p. 11) rather than a "referee [for] a basketball game" (p. 18) who is hired to decide who is "right" or "wrong".

A definition that points to the judgmental character of evaluation might create considerable anxiety among potential evaluees and raise resistance among opponents of evaluation. Obviously, a nonjudgmental definition of evaluation, such as "providing information for decisionmaking," might be accepted more favorably by

evaluees and clients. However, it may be unrealistic to create positive attitudes toward evaluation by ignoring one of its major features. Another approach intended to develop positive attitudes toward evaluation might be to demonstrate its constructive functions within the various domains of education.

2. *What are the functions of evaluation*? Scriven (1967) was the first to suggest the distinction between "formative evaluation" and "summative evaluation," referring to two major roles or functions of evaluation, although he was not the first to realize the importance of such a distinction. Later, referring to the same two functions, Stufflebeam (1972) suggested the distinction between proactive evaluation intended to serve decisionmaking and retroactive evaluation to serve accountability. Thus, evaluation can serve two functions, the "formative" and the "summative." In its formative function evaluation is used for the improvement and development of an ongoing activity (or program, person, product, etc.). In its summative function evaluation is used for accountability, certification, or selection.

A third function of evaluation, the psychological or sociopolitical function, which has been less often treated by evaluation literature (Cronbach et al., 1980; House, 1974; Patton, 1978), should also be considered. In many cases it is apparent that evaluation is not serving any formative purposes nor is it being used for accountability or other summative purposes. However, it is being used to increase awareness of special activities, motivate desired behavior of evaluees, or promote public relations. Regardless of our personal feelings about the use (or misuse) of evaluation for this purpose, we cannot ignore it.

Another somewhat "unpopular" function of evaluation is its use for the exercise of authority (Dornbusch & Scott, 1975). In formal organizations it is the privilege of the superior to evaluate his or her subordinates and not vice versa. In many cases a person in a management position might evaluate someone to demonstrate his authority over that person. We may refer to this as the "administrative" function of evaluation.

To summarize, evaluation can serve many functions: (a) the formative function for improvement, (b) the summative function for selection, for certification, for accountability, (c) the psychological or sociopolitical function for motivation and to increase awareness, and (d) the administrative function to exercise authority.

Some evaluators (Alkin, Daillak, & White, 1979; Cronbach et al., 1980) express a clear preference for the formative function of evaluation, but the general perception seems to be that there are no "right" or "wrong" roles of evaluation, and that it can serve deliberately more than one function. However, different functions can be served in various ways and by different evaluation methods. It is therefore important to realize the existence of the various evaluation functions and to determine the specific function(s) of a concrete evaluation at an early stage of its planning.

3. *What are the objects of evaluation*? Students and teachers have always been popular objects of evaluation in education. Almost all the measurement and evaluation literature in education up to the mid-sixties dealt with the evaluation of students' learning. Up to that time one could hardly find in the educational literature any substantial guidance regarding the evaluation of other objects such as educational projects or programs, curricular materials, or educational institutions. Various developments in the educational system of the United States (e.g.,

the Elementary and Secondary Education Act of 1965) led to a significant shift of focus regarding the objects of educational evaluation from students to projects, programs, and instructional materials, which have been since then most common in the writings of the major authors in the evaluation literature in education (Alkin, 1969; Provus, 1971; Scriven, 1967; Stake, 1967; Stufflebeam, 1969; Stufflebeam et al., 1971).

Two major conclusions can be drawn from the review of contemporary evaluation literature: (a) Almost everything can be an object of evaluation, and evaluation should not be limited to the evaluation of students or school personnel; and (b) the clear identification of the evaluation object is an important part of the development of any evaluation design.

In planning an evaluation it seems to be important to determine what is "the thing" (or "the evaluand," to use Scriven's, 1980, term) that has to be evaluated. It helps to determine what kind of information should be collected and how it should be analysed. A clear object identification helps keep an evaluation focused. It also helps to clarify and resolve value conflicts and potential threat among stakeholders and others likely to be affected by the evaluation (Guba & Lincoln, 1981).

4. *What kinds of information should be collected regarding each object*? After an evaluation object has been chosen, a decision must be made regarding the various aspects and dimensions of the object that should be evaluated. Information pertinent to such aspects must be collected. Earlier approaches to evaluation focused mainly on results or outcomes. Thus, to evaluate an educational object (e.g., a new curriculum) would mean to evaluate the quality of the results of its functioning (e.g., students' achievements). In recent years some interesting attempts have been made to extend the scope of evaluation variables in various evaluation models (Alkin, 1969; Provus, 1971; Stake, 1967; Stufflebeam, 1969, 1974; Stufflebeam et al., 1971). Stufflebeam's CIPP Model suggests that evaluation focus on four variables for each evaluation object; (a) its goals, (b) its design, (c) its process of implementation, and (d) its outcomes. According to this approach an evaluation of an educational project, for example, would be an assessment of (a) the merit of its goals, (b) the quality of its plans, (c) the extent to which those plans are being carried out, and (d) the worth of its outcomes.

Stake (1967) in his Countenance Model suggested that two sets of information be collected regarding the evaluated object: descriptive and judgmental. The descriptive set should focus on intents and observations regarding antecedents (prior conditions that may affect outcomes), transactions (the process of implementation), and outcomes. The judgmental set of information is comprised of standards and judgments regarding the same antecedents, transactions and outcomes.

Guba and Lincoln (1981), expanding Stake's Responsive Education Model (Stake, 1975) and applying the naturalistic paradigm, suggest that the evaluator generate five kinds of information: (a) descriptive information regarding the evaluation object, its setting, and its surrounding conditions, (b) information responsive to concerns of relevant audiences, (c) information about relevant issues, (d) information about values, and (e) information about standards relevant to worth and merit assessments.

Thus, the evaluation literature seems to suggest that a wide range of information

should be collected by evaluation regarding the evaluated object. It should not limit itself to the narrow scope of evaluation regarding outcomes or results. This does not mean that each single evaluation must always collect all possible kinds of information; it may focus on some of them according to identified evaluation priorities or practical constraints.

5. *What criteria should be used to judge the merit and worth of an evaluation object?* To choose the criteria to be used to judge the merit of an evaluation object is one of the most difficult tasks in educational evaluation. Those who think that evaluation should attempt to determine whether goals have been achieved (Provus, 1971; Tyler, 1950) make this task easy for themselves by partially ignoring the issue of evaluation criteria. What they actually do is use "goal achievement" as the evaluation criterion without having justified its being an appropriate criterion. What about trivial goals or all kinds of "stated objectives" that aren't worth achieving? Should they be used as evaluation criteria?

Another way to avoid the issue of evaluation criteria is to ignore the judgmental nature of evaluation. Those who defined evaluation as an information collection activity to serve decisionmaking or other purposes (Alkin, 1969; Cronbach, 1963; Stufflebeam, 1969) did not have to deal with the problem of choosing evaluation criteria.

Apparently, the achievement of (important) goals is one possible basis for evaluation criteria. Alternative bases for evaluation criteria suggested by the literature might be: identified needs of actual and potential clients (Joint Committee, 1981; Patton, 1978; Scriven, 1972b), ideals or social values (Guba & Lincoln, 1981; House, 1980), known standards set by experts or other relevant groups (Eisner, 1979; Guba & Lincoln, 1981; Stake, 1967), or the quality of alternative objects (House, 1980; Scriven, 1967).

Most evaluation experts seem to agree that the criterion (or criteria) to be used for the assessment of a specific object must be determined within the specific context of the object and the function of its evaluation. While in many cases the evaluator does not have the authority to choose among the various alternative criteria, it is the evaluator's responsibility that such a choice be made and that he be able to provide a sound justification for the choice, whether it is made by him or by somebody else.

6. *Who should be served by an evaluation?* Those who define evaluation as providing information for decisionmaking (Alkin, 1969; Cronbach, 1963; Stufflebeam et al., 1971) seem to have a clear opinion as to who has to be served by evaluation. They identify the relevant decisionmakers and attempt to determine their information needs. Others (Cronbach et al., 1980; House, 1980) reject the notion of serving "decisionmakers" because of the threat of co-optation or oversimplification of social and organizational processes. Cronbach and his associates (1980) are inclined to serve the "policy-shaping community" rather than some kind of managerial decisionmaker. Many authors refer to "evaluation clients" or "evaluation audiences" as those who have to be served by evaluation. Guba and Lincoln (1981) suggested the term "stakeholders" or "stakeholding audience" for the whole group of persons having some stake in the performance of the evaluand and therefore should be served by the evaluation.

If evaluation is to be useful at all, it has to be useful to some specific client or

audience. The evaluation literature does not suggest which is the "most appropriate" audience for evaluation, but three important propositions can be found in writings regarding this issue. They are: (a) An evaluation can have more than one client or audience; (b) different evaluation audiences might have different evaluation needs; and (c) the specific audiences for an evaluation and their evaluation needs must be clearly identified at the early stages of planning an evaluation.

Differences in evaluation needs might be reflected in many ways: for example, the kind of information to be collected, the level of data analysis to be used, or the form of reporting the evaluation results. Sometimes it is impossible to serve all identified evaluation needs at the same time, and some priorities have to be set regarding the specific evaluation needs to which the evaluation will respond.

7. *What is the process of doing an evaluation*? The process of doing an evaluation might differ according to the theoretical perception guiding the evaluation. A theoretical approach perceiving evaluation as an activity intended to determine whether goals have been achieved (Tyler, 1950) might recommend the following evaluation process: (a) stating goals in behavioral terms, (b) developing measurement instruments, (c) collecting data, (d) interpreting findings, and (e) making recommendations.

According to Stake's Countenance Model (Stake, 1967) the evaluation process should include (a) describing a program, (b) reporting the description to relevant audiences, (c) obtaining and analyzing their judgments, and (d) reporting the analyzed judgments back to the audiences. Later on, in his Responsive Evaluation Model Stake (1975) suggested a continuing "conversation" between the evaluator and all other parties associated with the evaluand. He specified 12 steps of dynamic interaction between the evaluator and his audiences in the process of conducting an evaluation.

Provus (1971) proposed a five step evaluation process including (a) clarification of the program design, (b) assessing the implementation of the program, (c) assessing its in-term results, (d) assessing its long-term results, and (e) assessing its costs and benefits.

The Phi Delta Kappa Study Committee on evaluation (Stufflebeam et al., 1971) presented a three-step evaluation process. It included (a) delineating information requirements through interaction with the decisionmaking audiences, (b) obtaining the needed information through formal data collection and analysis procedures, and (c) providing the information to decisionmakers in a communicable format.

Scriven (1972a) has suggested nine steps in his Pathway Comparison Model. Guba and Lincoln (1981) suggest in their recently published book that a naturalistic-responsive evaluation be implemented through a process including the following four stages: (a) initiating and organizing the evaluation, (b) identifying key issues and concerns, (c) gathering useful information, and (d) reporting results and making recommendations.

While there seems to be no agreement among evaluation experts regarding the "best" process to follow when conducting an evaluation, most of them would agree that all evaluations should include a certain amount of interaction between evaluators and their audiences at the outset of the evaluation to identify evaluation needs, and at its conclusion to communicate its findings. Evaluation cannot be limited to the technical activities of data collection and analysis.

8. *What methods of inquiry should be used in evaluation?* While challenging the usefulness of various research methods for evaluation studies (Guba, 1969; Stufflebeam et al., 1971), recent years have also introduced various methods of inquiry into the field of educational evaluation. In addition to traditional experimental and quasi-experimental designs (Campbell, 1969; Stanley, 1972; Cook & Cambell, 1976), naturalistic methods (Guba & Lincoln, 1981; Patton, 1980), jury trials (Wolf, 1979), case studies (Stake, 1978), art criticism (Eisner, 1977, 1979), journalistic methods (Guba, 1978), the modus operandi method (Scriven, 1974), and many others became legitimate methods for the conduct of evaluation. Some methodologists still advocate the superiority of certain methods such as experimental design (Boruch & Cordray, 1980; Rossi, Freeman, & Wright, 1979) at one extreme, or naturalistic methods (Guba & Lincoln, 1981; House, 1980; Patton, 1980) on the other extreme, but overall there seems to be more support for a more eclectic approach to evaluation methodology. At the present state of the art in evaluation it looks like "the evaluator will be wise not to declare allegiance to either a quantitative-scientific-summative methodology or a qualitative-naturalistic-descriptive methodology" (Cronbach et al., 1980, p. 7). It might be also true that for a complicated task such as the conduct of evaluation an approach is needed that seeks the best method or set of methods for answering a particular evaluation question, rather than assuming that one method is best for all purposes.

9. *Who should do evaluation?* Becoming a professional group, evaluators devoted much attention to identifying the characteristics of "good" evaluators and appropriate ways to train them (Boruch & Cordray, 1980; Cronbach et al., 1980; Guba & Lincoln, 1981; Stufflebeam et al., 1971; Worthen, 1975). To be a competent and trustworthy evaluator one needs to have a combination of a wide variety of characteristics. These include technical competence in the area of measurement and research methods, understanding the social context and the substance of the evaluation object, human relations skills, personal integrity, and objectivity, as well as characteristics related to organizational authority and responsibility. Because it is difficult to find one person possessing all these qualifications, it often becomes necessary to have a team conduct an evaluation or to choose the person with the most appropriate characteristics for a specific evaluation task.

The evaluation literature also suggests two important distinctions that should be taken into account when deciding who should do an evaluation. The first is the distinction between an internal evaluator and an external evaluator (Scriven, 1967, 1975; Stake & Gjerde, 1974; Stufflebeam et al., 1971). An internal evaluator of a project is usually one who is employed by the project and reports directly to its management. Obviously, the internal evaluator's objectivity as well as external credibility might be lower than those of an external evaluator, who is not directly employed by the project and/or enjoys a higher degree of independence.

The second distinction is between a professional evaluator and an amateur evaluator. This distinction, suggested by Scriven (1967), refers to two different foci of training and expertise rather than to a value judgment regarding the quality of an evaluation. An amateur evaluator is usually one whose major professional training is not in evaluation, and involvement in evaluation represents only part of his or her job description (e.g., the associate director of a new math curriculum development project conducting the formative evaluation of the project, who has

an M.A. in math education and some on-the-job training in evaluation). A professional evaluator is one with extensive training in evaluation and whose major (or even only) responsibility is conducting evaluation (e.g., the internal evaluator of a special education project, who has an M.A. in measurement and evaluation and 5 years experience evaluating special education projects). While the amateur evaluator's technical evaluation skills might be lower than those of a professional evaluator, he or she might have a better understanding of the project's unique evaluation needs and be able to develop better rapport with the members of the evaluated project.

These two distinctions are independent; there may be an internal-amateur evaluator, an external-amateur evaluator, an internal-professional evaluator, and so forth.

10. *By what standards should evaluation be judged*? Several attempts have been made in recent years to develop standards for evaluations of educational and social programs (Evaluation Research Society, 1980; Joint Committee, 1981; Stufflebeam et al., 1971; Tallmadge, 1977; U.S. General Accounting Office, 1978). In spite of the fact that some writers (Cronbach et al., 1980; Stake, 1981) have criticized the rationale for the whole standard-setting effort as being premature at the present state of the art in evaluation, there seems to be a great deal of agreement regarding their scope and content.

Boruch and Cordray (1980) analyzed six sets of such standards and reached the conclusion that there has been a large degree of overlap and similarity among them. The most elaborate and comprehensive set, and the one based on the largest amount of consensus, is probably the set developed and published by the Joint Committee on Standards for Educational Evaluation (1981). These standards have been developed by a committee of 17 members, chaired by Dr. Daniel Stufflebeam, which represented 12 professional organizations associated with educational evaluation. The committee suggested 30 standards, which are divided into four major groups: utility standards (to ensure that evaluation serves practical information needs), feasibility standards (to ensure that evaluation is realistic and prudent), propriety standards (to ensure that evaluation is conducted legally and ethically), and accuracy standards (to ensure that evaluation reveals and conveys technically adequate information).

Summary

Risking oversimplification, one could summarize the review of the literature with the following most common answers to our 10 questions. This could be one way to describe briefly the state of the art in the conceptualization of educational evaluation.

1. *How is evaluation defined*? Educational evaluation is a systematic description of educational objects and/or an assessment of their merit or worth.

2. *What are the functions of evaluation*? Educational evaluation can serve four different functions: (a) formative (for improvement); (b) summative (for selection and accountability); (c) sociopolitical (to motivate and gain public support); and (d) administrative (to exercise authority).

3. *What are the objects of evaluation*? Any entity can be an evaluation object. Typical evaluation objects in education are students, educational and administra-

tive personnel, curricula, instructional materials, programs, projects, and institutions.

4. *What kinds of information should be collected regarding each object*? Four groups of variables should be considered regarding each object. They focus on (a) the goals of the object; (b) its strategies and plans; (c) its process of implementation; and (d) its outcomes and impacts.

5. *What criteria should be used to judge the merit of an object*? The following criteria should be considered in judging the merit or worth of an educational object: (a) responding to identified needs of actual and potential clients; (b) achieving national goals, ideals, or social values; (c) meeting agreed-upon standards and norms; (d) outdoing alternative objects; and (e) achieving (important) stated goals of the object. Multiple criteria should be used for any object.

6. *Who should be served by an evaluation*? Evaluation should serve the information needs of all actual and potential parties interested in the evaluation object ("stakeholders"). It is the responsibility of the evaluator(s) to delineate the stakeholders of an evaluation and to identify or project their information needs.

7. *What is the process of doing an evaluation*? Regardless of its method of inquiry, an evaluation process should include the following three activities: (a) focusing the evaluation problem; (b) collecting and analyzing empirical data; and (c) communicating findings to evaluation audiences. There is more than one appropriate sequence for implementing these activities, and any such sequence can (and sometimes should) be repeated several times during the life span of an evaluation study.

8. *What methods of inquiry should be used in evaluation*? Being a complex task, evaluation needs to mobilize many alternative methods of inquiry from the behavioral sciences and related fields of study and utilize them according to the nature of a specific evaluation problem. At the present state of the art, an a priori preference for any specific method of inquiry is not warranted.

9. *Who should do evaluation*? Evaluation should be conducted by individuals or teams possessing (a) extensive competencies in research methodology and other data analysis techniques; (b) understanding of the social context and the unique substance of the evaluation object; (c) the ability to maintain correct human relations and to develop rapport with individuals and groups involved in the evaluation; and (d) a conceptual framework to integrate the above-mentioned capabilities.

10. *By what standards should evaluation be judged*? Evaluation should strike for an optimal balance in meeting standards of (a) utility (to be useful and practical); (b) accuracy (to be technically adequate); (c) feasibility (to be realistic and prudent); and (d) propriety (to be conducted legally and ethically).

Conclusion

As stated at the beginning of this article, a critical analysis of the various theoretical approaches to educational evaluation might have important implications for practitioners of evaluation as well as for theoreticians and researchers who are concerned with developing new concepts and better methods. All of them could benefit from the analytical scheme of the 10 questions, which guided our analysis, as well as from the review of the answers contained in the evaluation literature.

Evaluators could use the 10 questions to organize their own perceptions of evaluation using the evaluation literature to develop their own sets of coherent answers for the 10 questions rather than adopting piously one evaluation model or another. Understanding what others mean when they refer to evaluation could be another use of the 10 questions. Evaluators may encounter considerable difficulties if their perceptions of a concrete evaluation differ from those of their clients and audiences. It is appropriate before one starts planning an evaluation or even decides to do it at all to find out what is meant by evaluation by the various parties involved in the evaluation, what purpose it is intended to serve, what is to be evaluated, what are some feasible alternatives for doing it, and by what standards the evaluation is to be judged if it is to be conducted at all. In other words, addressing the 10 questions discussed in this article might help evaluators to develop a clear perception of evaluation and to get a better grasp of their evaluation problems before they get themselves into all kinds of dubious evaluation adventures.

Discussions among theoreticians of evaluation can be a fruitful contribution to the advancement of evaluation theory and practice. It could be even more so if those discussions focused on issues in disagreement rather than on competing models and paradigms. The contribution would be even more robust if the various theoretical propositions were substantiated by some research findings. The 10 questions reviewed here could provide a framework to delineate research variables for an empirical study of evaluation. Data on the actual relationships among those variables as well as their relationships with other variables (e.g., evaluation utilization or variables reflecting the context of evaluation) would be very much appreciated by the evaluation profession.

References

Alkin, M. C. Evaluation theory development. *Evaluation Comment,* 1969, *2,* 2–7.

Alkin, M. C., Daillak, R., & White, P. *Using evaluations: Does evaluation make a difference?* Beverly Hills, Calif.: Sage, 1979.

Boruch, F. R., & Cordray, D. S. *An appraisal of educational program evaluations: Federal, state, and local agencies.* Evanston, Ill.: Northwestern University, 1980.

Campbell, D. T. Reforms as experiments. *American Psychologist,* 1969, *24,* 409–429.

Cook, T. D., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology.* Chicago: Rand McNally, 1976.

Cronbach, L. J. Course improvement through evaluation. *Teachers College Record,* 1963, *64,* 672–683.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. E., & Weiner, S. S. *Toward reform of program evaluation.* San Francisco: Jossey-Bass, 1980.

Dornbusch, S. M., & Scott, W. R. *Evaluation and the exercise of authority.* San Francisco: Jossey-Bass, 1975.

Eisner, E. W. On the uses of educational connoisseurship and educational criticism for evaluating classroom life. *Teachers College Record,* 1977, *78,* 345–358.

Eisner, E. W. *The educational imagination.* New York: Macmillan, 1979.

Evaluation Research Society. *Standards for program evaluation.* Author, 1980.

Glass, G. V. *The growth of evaluation methodology* (Research Paper No. 27). Boulder: Laboratory of Educational Research, University of Colorado, 1969. (Mimeo)

Guba, E. G. The failure of educational evaluation. *Educational Technology,* 1969, *9,* 29–38.

Guba, E. G. *Metaphor adaptation report: Investigative journalism.* Research on Evaluation Project. Portland, Ore.: Northwest Regional Educational Laboratory, 1978. (Monograph)

Guba, E. G., & Lincoln, Y. S. *Effective evaluation.* San Francisco: Jossey-Bass, 1981.

House, E. R. *The politics of educational innovation.* Berkeley, Calif.: McCutchan, 1974.

House, E. R. *Evaluating with validity.* Beverly Hills, Calif.: Sage, 1980.

Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill, 1981.

Nevo, D. The evaluation of a multi-dimensional project. In A. Lewy, S. Kugelmass, G. Ben-Shakar, N. Blass, R. F. Boruch, D. J. Davis, B. Nevo, D. Nevo, P. Tamir, & I. Zak, *Decision oriented evaluation in education: The case of Israel.* Philadelphia: International Science Services, 1981.

Patton, M. Q. *Utilization focused evaluation.* Beverly Hills, Calif.: Sage, 1978.

Patton, M. Q. *Qualitative evaluation methods.* Beverly Hills, Calif.: Sage, 1980.

Popham, W. J. *Educational evaluation.* Englewood Cliffs, N. J.: Prentice-Hall, 1975.

Provus, M. M. *Discrepancy evaluation.* Berkeley, Calif.: McCutchan, 1971.

Rossi, P. H., Freeman, H. E., & Wright, S. R. *Evaluation: A systematic approach.* Beverly Hills, Calif.: Sage, 1979.

Scriven, M. The methodology of evaluation. In R. E. Stake (Ed.), *AERA monograph series on curriculum evaluation,* No. 1. Chicago: Rand McNally, 1967.

Scriven, M. Pros and cons about goal-free evaluation. *Evaluation Comment,* 1972, *3*(4), (b)

Scriven, M. *The pathway comparison model of evaluation.* Berkeley: University of California, January 1972. (a) (mimeo).

Scriven, M. Maximizing the power of causal investigations: The modus operandi method. In W. J. Popham (Ed.), *Evaluation in education.* Berkeley, Calif.: McCutchan, 1974.

Scriven, M. *Evaluation bias and its control* (Occasional Paper No. 4). Kalamazoo: Western Michigan University, 1975.

Scriven, M. *Evaluation thesaurus* (2nd ed.). Inverness, Calif.: Edgepress, 1980.

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

Stake, R. E. (Ed.). *Evaluating the arts in education: A responsiveness approach.* Columbus, Ohio: Merrill, 1975.

Stake, R. E. *Evaluating educational programmes: The need and the response.* Washington, D. C.: OECD Publications Center, 1976.

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

Stake, R. E. Setting standards for educational evaluators. *Evaluation News,* 1981, *2*(2), 148–152.

Stake, R. E., & Gjerde, C. An evaluation of TCITY, the Twin City Institute for Talented Youth. In Fred N. Kerlinger (Ed.), *AERA monograph series in curriculum evaluation,* No. 7. Chicago: Rand McNally, 1974.

Stanley, J. C. Controlled field experiments as a model for evaluation. In P. H. Rossi & W. Williams (Eds.), *Evaluating social programs.* New York: Seminar Press, 1972.

Stufflebeam, D. L. Evaluation as enlightenment for decision making. In W. H. Beatty (Ed.), *Improving educational assessment and an inventory for measures of affective behavior.* Washington, D. C.: National Education Association, 1969.

Stufflebeam, D. L. The relevance of the CIPP evaluation model for educational accountability. *SRIS Quarterly,* 1972, *5,* 3–6.

Stufflebeam, D. L. *Meta-evaluation* (Occasional Paper No. 3). Kalamazoo: Western Michigan University, December 1974.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammon, R. L., Merriman, H. O., & Provus, M. M. *Educational evaluation and decision-making.* Itasca, Ill.: Peacock, 1971.

Stufflebeam, D. L., & Webster, W. J. An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis,* 1980, *2*(3), 5–20.

Tallmadge, G. K. *Joint Dissemination Review Panel ideabook.* Washington, D. C.: U. S. Government Printing Office, 1977.

Tyler, R. W. *Basic principles of curriculum and instruction.* Chicago, Ill.: University of Chicago Press, 1950.

U. S. General Accounting Office. *Assessing social program impact evaluations: A checklist approach.* Washington, D. C.: U. S. General Accounting Office, 1978.

Wolf, R. L. The use of judicial evaluation methods in the formation of educational policy. *Educational Evaluation and Policy Analysis,* 1979, *1,* 19–28.

Worthen, B. R. Competencies for educational research and evaluation. *Educational Researcher,* 1975, *4*(1), 13–16.

Worthen, B. R., & Sanders, J. R. *Educational evaluation: Theory and practice.* Belmont, Calif.: Wadsworth, 1973.

AUTHOR

DAVID NEVO, Senior Lecturer, School of Education, Tel-Aviv University, Tel-Aviv 69978, Israel. *Specialization:* Evaluation theory, measurement and research methodology.