14. COURSE IMPROVEMENT THROUGH EVALUATION

LEE J. CRONBACH

The national interest in improving education has generated several highly important projects attempting to improve curricula, particularly at the secondary-school level. In conferences of directors of course content improvement programs sponsored by the National Science Foundation, questions about evaluation are frequently raised. Those who inquire about evaluation have various motives, ranging from sheer scientific curiosity about classroom events to a desire to assure a sponsor that money has been well spent. While the curriculum developers sincerely wish to use the skills of evaluation specialists, I am not certain that they have a clear picture of what evaluation can do and should try to do. And, on the other hand, I am becoming convinced that some techniques and habits of thought of the evaluation specialist are ill-suited to current curriculum studies. To serve these studies, what philosophy and methods of evaluation are required? And, particularly, how must we depart from the familiar doctrines and rituals of the testing game?

DECISIONS SERVED BY EVALUATION

To draw attention to its full range of functions, we may define evaluation broadly as the collection and use of information to make decisions about an educational program. This program may be a set of instructional materials distributed nationally, the instruc-
tional activities of a single school, or the educational experiences of a single pupil. Many types of decision are to be made, and many varieties of information are useful. It becomes immediately apparent that evaluation is a diversified activity and that no one set of principles will suffice for all situations. But measurement specialists have so concentrated upon one process—the preparation of pencil-and-paper achievement tests for assigning scores to individual pupils—that the principles pertinent to that process have somehow become enshrined as the principles of evaluation. “Tests,” we are told, “should fit the content of the curriculum.” Also, “only those evaluation procedures should be used that yield reliable scores.” These and other hallowed principles are not entirely appropriate to evaluation for course improvement. Before proceeding to support this contention, I wish to distinguish among purposes of evaluation and relate them to historical developments in testing and curriculum making.

We may separate three types of decisions for which evaluation is used:

1. Course improvement: deciding what instructional materials and methods are satisfactory and where change is needed.
2. Decisions about individuals: identifying the needs of the pupil for the sake of planning his instruction, judging pupil merit for purposes of selection and grouping, acquainting the pupil with his own progress and deficiencies.
3. Administrative regulation: judging how good the school system is, how good individual teachers are, etc.

Course improvement is set apart by its broad temporal and geographical reference; it involves the modification of recurrently used materials and methods. Developing a standard exercise to overcome a misunderstanding would be course improvement, but deciding whether a certain pupil should work through that exercise would be an individual decision. Administrative regulation likewise is local in effect, whereas an improvement in a course is likely to be pertinent wherever the course is offered.

It was for the sake of course improvement that systematic evaluation was first introduced. When that famous muckraker Joseph Rice gave the same spelling test in a number of American schools and so gave the first impetus to the educational testing movement, he was interested in evaluating a curriculum. Crusading against the extended spelling drills that then loomed large in the school schedule—“the spelling grind”—Rice collected evidence of their worthlessness so as to provoke curriculum revision. As the testing movement developed, however, it took on a different function.

The greatest expansion of systematic achievement testing occurred in the 1920s. At that time, the content of any course was taken pretty much as established and beyond criticism, save for small shifts of topical emphasis. At the administrator’s direction, standard tests covering this curriculum were given to assess the efficiency of the teacher or the school system. Such administrative testing fell into disfavor when used injudiciously and heavy-handedly in the 1920s and 1930s. Administrators and accrediting agencies fell back upon descriptive features of the school program in