EVALUATION AND THE IMPROVEMENT OF COMPENSATORY EDUCATIONAL PROGRAMS

Rodney W. Skager

CSE Report No. 54
September 1969

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California

Evaluation and the Improvement of Compensatory Educational Programs[1]

Those who comment on the manner in which instructional programs are evaluated usually begin by suggesting that vast improvements are needed in both conceptualization and practice. Consistent with this prevailing view, it certainly appears that the great majority of evaluations of compensatory education have not produced sufficiently useful and informative results. More than anything else this admittedly uncharitable assertion reflects a lack of awareness on the part of the majority of evaluation specialists and the decision-makers who are their clientele as to the extent to which improvements in educational practice can be facilitated by the optimal use of evaluation procedures. Not only is there a wider role for evaluation, but the application of appropriate methods of evaluation is an essential component in the development and operation of successful instructional programs, especially those intended for disadvantaged students.

---

[1]Revision of a position paper presented to the Seminar on Educating the Disadvantaged, University of Wisconsin, Madison, April 9-10, 1969.

While typical evaluation practice is deficient in many ways, achievement of more effective practice does not necessarily require the involvement of more educational researchers or the expenditure of large sums of money. Contrary to the beliefs of some individuals, we do not need teams of scholars constructing models of the relationship between research and decision-making in education. As far as the latter is concerned, in American education there has infrequently been a close relationship between educational research and educational decision-making, one of the reasons, perhaps, why evaluation research has been less effective than it might have been.

The sense of urgency which says that programs of compensatory education must be successful also implies that we will have to do better with the conceptual and technical tools which are either presently available or can be developed very soon. We admittedly have much to learn about the evaluation of instruction and have limited resources, especially in personnel with appropriate training and experience. Nevertheless, enough is now known in the methodology of evaluation to do a far better job than is now being done.

As it is used here, the term *evaluation* refers to the collection and interpretation of systematic information about the effectiveness of alternative educational practices. The word "systematic" is important. One common usage of that term refers to any

process that is "methodical in procedure or plan." By such a definition educational research as it is currently conceived is not the only source of systematic information about educational phenomena. Research should be an important source, but many knowledgeable people in the field have felt for some time that the prevailing conception of evaluation incorporates only the most rigorous (and restrictive) facet of the total spectrum of research activity--that of testing finished educational products by means of some approximation of the classic laboratory experiment (Guba & Stufflebeam, 1968). In order to understand the implications of the previous statement, a broader view of the functions of evaluation in educational decision-making is required. Such a view will also provide a context for suggestions as to how evaluation procedures may be applied more effectively in compensatory education.

Very recently, Provus (1969) reported a clear and comprehensive conceptualization of the various functions of evaluation in education. While based on experience gained in evaluating ESEA--Title I programs in the city of Pittsburgh, Provus' conception is consistent with views presented previously by others, such as Stufflebeam (1968), and reflects the germinal ideas advanced by Cronbach (1963). These ideas seem to represent something approaching a consensus among a number of individuals attempting to advance the theory and practice of evaluation.

Provus divided the various evaluation functions into stages, beginning with the design of a new instructional program and ending with a final decision about incorporating the finished product into regular operational use. If we add one more stage to those described by Provus (Analysis of Need, below), evaluation becomes a continuous cycle reflecting the informational requirements of a matching cycle of educational decisions.

Analysis of Need

The first information provided by evaluators which is relevant to decisions about change in educational practices involves determining which skills, attitudes, and values are deficient in the target group of learners. The suggestion that one begin by assessing the needs of students is hardly profound. Yet personal experience and many conversations with evaluation personnel have revealed that changes are often instituted in schools not because particular needs have been isolated but because in some communities the use of up-to-date methods is fashionable, because money is available, because the results of the changes would facilitate convenience for the staff, or because of other reasons unrelated to priorities based on the needs of students.

Some may argue that such analysis of need is mainly appropriate for affluent middle class schools since, when learners are drawn from severely deprived contexts, it can be assumed that

instruction relevant to *any* learning objective is appropriate. This argument is not valid. In the section of the most recent California annual report on ESEA--Title I (*Evaluation of ESEA--Title I Projects*, 1968), projects in large cities were criticized because diagnostic testing had rarely been used to identify specific weaknesses of children participating in compensatory programs. Even where learners are educationally deprived, the failure to apply appropriate evaluation techniques to establish the learning needs of students can often result in waste of instructional efforts on skills that are already developed in many students. One side effect of such misdirected effort, particularly in older minority children, is a reinforced sense of being patronized by a system that has abandoned instruction for care-taking.

For example, the author was involved in the evaluation of a secondary school compensatory mathematics program in which evidence from a variety of sources revealed that in the absence of diagnostic information on student achievement, even a group of highly able and dedicated teachers concentrated instruction on areas in which students initially were relatively strong rather than on skills achieved by few students at entry (Skager, 1969). The reasons for this behavior on the part of teachers are complex, but the message is clear: if we are to develop or select materials and procedures that will meet learners' needs, we first must determine what those needs are.

Assessment of Program Design

Provus' first stage of evaluation concerns itself with whether or not the design for a new program is satisfactory in light of such factors as completeness and potential for achieving the anticipated results. This stage naturally follows the analysis of need described above. Here the role of the evaluator does not involve the kinds of activities typically associated with educational research. Rather, the task is best handled in a way comparable to systems analysis as used in engineering. In evaluating the design of a program of instruction, expert consultants, administrators, teachers, and other officials need to be brought together to make judgments based on professional experience and prior research in other contexts. Parents or other interested members of the local community probably would have a great deal to contribute as well. The evaluation of a design or plan does not involve the collection of data; it amounts to the structuring of a dialogue. As such, different skills are required of the evaluator than those usually associated with research competence.

Congruence Evaluation

The next stage of evaluation recognizes the fact that plans generated on the educational drawing board are not necessarily used in the school. In education as elsewhere it is easy to paste new labels on outmoded practices. It is also true that all elements of

even a careful plan may not be practicable in the real instructional situation. Likewise, the instructional staff, required to utilize new materials and practices, will often tend to return to more familiar methods of operation. When an instructional design is first applied, the evaluator's job is to monitor the degree to which there is congruence between the design and the practice supposed to result from that design. Where such congruence does not exist, either design or practice must be modified.

The importance of evaluation for congruence cannot be overemphasized in the case of compensatory education programs. It is inevitable that many of the educational personnel involved in designing such programs and rendering them operational are faced with the real challenge of responding to the needs of children from an unfamiliar cultural context. Mistakes will inevitably be made, even given information about entry skills. Here, too, the special knowledge of the local community can be essential.

The report prepared by the American Institute of Research for the Title I--ESEA Fourth Annual Report (1969) pointed out that "...instruction irrelevant to the stated objectives of the programs..." was the most frequent reason for failure of programs at the elementary level, a result that could have been avoided by the use of congruence evaluation procedures. It was evident from the report that similar problems appeared in unsuccessful programs at the secondary level. Fox's (1968) evaluation report on the More

Effective Schools Program in New York City concluded with observation that only part of the program as originally proposed had been evaluated since only part had ever been made operational. Unless design and practice are brought into congruence, successful programs cannot be used elsewhere. A program design inconsistent with successful practice will only mislead attempts at wider application.

Program Improvement Phase

Congruence evaluation is concerned with measuring how much has been learned by students participating in the program. Once congruence between design and practice has been achieved, however, it is time to look at the *elements* of the instructional program to see which are effective with students. Here the evaluation researcher collects the type of data educational researchers usually collect, including scores on tests or observation of the learning process.

It is of critical importance to remember that the information produced in this phase of evaluation does not ordinarily make it possible to judge the total program as a final educational product but rather is directed at improving that program by providing as much information as possible about the relative success of its parts. While this evaluation function often involves the collection, analysis, and interpretation of data, it also requires a new

orientation on the part of the research personnel trained in experimental behavioral sciences. The evaluator engaged in program improvement intervenes directly in the educational process whenever the results of his research are used to improve the instructional practices being evaluated. The word "intervenes" is the key to the whole matter. When there are many changes resulting from such intervention during the first year or two of operation, it is not ordinarily possible to make summative statements about the effectiveness of the final program. The waters have been muddied by trial and error, and student achievement or other criterion data cannot be uniquely attributed to the finalized instructional procedures.

Unfortunately, this kind of experimentation to improve instruction does not produce very much information for the typical annual evaluation report, at least not as such reports are presently conceived, either in education or government. This situation is tragic because if there is one evaluation function that is supremely important for the success of educational programs, it is this deliberate use of evaluation procedures to improve practice.

Program Certification

Judgments about the instructional program as a whole are appropriate only when the developmental stages of program design and improvement have been successfully completed. As in the need analysis

phase, the evaluation researcher can here behave in the way that educational researchers are accustomed to behaving. He may be able to use pre- and posttest designs and sophisticated data analysis methods, which hopefully will allow him to make reasonably unequivocal statements on the extent to which the new program brings about improvements in those student characteristics selected as criteria for determining effectiveness. In contrast to the program improvement stage of evaluation, the evaluator must avoid intervention in the educational process. He does not want the instructional practices under study to be affected in any significant way by the fact that evaluation data are being collected. If the practices are thus affected, the results of the evaluation apply not to the program under normal conditions of operation but merely to the program at the time it is evaluated for certification. Those who have conducted evaluation research in schools know that complete non-intervention is a goal that can be striven for but never entirely achieved. Principals will urge teachers to expend maximum effort because the "school is being evaluated." Materials from experimental programs will appear mysteriously in "traditional" classes supposedly used for purposes of comparison. Brighter students will often be assigned to whatever program authorities hope will look best. Unkind as these observations may be, they do reflect one reality of research on operating educational programs. Fortunately, the business

of controlling such extraneous factors is the special domain of the trained researcher. In evaluation for program certification he is very much on his home ground. It is unfortunate that most educational researchers inexperienced in evaluation ordinarily will undertake *only* this certification function, disregarding the previous stages.

Cost Benefit Analysis

The final stage of evaluation, one which remains more of an ideal than a practical reality, confronts decision-makers with comparisons of alternative educational programs in the form of summary information portraying what the program has achieved against the reality of its cost. In practice there is usually insufficient information available on the various program alternatives to permit formalized procedures of cost-benefit analysis, but approximations are made based on available hard data plus judgment.

Summary

Six stages of evaluation have been summarized. Beginning with evaluation to determine the needs of the intended population of learners, we proceeded through evaluation of the design of the program, evaluation of the extent to which design and practice coincided, evaluation to improve practice, evaluation to judge

the program as a whole, and a final evaluation activity that involves the integration of program costs with conclusions as to program effectiveness. Not all programs of compensatory education must go through all of the stages, of course. But many should, particularly those programs that begin anew with the development or adaptation of practices not used previously in a given situation.

Taken as a whole, these evaluation practices involve active participation aimed at building effective educational programs, rather than simply the educational counterpart of a good housekeeping seal of approval. If this broader role for evaluation is as promising as it appears to be, then contemporary evaluation practice, as reflected in the content of federal and state summaries of evaluations of Title I programs, usually falls far short of its potential. Practice today is overwhelmingly limited to the program certification facet of the evaluation spectrum. At certain times such information is admittedly important to decision-makers. Still, if appropriate evaluation techniques are not utilized in program development, far too many of the certification decisions will be negative, reflecting waste of funds and skilled educational personnel, as well as confronting the students and parents concerned with yet another disappointment.

IMPROVING EVALUATION PRACTICE

I. The widespread tendency to overemphasize program certification at the expense of the other types of evaluation functions suggests that responsible officials, school administrators, developers, and researchers are either mainly unaware that these other functions are important or do not know how they are to be implemented. Certainly we are a long way from resolving all the problems in our conception of evaluation, but it does seem that practice would be vastly improved in the near future if, as a first step, *a set of guidelines were drawn up, elaborating the kinds of evaluation activities described above, and given suitable credibility via the endorsement of major fund granting agencies, especially the U. S. Office of Education.*

One device for establishing the suggested guidelines would be to convene, under the auspices of an appropriate agency, a suitable panel of persons knowledgeable in evaluation. While there will inevitably be some disagreements among the members of such a group as to the specifics of emphasis and terminology, there appears to be enough of a consensus at present to allow for a satisfactory resolution of conflicts, certainly one specific enough to greatly improve evaluation practice. Precedent for the development of such guidelines exists in the elaborate standards for the construction and evaluation of psychological tests published

by the American Psychological Association. Admittedly, there is the danger of a premature codification, which inhibits later developments in the conceptualization of evaluation. But in view of the social urgency underlying the development of effective programs for deprived learners, the need for improved evaluation practice is great enough to be worth the risk.

II. The preparation and dissemination of a set of guidelines covering the full evaluation cycle will have only a minor influence on program effectiveness unless the *specification of an evaluation design is held by fund granting agencies to be just as important as the description of the proposed instructional program.* The implementation of this suggestion would be far easier given the set of guidelines called for above. The majority of educators engaged in the planning and operation of compensatory education programs cannot be expected to know what is meant by an evaluation design without appropriate guidelines.

The requirement that evaluation designs be built into proposals would at last give evaluation personnel a meaningful role in program planning. A director of evaluation for the board of education of a large state expressed with irritation that "evaluators are never present at the beginning." His observation reflects the very real frustrations of experienced professionals at repeatedly being

called in when it is already too late to play a constructive role in the improvement of plans and programs. Without this opportunity experienced evaluators know that the later certification phase is far less likely to be brightened by the discovery that significant gains in achievement are associated with the program.

III. One barrier to the effective implementation of the previous proposals is the scarcity of available personnel having requisite skills and experience in evaluation. The requirement written into the 1965 educational bill that every project be evaluated at least annually is as laudable as it is unrealistic, if one conceives of evaluation as more than the administration of a standardized test of achievement. The last national annual report estimated that there were over 20,000 Title I projects funded at the time the report was compiled. The number of qualified evaluation personnel available is difficult to estimate, but it certainly does not in any way approach the number needed for Title I alone, particularly if all such projects utilized effective evaluation procedures. The final recommendation to be submitted in this paper will offer one strategy for the more effective use of available personnel. But in order to increase the number of such individuals, there is clearly a need for *the preparation of realistic and effective training materials for evaluators, developers, and administrators.*

In order to have an impact in the reasonably near future, such training materials must be relatively brief, exportable to places where conventions and workshops are held, and designed to confront the trainee with a sense that the learning experience offered has a direct relationship to the demands of his own work. Simulation offers an effective avenue for meeting these requirements. At the Center for the Study of Evaluation at UCLA, for example, we have constructed one such exercise which in two days takes trainees through the highlights of about one year of an evaluation project. The CSE Simulated Evaluation Exercise utilizes documents adapted from actual evaluation projects and incorporates slides and taped materials in an effort to provide a sense of reality. On the basis of initial information about a Title I curriculum development project, trainees plan evaluation research and then receive feedback in several stages as to the adequacy of their plans. Changes or constraints on the project occurring after initial plans were made are reported to the trainees so that evaluation designs can be modified.

It is important that such training devices be designed for developers and decision-makers as well as for evaluators. The expectations of the former as to what can be accomplished by effective evaluation strategies are just as important as the training of the evaluator. Until perspectives are broadened on both

sides, evaluation personnel will continue to be called in too late to be of service in program improvement. One way to broaden perspectives through the proposed training procedures would be to develop a simulation exercise in which evaluation phases would be worked through by teams of trainees composed of administrators, program developers, and evaluators, each playing their own professional role.

IV. Procedures for training those who conduct evaluations or use the results therefrom can be developed in a reasonably short period of time given suitable application of energy and experience. Raw material for simulated training exercises exists in abundance in the records of many compensatory education projects now in operation. The evaluation guidelines proposed earlier would provide another essential element for the development of such training devices. Still, even given the widespread use of effective training materials, evaluation personnel face the additional major problem of finding or developing the tools to measure program outcomes. Tests or other criterion measures consistent with program objectives are often as not unavailable.

The section on measuring devices of the New York State report on ESEA--Title I (Closing the Gap, 1968) concluded with the statement "...*regardless of whether or not the test was appropriate it*

is clear that most of the programs were measured by a standardized test." (Italics mine) The extreme overdependence on standardized testing characteristic of contemporary practice has been deplored many times in the past. The point is not that standardized tests are useless but rather that they are used to answer questions that they were not designed to answer. Such misapplication occurs in several ways. First, in searching for gains in students' achievement, standardized tests are often used without paying any attention to whether or not the content of the test is particularly consistent with the instruction being offered. Standardized tests often range rather broadly in content and reflect long-term educational experience. It is unrealistic to expect such tests to reflect short-term learning goals or even long-term goals where there is only a partial overlap between such goals and test content. Moreover, standardized tests are designed for middle class members of the majority society. Performance of members from other subcultures can show decrements based on unfamiliarity with test format and with the language used in instructions. Finally, Hunter and Rogers (1967) and others have warned that the norms by which standardized tests are interpreted are often grossly inappropriate for the rural poor and for urban residents in general.

Many bemoan this situation, particularly the problem of cultural bias in tests, but few suggest constructive solutions. The

problem will not be solved by throwing out standardized tests but by using them wisely with other kinds of measures. For example, all aptitude tests measure skills learned in a given subculture. If we wish to measure the learning abilities of members of minority subcultures, then we must formulate test content in terms of skills their children have had the opportunity to develop. An associate very active in the development of curriculum materials for deprived urban minority children recently suggested that the first step in making an inference about aptitude for learning in an inner-city minority child is to ask him what he does best. If it is playing pool, then the thing to do is to find out how well he has learned to play pool compared to other pool players his own age. The suggestion was not a facetious one, although we might hope to find somewhat more generalized measures. On the other hand, carefully constructed standardized achievement tests do reflect standards set by the culture as a whole.

Imperfections aside, such tests represent the realities of educational expectations in our society. It is because the educational opportunities afforded many children result in lower performance on such measures that we have programs like ESEA--Title I, and such programs will be necessary as long as the decrements exist. Without some kind of common standard, whether it be standardized tests or an as yet undefined approach that is superior, one

of the most convincing justifications of the need for compensatory programs would be removed.

Standardized tests are frequently used for purposes for which they were not designed and are interpreted in ways that were never intended. Naivete on the part of some individuals doing evaluations is only one of the reasons. More important is that alternative measuring devices are usually unavailable. Evaluation personnel often do not have the resources to develop measures appropriate for local objectives.

No one knows how to solve all of the problems encountered in the measurement of achievement. Moreover, our devices for measuring non-cognitive characteristics, such as the "positive self-image" cited in so many compensatory education proposals, are even less adequate. Some action can be taken to facilitate relatively easy and inexpensive assemblage of achievement tests with content closely related to the learning objectives of local programs. This can be achieved through the development of *central banks of instructional objectives with accompanying pools of tasks or items measuring each objective*. These objectives and items should cover pre-school through secondary levels in the most important target areas for compensatory programs, especially reading, language, and number skills. Of additional use would be an efficient item retrieval system to facilitate the prompt and inexpensive construction of tests for local use.

How might such a measurement system work? Briefly, local evaluation personnel, having interacted with program designers in the specification of instructional goals, would review learning objectives written for the content area in question, select those objectives compatible with local instructional goals, and order a test or tests measuring the desired skill. Unlike the standardized test, questions on such locally prescribed instruments would reflect local learning goals and provide relevant information for the evaluation stages of need assessment, program development, and program certification. The Center for the Study of Evaluation is working on one prototypic system of this nature (the Instructional Objectives Measurement System) in the area of mathematics. One factor facilitating the development of such systems is that many agencies and programs are independently developing pre-school and primary level instructional objectives, some with accompanying sets of test items. These materials can be gathered and incorporated into comprehensive sets of learning objectives, as the Instructional Objectives Measurement Exchange, also at CSE, is attempting.

For the measurement of achievement such systems are technically feasible now. Systems incorporating other types of objectives, particularly non-cognitive objectives, might be included later; but the need for this kind of resource in the cognitive area

is most pressing. Instruction in the use of such objective-item systems could be incorporated in the evaluation training programs suggested earlier.

V. Given progress on the problems discussed above, there remains a serious deficiency of personnel with pre-requisite skills in research methods, measurement, and other relevant disciplines. Given such a shortage of human resources one can either let the market somehow distribute haphazardly what resources are available among competing alternatives, or one can begin to establish priorities on the basis of where the resources are likely to do the most good.

The establishment of priorities would obviously be a more productive approach. One key to how this might be accomplished has already been provided by Congress in the 1968 amendment to ESEA which called for early identification of those programs with the highest promise of improving the achievement of participating children. This request was reflected in the most recent Title I annual report in the comparisons made between successful and unsuccessful programs and in the list of generalizations derived from those comparisons. Congress apparently feels that one important, if not the most important outcome of ESEA--Title I is the development and dissemination of better educational practices. If

authorities agree, as suggested earlier, that evaluation makes the greatest contributions where it can be applied in the assessment, design, and program improvement phases, then *priorities for funding proposals for the development of new educational programs should emphasize large-scale programs with well-conceived evaluation strategies. During program design and development phases, refunding decisions should be based on evaluation strategies relevant to those phases rather than on premature attempts at program certification.*

Title I projects vary considerably in the extent to which developmental activities are undertaken. The most original programs, particularly those ambitious in scope, require the full spectrum of evaluation activities described earlier. Other kinds of Title I projects, laudable though they may be, often amount simply to the provision of extra reading specialists or other resources of a traditional type. These kinds of programs do not require total evaluation effort because they do not incorporate innovations. In most cases, program certification activities are sufficient after an initial assessment of need. The greatest expenditure of evaluation resources, then, should go into large-scale developmental programs.

Unless funding agencies permit evaluation activities to be pertinent to program needs, evaluation practice will not change.

As a result, most evaluations will continue to contribute little or nothing to the quality of the educational practices developed under Title I. In fact, the effect of evaluation will sometimes be to lower the quality of instructional programs by placing constraints on program development. Even the annual program certification evaluation report now typically required from evaluators is of little use to those making decisions as to refunding since deadlines established at higher levels usually require that such decisions be made before the reports are available.

Conclusion

The ideas presented in this paper are neither completely original nor especially controversial. All can be implemented; indeed, some already are being at least partially implemented. Surely it is time to utilize evaluation procedures that do more than establish grounds for final judgments about educational programs. Evaluation will be most productive when it is seen as part of a process helping to render those final judgments favorable.

REFERENCES

Closing the gap. (Report on the first two years of ESEA--Title I in state of New York). Buffalo: University of the State of New York, State Education Department, 1968.

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

Evaluation of ESEA--Title I projects of California schools. Annual Report, 1967-1968. Sacramento: California State Department of Education, 1968.

Fox, D. J. Conclusions of the center's MES evaluation. *Urban Review*, 1968, *2*(6), 17.

Guba, E. C., & Stufflebeam, D. L. Evaluation: The process of stimulating, aiding, and abetting insightful action. (Address to the Second National Symposium for Professors of Educational Research, Boulder, Colo., 1968).

Hunter, L. B., & Rogers, F. A. Testing: Politics and pretense. *Urban Review*, 1967, *2*(3), 5, 6, 8, 25-26.

Provus, M. M. *The discrepancy evaluation model: An approach to local program improvement and development.* Pittsburgh: Pittsburgh Public Schools, 1969.

Skager, R. W. *Student entry skills and the evaluation of instructional programs: A case study.* CSE Report No. 53. Los Angeles: Center for the Study of Evaluation, 1969.

Stufflebeam, D. L. Evaluation as enlightment for decision-making. (Address to the Working Conference on Assessment Theory, ASCD, Sarasota, Fla., 1968.

Title I--ESEA: A review and a forward look--1969. Fourth Annual Report. Washington, D. C.: National Council on the Education of Disadvantaged Children, 1969.