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

Evaluating programs for the gifted: four questions about the larger issues

Abstract

This paper discusses four major issues concerning the evaluation of programs for the gifted:

1. The need to evaluate gifted programs so that planning, development and accountability form a natural sequence of educational activities.

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 program.

4. It must be recognised that the evaluation of gifted programs requires awareness of the problems associated with assessing higher level abilities and to use standardized tests and the practical demands 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 school 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 reports 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 that accurately reflects the needs of the gifted programs have been launched as last-ditch efforts to save programs that are in danger of being eliminated or simply reduced in the amount of funds 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 survival is a carefully planned and comprehensive evaluation that will accurately document all aspects of the services being provided for gifted and talented students. 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 equalled 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 existence of a particular program was frequently locked upon as some sort of malcontent, especially if the program in question was placed in the mantle of innovation, launched in the wave of "defining", 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.
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 evaluation? How do actions of decision makers affect the program?
2. What actions do decision makers have control over?
3. What information is necessary for making decisions?

Perhaps the best way to illustrate the interconnectedness and the importance of these considerations is by developing a hypothetical example. Let us suppose that we are evaluating an anthropology program for gifted students at a junior high school. The school administration is concerned about the program's effectiveness and the dropouts that occur in the junior high years. A team of research assistants is brought in to study various dropouts where students attempt to study differences between communities by analyzing the types of objectives that are effective. 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 continue the program, perhaps even reject the entire concept, because some question whether the board is concerned with 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 questionnaires and interviews with a random sample of parents.

Another decision maker in this situation might be the school principal. He may be up to him or her to decide when students should be on school trips or where they travel, and the principal might set the standards for the number of days the school can travel. Rules are often set as a result of regulations but could be established as a result of the principal's decisions. Thus, the principal might be found at fault because students are not excused from some of their regular classes. The principal may require information about the availability of educational instruments such as the SEA Test (Callahan, Covert, Aylesworth, Vanco, 1961) and the Class Activities Questionnaire (Steinle, 1969). These instruments can be used specifically for evaluation purposes 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 provided, although the evaluation process involves decisions about which educational programs to eliminate which effect learning.

It should be emphasized that the conditions which affect 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 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 collecting data to determine the frequency of our year-long effort to implement this objective. Further data might result from interest questionnaires distributed during the school year. 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.

Decision making is a fundamental goal of evaluation. Unless persons being evaluated can see some value and benefit for themselves as a result of participating in an evaluative study, they will have no stake in the process halfheartedly; or even worse, they may actually try to distort evaluative information.

An evaluation plan should focus on the ‘good things’

The third objective of program evaluation listed above— to determine the overall effectiveness of programs; that is, to determine which programs affect a program— calls attention to the fact that a successful program is frequently the result of policy decisions and actions that may influence institutional and 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 since the selection of teachers (and the 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; that is, any 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 deal of attention is given to the experimental situations while the prescribed activity is taking place. Evaluation on the other hand, is concerned with program improvement and providing feedback to the program. The evaluation of objectives is essential, they would also support the position that an evaluation study should investigate any and all conditions that give rise to the 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 ineffective 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 drop out of any evaluation attempts. 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 schematically 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 pinpoints by careful examination the circumstances and conditions that result in identifiable changes in performance, attitude, or other indicators of program effectiveness. In order to evaluate an act as constructive and positive in the overall area of education, it is essential to evaluate as many of the following objectives as possible.

1. To determine whether and how effectively the objectives of a program are being fulfilled.
2. To determine unanticipated or 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 achievement of objectives is essential, they would also support the position that an evaluation study should investigate any and all conditions that give rise to the 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 ineffective 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 drop out of any evaluation attempts. 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.

The third objective of program evaluation listed above— to determine the overall effectiveness of programs; that is, to determine which programs affect a program— calls attention to the fact that a successful program is frequently the result of policy decisions and actions that may influence institutional and 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 since the selection of teachers (and the 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; that is, any 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 deal of attention is given to the experimental situations while the prescribed activity is taking place. Evaluation on the other hand, is concerned with program improvement and providing feedback to the program. The evaluation of objectives is essential, they would also support the position that an evaluation study should investigate any and all conditions that give rise to the 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 ineffective 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 drop out of any evaluation attempts. 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 schematically 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 pinpoints by careful examination the circumstances and conditions that result in identifiable changes in performance, attitude, or other indicators of program effectiveness. In order to evaluate an act as constructive and positive in the overall area of education, it is essential to evaluate as many of the following objectives as possible.

1. To determine whether and how effectively the objectives of a program are being fulfilled.
2. To determine unanticipated or 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 achievement of objectives is essential, they would also support the position that an evaluation study should investigate any and all conditions that give rise to the 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 ineffective 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 drop out of any evaluation attempts. 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.
However, the nature of gifted programs and their concern for developing higher thinking processes may make this model too rigid for the gifted and talented. In fact, when the behavioral objectives approach is used in its most rigid form, the gifted student, when shown evidence that performance standards is high, may lose interest in the trivial rather than important behaviors of superior learners.

A gifted educational objectives model is inappropriate for programs for the gifted because it forces us to be primarily concerned with those behaviors that are easily measured and that can be broken down into discrete teachable tasks. For example, the goal of teaching the dog to give a hand. McSweeney, the 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 their 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 and selecting meanings in the light of different points of view. We talk about producing unique plans, communication forms, and products. And we also talk about some less elaborated 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 are not as specific and practical requirements necessary for good evaluation studies. While virtually hundreds of relatively valid and reliable tests have been developed 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 not be or cannot be justified.

A second dimension of this problem is that gifted programs are frequently characterized by highly individualized, self-paced, self-directed reading skills program for average or slow learners may have enough uniformity in its objectives to warrant large scale standardized tests. 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. A test used in a small school district may be statistically significant pre-test to post-test gains only when a few students are being evaluated with a given instrument. Standardized tests can, of course, be used effectively in evaluating programs for the gifted if they (a) are valid (appropriate) measures of particular objectives, and (b) if they are adjusted to 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 preschool education, it is impossible to equate standardized tests with the appropriateness of tests based on systemwide or nationwide objectives.

The problem has been a great deal of concern in education about the specification of objectives in terms of observable and measurable student behaviors. Many research, however, has developed the behavioral objectives model as a panacea for conducting evaluation studies.

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

A. The Problem of Lower Level Objectives

Programs for the gifted are often characterized by a commitment to developing higher order skills of mind and advanced levels of awareness, interest, and other affective behaviors. This presents a somewhat unique evaluative 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 processes, it becomes increasingly difficult to find measures of the type of specific and practical requirements necessary for good evaluation studies. While virtually hundreds of relatively valid and reliable tests have been developed 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 not be or cannot be justified.

B. Measurement and Statistical Problems

Measurement and formal testing often play a major role in a gifted program. 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 reason for this is that the normal distribution curve does not assume that a year’s growth or growth in a given number of percentile points is a uniform unit. Thus, for example, if the norm referenced test is 10th grade level 25th percentile, it does not mean that a year of growth in the 40th to 50th percentile over the course of a school year, we cannot assume that this is a greater gain than that 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 than to maintain the same score level.

Generally, there is a slowing down of gains at the upper levels of most performance tests that were normed on the general population. For example, when the evidence-writer uses standardized tests, he should avoid making comparisons between gifted students and other populations. The population of gifted students varies from one to another; 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 gains is. 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 quality and potential, for this reason most of the test. If 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 may cause serious problems in the standard method of 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 homogeneous 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, many gifted teachers fail to report reliability data for subpopulations within their standardization sample and therefore it may be necessary to conduct a ‘local’ reliability study using the same conventional tests are used with standard populations.

One of the major statistical problems encountered when working with the gifted and superior students is the well-known ‘regression toward the mean’ effect. Although this is complicated statistical phenomenon, simply stated it means that there is a tendency for test scores to move toward the mean when the group scores are high; it is quite likely their post-test scores will actually decrease due to the regression effect. It is for this reason that many educators place a great deal of caution on the interpretation of the pre/post design and other statistical designs that do not take into account the lack of normal distribution curve. If a program is testing with a pretest and post-test scores are used, it may be necessary to explore the use of non-parametric statistics or multi-variate 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 ‘lacked’ on to a program as an add-on or extension of the human and financial resources necessary for carrying out a comprehensive evaluation are not available, program personnel may very well adopt ‘the method not the principle’ attitude. This is impossible. Although these are practical problems they can, nevertheless, have as much influence on the value and quality of an evaluative study as the methodology and made a major impact on the validity of critical problem identified in the discussion 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, writing, and implementing portions of the program who actually do the programs. While it seems almost futile to say that provisions and resources for evaluation should be part of the program, the fact is that without such provisions evaluation becomes a game (sometimes even a farce) that can serve very limited purposes. Many evaluators and specialists have now reached the point where they are simply saying that a program may be ‘unevaluable’ because program planners have not given adequate thought and attention 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 planners problems that must be taken and 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 tackled 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 it. A good evaluation plan can be a very threatening affair that might result in some rather harsh actions, especially if the evaluation is made a part of a decision-making body or outside funding agency.

Although negative or at least cautious attitudes toward evaluation are not to be wished away, the next steps toward dealing with this problem. The most obvious action that can be taken with regard to this program is to create an atmosphere that is more favorable and constructive. An evaluation plan should focus on the ‘good things’ that are happening in a program just as much as those that need improvement. If the evaluation plan is to be successful, it should also explain this positive focus to teachers and administrators in terms of ways in which an evaluation can help promote good teaching and learning. An evaluation plan can see some value and benefit for themselves as a result of participating in an evaluative study, they are like other professionals. This can help to reduce the worse, they may actually try to distort evaluative information.

References

Cullahan, C. M., Covell, R., Ayersworth, M. S. & Voico, P. SEA Test. Evaluation Research Center and Department of Foundations of the University of Chicago, 1963.


Sims, M. Perspectives of curriculum evaluation. AERA Monograph Series, 1966.
