

## CURRENT PROBLEMS AND ISSUES IN EVALUATION\*

R. Emile Jester

*University of Florida*

The issues in evaluation listed by Dr. Cahen are certainly timely and appropriate. His definition stresses the importance of the value or the worth of something as central to evaluation. He defines evaluation as "a rational process of reaching decisions about the worth of something." He then goes on to state that the "something" may be the quality of a science program or curriculum, a film strip, a new model of a tape recorder, teaching method A, etc. His stress on values is an important one and should be kept in mind during this discussion. In addition, he places great importance on the decision-making aspect of evaluation. It seems to me that this point is the most critical; that is, that the most important aspect of evaluation is decision making. Cahen cites Scriven's distinction between formative and summative evaluation. I wonder how important that distinction really is. It seems to me that the only real importance to be ascribed to the distinction lies in how the two form parts of a continuous sequence in evaluating, decision-making, and subsequent modification, evaluation, etc.

The discussion of the affective dimensions in evaluation reflects a current and popular concern of many educators. Webster's Collegiate Dictionary defines affective as "relating to, arising from, or influencing feelings or emotions: emotional." Affective in this sense may also refer to expressing emotion. If this is the kind of definition being used for the affective component in this discussion then I suppose what Cahen is calling for is an evaluation of the emotional side of education. It is possible, of course, that the emotional aspect deserving evaluation is only that which may be thought of as producing negative outcomes. That is, the emotional kinds of activity in the classroom producing ill feelings or bad feelings towards the material being learned. If this is the case then perhaps affective is the appropriate term.

\*Discussion of the invited address by Leonard S. Cahen, Florida Educational Research Association—NCME meetings at Jacksonville, Florida, January 23, 1970.

There is a distinct possibility that the affective dimensions refer more to attitudes of the person doing the evaluation rather than to the emotional side-effect produced upon children in the educational system. For example, I have been working in a number of pre-school programs. Many persons involved in the education of the very young child attach an evaluation of "not good" to programs designed to teach children in the age range of three or four years. The reason for this "not good" evaluation seems to be that some people think it is mean or harsh or unkind to actively teach the children, and that the children should be left to their own devices so that they can discover the features of their environment. Persons making evaluations in these terms probably consider the preschool to be a place where the child learns to "socialize" and to be "happy" with no regard for the possibility that he could also be learning to read or write, or perhaps even do fairly complex arithmetic operations. The whole point of the nursery school to such evaluators apparently is to avoid complicating the youngster's life under the premise or the assumption that such complications will produce an abnormally adjusted child. It is implicit that if one actively intervenes in the youngster's life it will automatically produce a youngster who is affectively disturbed. The important point here is that evaluation of programs need to be made in terms of previously established criteria rather than in terms of the biases of the evaluator.

Cahen cites observation as a way of identifying side-effects of a program. One of the side-effects, of course, relates to the affective dimension of classroom instruction. As suggested earlier, if the most important affective side-effect belongs or resides within the observer, rather than as a function of the observed, then it is entirely possible that the observer's biases during the process of evaluation might overcome any objectivity inherent in the observation system. The fact that the observations are often played down as too soft may be related in part, to the fact that so often the observer has been asked to infer the psychological attributes; for example, happiness, adequate self, anxiety, etc. on the basis of his observations. This kind of data is generally nonreplicable and must be considered very soft indeed. If the observations have been made in the tradition of those observers who have concerned themselves with systematic counts of occurrence of clearly observable behavior, then the observations are probably somewhat more reliable. I think it is important that this distinction be made with respect to the softness or hardness of observations.

The issue of comparative evaluation raises some very interesting points. Simply comparing the application of curriculum A with curriculum B; or method A with method B seems to me to be a relatively useless activity. As Cahen stresses when he quotes Cronbach it is usually the case that the two methods or two curricula have different goals and different expected outcomes and therefore comparison is logically indefensible. Often, such comparisons are done as an effort to "prove" that the new innovative method is somehow superior to the old traditional method.

Some years back a candidate for a Masters' degree came to me for help on his thesis. It seems that he was having trouble "proving" that his method of reading instruction was better than the traditional method. He wanted me to come up with a statistical technique that would prove his point. I didn't, but he went on to write in his thesis that since his method did not produce significant differences, that it was "as good as" the traditional method and therefore, since it was innovative, it should be preferred. At this point I took my money and withdrew back to my cell to finish my work. I did notice that this person was awarded his Master's degree, that his thesis was passed on just as it had been shown to me, and, as nearly as I could determine, he had not only proved nothing, he had probably been asking the wrong questions. He was clearly out to prove that he could do better than the old hands at reading instruction, but had failed to adequately determine how or what it was he was going to do better. Had he specified that the children learning to read under method A would be able to read as well as those taught by method B and in addition would like to read more or would have more reading appreciation, he might have demonstrated a superiority in his innovation.

Only when the evaluator starts out with an adequate specification of what is expected as an outcome from a program can he make a decision as to whether or not that outcome has been reached. Once again the business of stating goals specifically and operationally becomes a major issue. *Only when the goals have been precisely stated is it possible to produce a meaningful evaluation.* It must be remembered that evaluation as being used here is primarily for the purpose of decision-making relating to one particular program. Does the program reach, or does it not reach the goals it set out to reach.

In making any comparative evaluation, it seems obvious that the evaluator is now approaching what may be called "research" and this should be distinguished from evaluation. The reason for the distinction is simply because the rules are different. In performing an evaluative function the primary task is to decide if the prescribed goals have been reached. Research involves much more, and, it seems to me, eventually leads to "scientific" activities including the construction of systems, models, and theories. Evaluation has no real need for this process of research although as Cahen mentioned it certainly requires the rigor. The essential difference between research and evaluation does not lie in the degree of rigor required of research versus that required of evaluation; both, if they are to be considered valuable, require a high degree of rigor. The difference lies in the expected outcome and procedures for attaining the outcomes of either evaluation or research. Neither is "better" or "worse" than the other—they are simply different activities.

The kind of comparing that Cahen seems to have done with research and evaluation is a bit like comparing plumbers and electricians. As soon as an attempt is made to decide which is more rigorous, more important, more valuable, or more skillful, then nonrelevant issues are being raised. Questions being asked are completely irrelevant to the issue. Neither is better than the other, they are both important but they are different. It seems to me that the same applies to evaluation and research. Both activities are important. I am somewhat concerned that Cahen has felt the necessity to cite the words of Stake and Denny when he says: "The evaluators sacrifice the opportunity to manipulate and control (a basic in research endeavor) but evaluation gains relevance to the immediate situation." Why is it so important that the evaluator has "sacrificed the opportunity?" If the evaluation is proceeding satisfactorily, it has no need of manipulation and control but only of assessing the degree to which a set of objectives has been obtained. The goal of the evaluator, it seems to me, is not to attempt to do everything to everybody but rather to perform as adequately as possible the decision-making process implied in the title of "evaluator."