Evaluating Programs for the Gifted:
Four Questions About the Larger Issues

Joseph S. Renzulli
Associate Director, Bureau of Educational Research
School of Education, University of Connecticut

Abstract

This paper discusses four major issues concerning the evaluation of programmes for the gifted.

1. The need to evaluate gifted programmes so that planning, development and accountability form a natural sequence of educational objectives.
2. Decisions for change and improvement can only be made in the light of sound evaluation and this evaluation must consider the practical as well as the educational issues.
3. Evaluation should be diagnostic and specific to a particular programme.
4. It must be recognized that the evaluation of gifted programmes requires awareness of the problems associated with assessing "higher level" objectives; the unsuitability of conventionally standardized tests and the practical demand on time, money and trained personnel.

There are probably dozens of questions that could be raised about both the theoretical and practical aspects of evaluating programs for the gifted. The remarks in this article, however, will be restricted to what I consider to be four of the larger issues associated with evaluation and the kinds of "big" questions that should be discussed and debated whenever program personnel decide to embark upon an evaluation study. Reaching a general consensus about the larger issues will help to place total evaluation efforts into a framework or context that hopefully will give some direction to the specific instruments and procedures that represent the "nuts and bolts" of evaluation studies.

Why Do We Need to Evaluate Our Program?

Although special provisions for the gifted and talented are an essential part of any school program that truly respects the principle of individual differences, the competition for limited resources among all types of supplementary programs frequently causes the needs of the gifted to be relatively low on the list of educational priorities. When school budgets are "cut" it is not unusual for the gifted program to be one of the first items to be eliminated. And when boards of education, legislatures, and other sponsoring agencies review the many special programs for which they are called upon

to support, the very survival of programs for the gifted may depend on having evaluation information readily available. All too often, evaluations have been launched as last-ditch efforts to save programs that are in danger of being eliminated or sharply reduced in the amount of support they receive from sponsoring agencies. Although a hastily conducted evaluation may be better than no evaluation at all, the best weapon in the battle for program support and survival is a carefully planned and comprehensive evaluation that will accurately document all aspects of the services being provided for gifted and talented youngsters. Evaluation should be an essential and ongoing part of total programming and each step of the planning and development phases of a program for the gifted should give careful attention to the ways in which evaluative information can be gathered, organized and presented to decision making individuals or groups.

The need for program evaluation in gifted education has grown out of a general concern on the part of decision makers for greater accountability in all aspects of education. In the past, innovation in education and especially efforts to help youngsters with unusual needs such as the gifted and talented were looked upon with a strangely philanthropic attitude. We accepted the notion that innovative efforts equaled innovation itself—that sincere and honest attempts to improve the education of gifted students were *de facto* indicators of favorable results. In other words, the attitude of "trying equals success" often caused us to minimize the need for program evaluation and, indeed, this attitude sometimes served as a substitute for evaluation. The person who was bold enough to raise serious questions about the value or equality of a particular program was frequently looked upon as some sort of malcontent, especially if the program in question was cloaked in the mantle of innovation, launched with great fanfare, and happened to be the "brain child" of an influential group or well-known "expert" in the educational establishment.

Programs for the gifted have been especially vulnerable to substituting the "trying equals success" attitude for rigorous attempts to evaluate program effectiveness. Innovation approaches to learning are frequently used with gifted youngsters and many programs for highly able students are characterized by "flashy" activities, materials, and student products. Gifted students, *because* they are gifted, have the capacity for high level performance and their products are often of superior quality. But the nagging questions the evaluator must always examine are: "What types of programmatic learning experiences fostered this outstanding performance? Was the performance really attributable to the program or did it emerge simply because the child was gifted and we had the good sense to stay out of his way?"

In recent years the unprecedented expansion of programs for the gifted has provided educators with a wide variety of alternatives that can be used for serving students with special needs. But the availability of these alternatives has also given rise to many questions that can only be answered through systematic program evaluation. These questions generally deal with the appropriateness and effectiveness of various learning materials, teaching strategies, and approaches to program organization and management. Fortunately, the technology in educational measurement and evaluation has also undergone a period of increased growth and development. The greater

sophistication and availability of evaluation methods has provided us with some of the tools necessary for documenting the value and effectiveness of programs for the gifted. But if we are going to put this technology to work it becomes contingent upon us to seek out and analyze the instruments that are currently available. I recently visited a program that is almost totally directed toward the development of thinking processes. The program coordinator had no knowledge whatsoever about the availability of excellent process assessment instruments such as the SEA Test (Callahan, Covert, Aylesworth, Vance, 1981) and the Class Activities Questionnaire (Steele, 1969). These instruments were developed specifically for use in programs for the gifted, and yet we have made only minimal progress in putting them to work in our evaluation studies.

What Is the Relationship Between Evaluation and Decision Making?

The general purpose of evaluation is to gather, analyze, and disseminate information that can be used to make decisions about educational programs. Evaluation should always be directed toward *action* that hopefully will result in the improvement of services to students through the continuation, modification, or elimination of conditions which effect learning.

It should be emphasized that the conditions which effect learning are not necessarily restricted to cognitive growth measures. For example, if one of the objectives of a program is "to provide students with a wide variety of exposure to ideas, topics, and areas of study not ordinarily covered in the regular curriculum" (Renzulli, 1977), we can gather one aspect of evaluation data by simply preparing a matrix and descriptive summary of our year-long effort to implement this objective. Further data might result from interest questionnaires and feedback forms about the impact of new exposure experiences upon students and we can further examine the extent and nature of any student follow-up that might grow out of specially planned experiences.

Economy and efficiency can be improved in an evaluation design if we begin by raising three interrelated questions:

1. Who are the decision makers at various levels of possible action?
2. Over what actions do decision makers have control?
3. What information is necessary for making decisions?

Perhaps the best way to illustrate the interrelatedness and the importance of these questions is by developing a hypothetical example. Let us suppose that we are evaluating an anthropology program for gifted students in a junior high school. The program involves several trips to various city dumps where students attempt to study differences between communities by analyzing the types of objects people discard. Supplementary expenditures are necessary for transportation, insurance, reference books, and a consultant in anthropology. Who are the decision makers and over what actions do they have control? The board of education must approve the supplementary funds and therefore continuation of the program (action) rests with them. Let us speculate that one question with which the board is concerned is whether or not parents

are satisfied with the program. One segment of the information necessary for decision making thus becomes parental attitudes. This information might be gathered through the use of a questionnaire and/or interviews with a random sample of parents.

Another decision maker in this situation might be the school principal. It may be up to him or her to decide when the trips take place and whether or not any problems are resulting because students must be excused from some of their regular classes. The principal may require information about student performance and conscientiousness in completing regular classroom work that they may have missed as a result of participating in the special program. This information will help determine if the program should take place during school hours or at some other time. It may also reveal some unexpected findings such as hostility toward the program on the part of certain teachers. In this case the evaluator might want to recommend that greater efforts be made to familiarize the general staff with the nature of the program, why it is being provided for gifted students, and how the program fits into the general philosophy and objectives of the total school program.

Finally, the teacher and consultant in anthropology are decision makers in this situation because they have control over actions which relate directly to possible modifications in the instructional process. They may require information about the students' knowledge in anthropology and their ability to analyze data in social science. The teacher may also want to obtain information about student satisfaction with the program and his or her own success in working with gifted youngsters. Thus, tests, student questionnaires, and rating scales will be needed to help the teacher and consultant make decisions about how they might like to modify the instructional process.

By way of summary, decision making is a fundamental goal of evaluation and therefore it is important to identify decision makers and the actions over which they control at the beginning of any evaluation endeavor. Since the structure and focus of an evaluation should be guided by the decision making process, it is recommended that this process be analyzed during the planning phase of an evaluation.

What Are the Objectives of Program Evaluation?

Within the general decision making purpose of program evaluation there are a number of more specific objectives which help to give direction to the actual design of an evaluation. An evaluation is scarcely worth the paper it is written on if it does not provide relatively specific information that supports the maintenance, modification, or termination of *particular* program components. Thus, an evaluation should be "diagnostic" in the sense that it pin-points by careful examination the circumstances and conditions that result in identifiable changes in performance, attitude, or other indicators of program effectiveness. In order for an evaluation to play a constructive and positive role in the overall process of education, it should attempt to fulfil as many of the following objectives as possible.

1. To discover whether and how effectively the objectives of a program are being fulfilled.
2. To discover unplanned and unexpected consequences that are resulting from particular program practices.
3. To determine the underlying policies and related activities that contribute to success or failure in particular areas.
4. To provide continuous in-process feedback at intermediate stages throughout the course of a program.
5. To suggest realistic, as well as ideal, alternative courses of action for program modification.

Although most contemporary evaluation theorists would agree that measuring the attainment of objectives is essential, they would also support the position that an evaluation study should investigate any and all conditions that may influence the effective operation of a program. If an evaluator is told "where to look" and "what to look for" he may very well overlook important factors contributing to the success or failure of particular aspects of a program. For example, in an evaluation of a special program for gifted high school students it was found that the program was influential in helping a large number of students to clarify their career choices. A small number of students also reported that their involvement in the program had caused them to give up the use of marijuana! While neither of these outcomes were included in the stated objectives, they nevertheless provided powerful support for the continuation of the program.

The third objective of program evaluation listed above—to determine underlying policies and related activities that affect a program—calls attention to the fact that a successful program is frequently the result of policy decisions and actions that may influence instruction, but are not a direct result of instruction itself. For example, the procedures by which teachers are selected for a program may be based on a policy underlying teacher selection. This policy may be formal and written or it may simply exist in the minds of persons who are responsible for selecting teachers. But nevertheless, the policy (or lack of a policy) could have a serious impact on the program. Take, for example, a program in which teachers are selected on the basis of longevity or seniority rather than superior teaching ability. This policy may result in the haphazard selection of teachers, and if ineffective teachers are chosen the program may be doomed to failure long before students enter the classroom.

The fourth objective of program evaluation—to provide continuous feedback throughout the course of a program—calls attention to one of the basic distinctions between evaluation and research. This distinction is concerned with the responsibility for suggesting changes in program activities while the program is in progress. Generally, research is directed toward judging the effectiveness of a predetermined, carefully controlled, and relatively specific "treatment." It is concerned with the generalization and replication of a prescribed activity, and therefore a good researcher does *not* interfere or make suggestions while the prescribed activity is taking place. Evaluation on the other hand, is concerned with program improvement and providing continuous feedback so changes and modifications can be made as a program

progresses. Just as a ship's captain can miss his final destination by a wide margin if she or he fails to make small navigational corrections when his boat is off course by a few degrees, so also can a program "miss" its goal if the director does not recognize problems when they occur and take whatever action may be necessary to correct them. Continuous monitoring will help to nip many problems in the bud and avoid the kind of dissatisfaction that builds up when even the smallest problem is left unattended.

The final objective listed above—to suggest realistic as well as alternative courses of action—is concerned with the usefulness of evaluative findings. Realistic suggestions are those that take into consideration such predetermining conditions as the availability of funds and human resources, the prevailing political climate, attitudes toward certain types of programming on the part of key decision makers, and how a program for the gifted can work harmoniously within the overall framework of a particular school or school system. There would be little value in making a recommendation that requires a $100,000 outlay if the evaluator knows full-well that this money is not available. Similarly, an evaluator is being unrealistic if he recommends that an entire school system change its scheduling procedures because it is not meeting the needs of the gifted. Such recommendations are evaluation "cop-outs" unless the evaluator also suggests some alternatives that are within existing resources or within the realm of *possible* action on the part of the decision makers.

What Are the Special Problems in Evaluating Programs for the Gifted and Talented?

A. The Problem of "Higher Level" Objectives
Programs for the gifted are often characterized by a commitment to the development of higher powers of mind and advanced levels of awareness, interest, and other affective behaviors. This presents a somewhat unique evaluation problem because these objectives cannot be measured as easily and precisely as those objectives which deal mainly with the acquisition of basic skills. As we move up the scale of learning behaviors, from the simple acquisition of knowledge to the development of higher mental process, it becomes increasingly difficult to find measuring instruments that meet the scientific and practical requirements necessary for good evaluation studies. While virtually hundreds of relatively valid and reliable tests are available to measure skills in the traditional areas of school achievement, instruments of evaluating higher level objectives are not so readily available. In areas where these instruments have been developed they are often expensive to administer and/or score, and therefore their use in an evaluative study may be economically unfeasible.

A second dimension of this problem is that gifted programs are frequently characterized by highly individualized objectives for each student. Whereas a reading skills program for average or slow learners may have enough uniformity in its objectives to warrant large scale standardized testing, a program for gifted students may have many *different* objectives for each student. The reliability of most standardized tests is a function of group size and it is extremely difficult to show statistically significant pretest to post-test gains when only a few students are being evaluated with a given instrument.

Standardized tests can, of course, be effectively used in evaluating programs for the gifted if they (a) are valid (appropriate) measures of particular objectives, and (b) if they are used in situations where reasonable levels of reliability can be obtained. But when a teacher devises individualized objectives for each child, as is often the case in programs for the gifted, then we must seriously question the appropriateness of tests based on systemwide or nationwide objectives.

In recent years there has been a great deal of concern in education about the specification of objectives in terms of observable and measurable student behaviors. Many evaluators have looked upon the "behavioral objectives model" as a panacea for conducting evaluation studies. However, the nature of gifted programs and their concern for developing higher thought processes may make this model too cumbersome to be practically applied to programs for the gifted and talented. In fact, when the behavioral objectives approach is used in its most rigid form, it may even force program developers to focus their attention on the trivial rather than important behaviors of superior learners.

The rigid behavioral objectives model is inappropriate for programs for the gifted because it forces us to be primarily concerned with those behaviors that are easily measured. Such a situation may result in the tail wagging the dog. Michael Scriven, the single-most influential person writing on educational evaluation today, has pointed out that "putting pressure on (a person) to formulate his goals, to keep to them, and to express them in testable terms may enormously alter his product in ways that are certainly not always desirable" (Scriven, 1967, p. 55). In gifted programs we talk about types of learning that may lead to analyzing a moral principle, or synthesizing a political argument, or evaluating a philosophic point of view. We talk about producing unique plans, communication forms, and products. And we also talk about some relatively sophisticated non-cognitive objectives such as developing favorable attitudes toward learning, or developing acceptance of and even appreciation for opposing points of view, or showing a commitment to a cause by taking affirmative actions. Although many experts in the testing business believe that these complex objectives can be evaluated, Robert Stake, one of America's foremost authorities on evaluation (Stake, 1967) has suggested that the total cost of measuring such objectives may be one hundred times that of administering a forty-five minute standardized paper-and-pencil test; and the amount of time, personnel, and facilities necessary for such evaluation may be astronomical. Stake also points out (Stake, 1967) that the errors of testing increase markedly when we move from highly specific areas of performance to items which attempt to measure higher mental processes and unreached human potential. According to Stake, the only reason we have tolerated the test error in standardized instruments is because very few important educational decisions are ever based on test scores alone.

B. Measurement and Statistical Problems
Measurement and formal testing often play a major role in evaluation studies, but certain cautions are necessary when we consider the use of standardized tests in evaluating programs for the gifted. In addition to the measurement problems implicit in

the above discussion on Higher Level Objectives, problems often arise when we attempt to use norm referenced tests developed for general populations. Conventional standardized tests are based on the normal distribution curve and for this reason the equality of units of measurement is open to serious question. The main issue in using age, grade, or percentile norms is that we cannot assume that a year's growth or growth in a given number of percentile points is a uniform unit. Thus, for example, if the performance of an average student increases from the 40th to the 50th percentile over the course of a school year, we cannot assume that this is a greater gain than that made by a gifted student whose score increased from the 90th to the 95th percentile. The gifted student initially scored at the upper end of the normal curve where it is much more difficult to show an increase in percentile score points. The same is true for age and grade scores. Generally, there is a slowing down of gains at the upper levels of most performance tests that were normed on the general population. For this reason, when the evaluator uses standardized tests, he should avoid making comparisons between gifted students and other populations. This can be done by developing separate sets of norms for each distinct population whose growth is being evaluated, provided of course, that the test has a broad enough range to allow students to show maximum growth. If a test does not have enough "top" in it, highly able students may score at the upper limits, but we will be unable to determine their true growth because of the low ceiling of the test. Since many standardized tests are designed to provide achievement information for the vast middle ranges of ability, their content and interpretive data may not be valid for children who deviate markedly upward from the mean.

The use of conventional tests with gifted and talented students also presents some problems in the statistical treatment of evaluative data. As was pointed out earlier, test reliability is a function of group diversity—the more heterogeneous the group the higher the reliability. Since gifted groups frequently are, by definition, relatively homogenous groups, and therefore frequently show a narrower range of test scores than the population in general, we should be extremely cautious when viewing the reported reliabilities of standardized tests. Unfortunately, most test publishers do not report reliabilities for subpopulations within their standardization sample and therefore it may be necessary to conduct a "local" reliability study whenever conventional tests are used with special populations.

One of the major statistical problems encountered when working with the test scores of superior students is the well-known "regression toward the mean" effect. Although this is complicated statistical phenomenon, simply stated it means that predicted scores tend to "move-in" toward the mean of the distribution. Thus, if we are using a pretest and post-test design to evaluate the effects of a program for the gifted, and if the students" scores on the pretest are initially high, it is quite likely their post-test scores will actually decrease due to the regression effect. It is for this reason the evaluator must exercise a great deal of caution when considering the pre/post design and other statistical designs that do not take into account the lack of normality in the distribution of gifted students" test scores. When pretest and post-test scores are used,

it may be necessary to explore the use of non-parametric statistics or multivariate methods of analysis.

      C. *Practical Problems*
      The evaluation of programs for the gifted, like evaluation in all other areas, requires time, money, and trained personnel. When evaluation is "tacked on" to a program as an afterthought, and when the human and financial resources necessary for carrying out a comprehensive evaluation are not available, program personnel may very well end up being asked to do the impossible. Although these are practical problems they can, nevertheless, have as much influence on the value and quality of an evaluative study as the measurement and statistical problems discussed above. Indeed, practical problems more often than not underlie or give rise to more complicated problems in measurement and design.

      What can be done about practical problems in evaluation such as time, money, and personnel? The answer to this question often rests with persons who are responsible for drafting the guidelines for special programs and/or those persons who actually develop the programs. While it seems almost trite to say that provisions and resources for evaluation should be included or "written into" guidelines and proposals, the fact is that without such provisions evaluation becomes a game (sometimes even a farce) that can serve very limited purposes. Many evaluation specialists have now reached the point where they are simply saying that a program may be "unevaluatable" because program planners have not given serious attention and resources to the evaluation component.

      Almost all writers in the field of evaluation have stressed the importance of planning evaluation activities from the very beginning of any educational endeavor. A good evaluation plan can continually bring to the attention of program developers the steps that must be taken and the resources that must be allocated if evaluation is to serve useful purposes. Early and continuous concern for evaluation will help to overcome many of the difficulties that arise when evaluation is tacked-on as an afterthought. Another practical problem relates to the attitude that many educators hold toward evaluation. Teachers and other professional personnel often view evaluation as a means of controlling or checking up on a program and the persons responsible for operating a program. In short, evaluation can be a very threatening affair that might result in some rather harsh actions, especially if the evaluation is mandated by a decision-making body or outside funding agency.

      Although negative or at least cautious attitudes toward evaluation are not easy to overcome, it is important to take steps toward dealing with this problem. The most obvious action that can be taken with regard to this problem is to create a positive atmosphere of helpfulness rather than destructiveness. An evaluation plan should focus on the "good things" that are happening in a program just as much as those aspects of a program that are in need of change. It should also explain this positive focus to teachers and administrators in terms of the ways in which an evaluation can help each particular group. Unless persons being evaluated can see some value and benefit *for*

themselves as a result of participating in an evaluative study, they are likely to approach the process halfheartedly; or even worse, they may actually try to distort evaluative information.

References

Callahan, C. M., Covert, R., Aylesworth, M. S. & Vance, P. (1981). *SEA Test.* Evaluation Research Center and Department of Foundations of the University of Virginia, Charlottesville, Virginia.

Renzulli, J. S. (1977). *The Enrichment Triad Model: A guide for developing defensible programs for the gifted.* Mansfield Center, CT: Creative Learning Press.

Scriven, M. (1967). *Perspectives of curriculum evaluation.* AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally.

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*(7), 523–540. https://doi.org/10.1177/016146816706800707

Steele, J. M. (1969). *The class activities questionnaire.* Mansfield Center, CT: Creative Learning Press.