Basic to improved instruction is

Sound Evaluation

A BOARD member and his superintendent review school-wide testing data; several parents exchange impressions about their children's schooling, or sign—with proper admonishments—report cards; teachers exchange opinions about classes they have known. Pupils are dubbed dull or bright; textbooks are rated adequate or difficult. Whether to review instructional materials for classroom, choose test items, select leaders, or set up classroom subgroups, each person performs an act of evaluation.

A counselor interprets information about a counselee. Administrators weigh plans for some instructional innovation, alter budget allocations, hire teachers, advance others to new responsibilities. Every educational act will be related in some way to the process of evaluation.

A counselor interprets information about a counselee. Administrators weigh plans for some instructional innovation, alter budget allocations, hire teachers, advance others to new responsibilities. Every educational act will be related in some way to the process of evaluation.

What are the requirements for sound evaluation? Here is a question basic to all educational planning. It is one for which simple answers, single-dimensional ones, mislead—but are often proposed.

Characteristics of Sound Evaluation

"Evaluation includes the entire process through which teachers attempt to appraise pupil achievement and pupil growth. . . . Evaluation is an inclusive term which embraces all kinds of objective measures and subjective appraisal" (1). A central task of education is effective choice making. "Where choices are to be made, evaluation plays an important part," explains Raths (2). Taba places evaluation in the mainstream of activities that expedite the educational process (3). Given impetus by the challenges of the Eight-Year Study, educators Ralph Tyler, Hilda Taba, Maurice Troyer, Louis Raths and others proposed a concept of evaluation which stressed purpose, system, comprehensiveness and continuity.

Whether designed for a single classroom or for a system-wide assessment, evaluation is more than a battery of instruments, an array of standardized scores or a collection of data and records. While evaluation depends on measurement, its concern is with a
broader profile of characteristics and attainments.

Adequate evaluation necessitates the use of each part of the evaluating process without overemphasis on any one aspect. A program heavy with external testing or geared to routine weekly quizzes, for example, cannot be called sound according to criteria of consistency or comprehensiveness. To develop an evaluation program, several questions need answering (4):

1. What are the aims of the program to be evaluated? What specific behaviors will pupils demonstrate as a result of their learning?
2. Under what conditions or in what ways will pupils have opportunity to demonstrate those behaviors and thus let their achievements be known? What ways to get evidence?
3. By what criteria would one appraise pupil achievements? pre- and post-testing? norms? cooperatively determined standards?
4. What factors control the attainment—background of pupils, nature of instruction? How can we assess those factors so their influence on the learning being evaluated can be known?
5. What implications do findings in interpreted in light of influential factors (step 4) have for plans?

Evaluating the State of Evaluation

Reason exists for the growing concern about the state of evaluation in schools. Because evaluation takes place daily, at every stage in the teaching-learning process, in every level of the school, and by all persons directly or indirectly concerned, its principles must be understood and put into practice. If one or more steps of the process is weak or omitted, learnings are doubly influenced, for evaluation both affects and is affected by the process.

In a discussion for Educational Leadership's section, "Ideas Into Action," Taba and Sawin (5) summarize the concerns of the ASCD Commission on Evaluation. As a member of the Commission, the writer has taken liberty to elaborate at several points. Important curricular decisions are made with inadequate research, often with meager evidence or none at all to support decisions. Many new developments in curricular organization, some of them national in scope, are justified solely on the ground that introduction of these programs does not reduce the amount of information learned while reducing the number of teachers per student. This implication seems to be advanced, for example, with proposals for team teaching and the use of television. Seldom is it questioned whether important educational objectives other than the acquisition of information may be jeopardized. There is no assessment of possible negative side effects programs may be having on such important educational objectives as mental health, ability to think, creativity, or sheer individual expression.

Five major areas of deficiency in the designing of evaluation programs are identified as follows:

First, the objectives on which the evaluation is based are often too narrow—despite, in many cases, a preliminary broad statement of aims having been prepared. The range of objectives actually evaluated may be limited to recall and a few academic skills.

Second, not only are objectives limited, but the tools to be used in their assessment are frequently limited. A broad range of evidence-securing devices also available has been ignored. Many times,
reliance is placed on published tests whether appropriate to the learnings or standardized to be applicable to the pupils in question. Limitation on kinds of evaluation tools used affects both the outcomes of and also the diagnosis of learning difficulties. Both are important preliminaries to individualizing learning. Both are essential in the educative process, if one agrees to the assumption that learners differ (6).

Third, much attention has centered on evaluation as an end product and an end in itself—a score, a comparison, a status study, a reward-punishment device, or even a tool for segregation. Evaluation is frequently considered as a final judgment to be applied primarily at the close of a learning experience and dissociated from the full range of objectives sought by the curriculum. Thus, little attention is paid to the process of learning or thinking by which the end product was secured. Little attention is generally given to the diversity which identical scores may mask. Conversely, differences in scores, for instance, are often assumed to be real differences. A focus on the end product of evaluation encourages confusion between “effect” and “cause.” Such focus invites decisions which call for “more and more” or “less and less.” If the world of grades, scores, or external testing becomes a major end in and of itself, educators may find more profound values have given way and a vision lost of “education for what is real” (7).

Fourth, interpretation of results without adequate related data leads to misinterpretation. A lack of systematic information on the nature of the learner, his cultural backgrounds, his drive for schooling, the socioeconomic nature of his environment, for example, will hamper sound conclusions. Without accurate information about teacher behaviors or teaching-learning operations in the classroom, data about achievement may not yield sound implications for educational plans.

Despite the existence of an extensive literature and a maturing discipline of evaluation, major studies of educational innovation tend to test their outcomes with much less than a balanced program of evaluation. Because of the influential status and extensive financial underwriting of studies in question, their outcomes will draw attention from all levels of the public. Needless to add, where such studies will advocate expensive and extensive reordering of school arrangements, whether material, personnel, or both, misleading or incomplete evaluation is especially serious. Yet, without the inclusion of sound evaluation in their research designs, scholars subject their generalizations and implementation proposals to grave question.

Fifth, a major deficiency in modern evaluation programs lies in the failure to translate data into action, whether at district, school or classroom level. This is the hardest step of all: the rational use of data to govern decision making and to pattern changes in behavior. There is, moreover, no simple or direct line from rational proposals to expeditious change in educational procedures. Human relations skills become a requisite to effective action as much as do sound evaluation data.

In its closing report to the Executive Committee of ASCD, the Commission on Evaluation added several other points: Because standardized testing is being concentrated in large institutions (evidently a few test bureaus dominate the field), it would appear . . . these institutions have an inordinate chance to dictate what shall be taught and how.
They may thereby have undue influence over determining what local needs are acknowledged, which characteristics of a student body are recognized, or what kinds of success shall be rewarded.

An incipient trend toward state-wide mandatory testing adds further pressure on sound curricular decisions. These programs concentrate exclusively on achievement of factual knowledge. In a few instances so far, these programs have eliminated all testing except that of intelligence and achievement. This limits the possibility of interpreting achievement in terms either of cultural or personality factors, or of factors involving the conditions of learning. Legislation is being proposed in some states which makes it necessary to compare schools against each other and against national norms. If one contemplates the consequence of such distorted evaluation procedures on particular programs, such as compensatory education for culturally disadvantaged children, one can see the seriousness of faulty evaluation techniques (8).

Overcoming Obstacles

Surmounting obstacles to adequate evaluation through a focus on information and demonstration: To this end the Evaluation Commission of ASCD moved from criticism into the development of a proposed model. It was the Commission's intent to try out a comprehensive evaluation program in a specific setting to locate or devise truly workable procedures. A description of the proposed model is included in an Educational Leadership article, "Ideas Into Action—A Proposed Model for Evaluation" (5). This article contains a wealth of specifics, as does Chapter 19, "Evaluation of Outcomes of Curriculum" in the 1962 volume by Hilda Taba (3). The proposed model is also outlined there.

Overcoming obstacles through efforts to broaden understanding: Five teacher education needs have been identified by Taba, Brady and Robinson in consultation with schools having acculturation problems for which broad diagnostic information about their pupils was essential (9). Teachers needed: (a) to become acquainted with children's experiences and their opportunities to learn as a basis for deciding what the school should supply; (b) to understand the realities in children's lives so school experience could relate to these; (c) to learn the values and sanctities in the community and its groups so these could be respected; (d) to acquire new feelings about the community and new insights into its problems and ways of doing; and (e) to know what feelings people have about themselves.

These needs for evaluation information will scarcely be met with limited academic kinds of data or with those obtained primarily from paper-and-pencil or standardized test situations. Nor will depth in subject disciplines as now perceived supply the kinds of "empathy" and "group participation skills" the writers found necessary for teachers in schools serving pupils with widely heterogeneous backgrounds.

One may safely predict a continuing and increasing need for all school people to be thoroughly aware of sound evaluation principles and procedures. In 16 recommendations about staff competence in testing, Hagan and Lindberg (10) make a strong case for an increasing skill and understanding of testing at all levels of the school staff. Their recommendations are equally valid in the broader context, evaluation. They urged extensive in-service education.
As a means of disseminating experience and skill, the creation of a panel of experts in evaluation has been suggested. They would be available to consult with school systems about evaluation design, interpretation and implementation. In light of the impact of systematic review in the testing field, such as those in the Mental Measurements Yearbooks (11), feasibility of adapting the idea to a review of evaluation programs and techniques should be studied. There may be opportunity for school systems to adapt the principles in "self-study" and "accreditation" in order to stimulate close examination of evaluation programs and their relation to educational goals.

Seminars or institutes have been suggested to stimulate interest and increase competency. Opportunities should be available to learn about advanced thinking in the discipline of evaluation and particularly to learn about newer technical tools.

Most of the new devices have within them the potential for stereotyping students, curriculum and teaching. They also offer a golden promise of adding new dimensions to education and arming the teacher with powerful tools in individualizing his work. Where we place our emphasis in using these new aids will determine whether they are used for ill or for good (12).

Overcoming obstacles through rethinking some basics: Growing acknowledgment of social forces—urbanization, impact of technology on occupational structure, ghettoization—some key but not the only factors, will continue to press educators to see the inadequacies of old social and school arrangements which deny productive access to educational opportunity for large segments of the public. There is much rethinking to be done. Evaluation in all its steps will need sensitive restudy. Older answers may not suffice (13).

Much attention to knowledge "explosion" in subject areas has made disproportionate impact on teacher preparation requirements. Emphasis is on education of young teachers in subject disciplines to the virtual exclusion of other needed insights and skills. Such an emphasis ignores parallel knowledge explosions occurring in those disciplines forming the foundations of education. The assumption that subject matter preparation in depth will bring with it the skills and insight necessary for adequate evaluation of learnings in that discipline needs careful study. There is a great gap between "knowing" and being able to "induce that knowing in others" and further—to be skilled in a sensitive evaluation of learning needs and outcomes. These are gaps scholars have thus far chosen to ignore.

The education profession will need to deal with an important question: the division of responsibility between preservice and in-service education. At present, a new teacher's introduction to competency in evaluating is piecemeal. Education for sound evaluation will need to be planned systematically and seen as a continuous program ready to handle new training needs as they arise.

A second question implicit in professionalization is one of locating balance between specialization and general preparation. A danger lies in, "let George Expert do it," or its corollary, "George is the only one who knows how. . . ." These are questions of balanced use of all steps in the evaluation process.

What is of most worth? Will it make sense to completely evaluate the flea—
only to have ignored his host, the elephant? Evaluation and education, part and whole, and the people, but especially the people, who influence and are influenced by both part and whole, will be developing increasing skill and understanding as each and all continue to grow in their grasp of the basic meanings of education in a pluralistic and changing America.

Bibliography
