

Asking The Right Questions: The Central Issue In Evaluating Programs For The Gifted and Talented

Carolyn M. Callahan
Associate Professor
University of Virginia

Abstract

What is an appropriate focus for gifted program evaluation? How can we justify focusing program evaluations on questions connected to improving the program rather than judging its success? The author suggests and justifies criteria for framing evaluation questions.

Within every field of scholarly inquiry the most significant advances have come from the inquirer's unique ability to ask different and extending questions, to focus on the appropriate hypotheses and to select a strategy that most effectively and accurately answers the questions posed. The field of evaluation theory and practice is no exception. If the evaluation of gifted and talented programs is to yield valid assessment data and is to have a significant impact on the improvement of gifted programs then more serious attention must be directed toward framing evaluation questions that address the **relevant, useful and important** issues facing programs.

Relevancy, Usefulness and Importance

The concepts of relevancy, usefulness and importance are at the core of the problems facing the development of significant evaluation questions. Relevancy refers to the degree to which the questions actually address the functioning of the program under consideration, its components, its activities, its goals, and its structure. Evaluation questions are *not* research questions. Our purpose is *not* to address generalizability, but to address specificity—to the program under consideration. We do not seek to validate a universal identification system; we wish to validate the effectiveness and efficiency of a given system for identifying the students appropriate for a given program. Useful questions provide data that some audience can actually use in the process of making decisions about a program. To know that students failed to achieve specified objectives is more useful data if we can also identify problems in the curriculum or delivery which might be remedied. Important questions yield data helpful in making decisions which will have impact.

Issues/Constraints in Identifying Appropriate Evaluation Questions

In the formulation of evaluation questions for gifted and talented programs, one is faced with a particular set of

issues/constraints which overlap those of all other education programs, another set of issues which are shared in the evaluation of any special project and then a set of issues unique to the evaluation of programs for the gifted

Similarities to the evaluation of programs for other exceptionalities. Like programs for the handicapped, many programs for the gifted are predicated on the development of individualized programs of study, on the establishment of goals which are particular to the interests and abilities of the individual child, e.g., Renzulli's Enrichment Triad Model (1977) and Renzulli and Smith's Individualized Educational Programs for the Gifted (1979). In such programs it is unreasonable to set group behavioral objectives against which one can assess accomplishment. In addition, the provision of programs for a special population makes the establishment of control groups which can be used as a comparison a very difficult, and sometimes impossible, task. Ethical and political considerations nearly always dictate that all eligible children be served.

Unique issues in evaluating program for the gifted. Unlike programs for other exceptional students, programs for the gifted and talented face the difficulty of setting standards of performance against which achievement can be measured, or even against which program operation can be judged. First of all, the kinds of programs offered to gifted students vary considerably from school to school and no agreed upon standards of good programming exist within the field of gifted education for use in establishing criteria for specific program functioning. Secondly, many of the objectives in programs for the gifted are very complex and not easily defined. For example, the concepts of creativity and critical thinking are hypothetical constructs whose definitions have been the subject of much debate in psychology for twenty years. Furthermore, there are no empirical studies which provide norms or guidelines for "expected growth" over a specified time period in either the traditional academic areas or the more unique areas such as critical thinking, creativity, or higher level thinking skills. Therefore, it is unsound practice to phrase evaluation questions in terms of expectations of the attainment of certain gains on a specific achievement test

An even more fundamental issue in phrasing questions when evaluating programs for the gifted is the validity issue. Validity of both the questions asked and the instruments selected to use in gathering data relative to these questions are often not given serious consideration in the development of evaluation designs. First, why is this a particular

problem in evaluating programs for the gifted? The reasons lie in the uniqueness of the types of goals and objectives set for these programs. Because many goals of gifted and talented programs are both product oriented and individualized, we often must reject altogether questions phrased in terms of test score gains because paper and pencil assessments simply will not validly assess the program's goals. For example, a goal of developing independent study skills and the use of alternative resources is often translated into a question focusing on gains on a "Use of Sources" subtest of a standardized achievement test. Whereas the program staff intended to develop students' abilities to identify appropriate new, non-standard sources, to use a variety of types of resources and to use more sophisticated resources, the tests assess students' ability to use an index or table of contents. Second, many of the statements of goals and objectives of programs for the gifted tend to be holistic and long-term, making their evaluation very difficult and often resulting in the formulation of very short term, but invalid evaluation questions. If, in fact, these short term questions relate to enabling objectives then collection of data about those issues will be useful. But too often we stop at the point of evaluating students' increased fluency scores and fail to examine completely the question of an increase in creative productivity. Or we evaluate whether or not identification procedures were carried out as prescribed but fail to evaluate whether *gifted* students were actually identified by those procedures.

Issues common to all educational program evaluation. The last major issue is neither new nor unique to gifted programs. That is the issue of the *focus* of the evaluation and the questions which emerge from that focus. This issue was first addressed in depth in the field of evaluating programs for the gifted and talented by Renzulli (1975). Bemoaning the generation of evaluation questions which focused only on end products of programming efforts and which tended to use inappropriate standards for judging the quality of gifted program activities, he pointed to the importance of constructing both formative and summative evaluation questions and questions relating to process, product and presage information. His comments reflected the need not simply to judge a program as "successful" or "unsuccessful," but to help identify those aspects of a program which were functioning as they should and, therefore, would be likely to contribute to a successful program.

Current Directions

The current direction taken by evaluation theorists and practitioners is to extend these ideas further by expanding on the *purposes* of program evaluation in order to address the relevancy, utility and importance issues mentioned above. The practice of formulating evaluation questions using an expanded concept of the purposes of evaluation originates from the time that Scriven (1967), Stufflebeam (1968), and Stake (1967) first began exploring the concept of evaluation as the process of gathering data for the purpose

of decision-making. Up until that time, evaluation paradigms had most often mimicked research paradigms. Both evaluators and those being evaluated tended to think only in terms of very reductionistic, experimentally phrased and outcome oriented evaluation questions such as: Did the students become better independent learners as a result of this instructional program? Are the teachers using questions which require higher level thinking as a result of the staff development program? etc

Although such outcome questions legitimately remain one of the foci of evaluation efforts, the work of the late 1960's and the 1970's began a trend in evaluation directed toward asking questions and providing information which are of greater utility to the program being evaluated, to increasing communication and to addressing those issues fundamental to program planning. The application of these expanded purposes to the evaluation of programs for the gifted and talented has been elaborated on by Callahan and Caldwell (in press) and include: documentation of the need for a program, documentation of the case for a particular approach, documentation of the feasibility of a program, documentation of program implementation, identification of program strengths and weaknesses, provision of data for inprogress revisions of the program, documentation of the results or impacts of the program, and explanation and description of the program to interested and uninformed audiences. Each of these program purposes may serve as the basis for the generation of a unique set of evaluation concerns and questions depending on the context of the evaluation and the audience to be served (see Audiences). Questions derived from these evaluation purposes address specific program outcomes in the minority of cases. More often they address the process of describing and explaining a program, provide information rather than value judgments, and leave the evaluation process open to input according to the needs of the program.

New Methodologies The introduction of new methodologies borrowed from the social science research arena (e.g., ethnographic research borrowed from cultural anthropology) (Guba, 1978) has also contributed to a dramatic revision of the types of evaluation questions posed as well as to the fundamental approach to evaluation. Because these "naturalistic" approaches differ from traditional methodologies in philosophical base, inquiry paradigm, purpose, framework/design, style, view of reality, value structure, setting, scope, context, conditions, treatment definition, and methods (Guba, 1978), it is inevitable that the questions which emerge from designs based on these methodologies will differ from traditional modes of inquiry. For example, naturalistic methodologies tend to be emergent and variable in nature rather than preordinate and fixed. Thus an evaluator who is applying a naturalistic strategy may or may not start an investigation with a series of specifically stated hypotheses (evaluation questions), but it is almost inevitable that the process of studying the program will yield new questions and new strategies for answering

those questions as the investigator uncovers new information. An individual pursuing a non-traditional education approach would be likely to phrase evaluation questions in a more general, open-ended manner.

At the risk of oversimplifying the process of naturalistic inquiry and exaggerating the distinctions between qualitative and quantitative methodologies, the following is offered as an example of the contrasting questions which emerge. A traditional evaluation design might pose the question: "Do the scheduling process and curricular adaptations reflect the individual needs of gifted students?" A traditional evaluator is likely to define the "need" very specifically and may focus on aptitude as the variable to consider. The methodology selected may be to rate student programs on the degree of individualization in student records (programs) and to present students with surveys to complete. A naturalistic inquirer would probably begin with a question more like: "On what bases are the schedules and curricula for gifted students modified? Are student needs being met by these modifications?" It would then be appropriate to conduct in-depth interviews with students, teachers, counselors, parents and administrators. In the process, one individual might mention that "Yes, my schedule is different, but it's not what I want to do." The evaluator's question may then become, "Are relevant variables (i.e., student interest in this case) being considered as part of the scheduling process?"

Phrasing of evaluation questions. The phrasing of the evaluation questions also suggest the difference between the traditional evaluation questions and the more current notions of evaluation. The differences between the type of question asked for research purposes and one asked for evaluation purposes are those of generalizability and those described by Yavorsky (1984) as a distinction between asking "How should a program be planned to work?" and "How well is the planned program working?"

Audiences. The audience(s) to be served by the evaluation is also a key factor in determining the appropriate evaluation questions. As Dinham and Udall (in press) point out, the purpose of evaluation should be formed by looking at the specific needs of specific audiences. It follows that the evaluation questions should focus on issues which are of specific relevance and importance to specific audiences. Yavorsky (1984) identifies two distinct groups of audiences when she gives directions for the selection of appropriate evaluation concerns—external and internal audiences. External audiences consist of such individuals and groups as funding agencies (local, state or federal government; private foundations, etc.), higher levels of administration, professional associations and certification agencies, consumer agencies (e.g., parent groups), community and assorted interest groups. Internal audiences are those which are directly involved with the functioning of the program and would include such groups as program administrators, teachers in the program and in the school division, central office and building administrators, stu-

dents in the program, etc.

It is likely that the interests of each of these groups would overlap to some degree and the same evaluation question may provide information to a variety of audiences. For example, those individuals who work very closely with the students (program staff, regular classroom teachers, parents, etc.) will identify questions that relate to the impact of the program on the student both academically and psychologically. On occasion, however, different audiences have very different information needs. Funding agents and administrators often express different concerns than staff, parents and students—concerns about the cost of the program.

It is important that an evaluation designer give all relevant audiences the opportunity to voice their evaluation concerns and questions. Renzulli (1975) suggests the use of an input questionnaire and interviews. His recommendation is that any audience which is either directly involved in the program (students, parents, staff) or indirectly involved (non-program staff, counselors, etc.) should have the opportunity to raise concerns, to suggest questions that should be addressed, and to voice their opinion about types of information that should be collected in evaluating the program.

There may also be concerns and questions which should be addressed even though they may not be specifically mentioned by one of the audiences addressed in the original evaluation planning. For example, one should consider the kinds of data such as unintended outcomes as described by Scriven (1967) when he talks of goal-free evaluation. In the process of examining effects of a computer class on gifted students one may uncover telltale grumbling by other teachers that this class interferes with the equal access to computers that they believe all students should have or one may discover that assignments in other courses are left incomplete. It can be predicted that such situations, if not addressed, will have negative impacts on the continuation of the class. Even though no audience might have anticipated such a situation, it is important to attend to such evidence even though it was not included in the original planning.

Several of the evaluation paradigms mentioned earlier as part of the group labeled naturalistic inquiry have potential for addressing specific audience needs as well as uncovering new questions. For example the Responsive Model of Stake (1975) "orients more directly to program activities than to program intents" and "responds to audience requirements for information" (p. 14). Other models mentioned by Guba which seem to fall in the same category are the Judicial Model (Wolf, 1975), the Transactional Model (Talmadge, 1975), the Connoisseurship Model (Eisner, 1979), and the Illumination Model (Parlett and Hamilton, 1976).

The Importance of Program Description in Formulating Evaluation Questions

Reaching the point where we can identify questions

which are, in fact, audience responsive, appropriate for assessing the functioning and goals of gifted programs, and measurable is dependent on being able to describe the program. Experience in evaluating programs for the gifted demonstrates very quickly that one cannot evaluate what one cannot describe.

The most thorough consideration of all possible evaluation questions will result if care is taken to describe the program in terms of its components (e.g., identification process, management, curriculum and instructional strategies, programming options, etc.), the functional relationships between these components, the resources which support each component, the activities of each component, and the expected outcomes of the activities carried out as a function of each component. (See Yavorsky, 1984). If the evaluator and program administrator can identify each of the aforementioned attributes of the program, then evaluation questions will naturally emerge. Further evaluation questions should emerge from an examination of this description by those other audiences who have a need for evaluative data. Very simple questioning by the evaluator should yield a list of areas in which individuals and groups can reflect their concerns.

Guidelines for Formulating Evaluation Questions

Guidelines for formulating the evaluation questions from these concerns and prioritizing questions are suggested by Yavorsky (1984). The first of these guidelines is to construct and give priority to questions of concern to internal and external audiences. The second is to attend to questions relating to areas of the program that are of central functional importance. That is, success of the program is highly dependent on the success of that component. Third, one should identify questions that are suggestive of problems. These are often generated around areas that are not well defined, where program design is in question or is in controversy, where there has been a history of problems—either in that particular program or gifted programs in general, or areas marked by staff disagreement. Finally, one should identify those questions where information is needed soon.

Covert (personal communication, July 5, 1985) suggests that one also should generate questions relative to the adequacy and availability of resources which are needed to execute the program, the adequacy of planned activities, the degree to which implementation follows plans or is appropriately modified, and the degree to which the components of a program are in accord with one another and operate as a unit. Questions of resources are important in helping programs deliver what is expected or identifying reasons why the program fails to deliver. For example, if the identification of a musically talented student is dependent on past performance and the school has no regular music program, then identification is a more difficult and expensive process. Questions of the adequacy of planned activities are important for helping prevent unnecessary failures.

If one can identify faulty logic in planning, a considerable savings in time and energy are possible. If administrative staff development does not precede teacher staff development, the effectiveness of the latter may be seriously impaired. Questions of implementation are necessary to determine whether the achievement of expected outcomes (or lack of achievement) is attributable to the program as planned. Questions of accord are necessary to help identify potentially conflicting plans or implementation problems. A school may plan for students to engage in a mentor program but so tightly schedule advanced placement classes that students have no opportunity to participate.

Guba (1978) points out that the naturalistic evaluator is "materially assisted by keeping in mind that there are certain standard situations in which the persons involved may see the entity being evaluated as giving rise to issues or concerns" (p. 51). Among those he lists are:

- Undesirable consequences of an interaction or course of action (Parents may feel that a resource room may have a negative effect on other students' self concepts.)
- Confusion regarding courses of action (Teachers may not understand their role in the identification process.)
- An undesirable deviation from older practice (After all, the regular classroom was good enough for me.)
- Conflicts with traditional values (The gifted program may be perceived as contrary to the "American value of equal education.")
- Conflicts with personal values (Activities focusing on social issues in a gifted program may conflict with community values.)
- Potential loss of power (Principals may see the coordinator of gifted programs as having too much authority over the teachers in his or her building.)
- Potential economic threats (Teachers of regular classrooms may perceive that the resource room drains resources from the regular class instruction.)
- Perceived inconsistency with a suggested course of action (If the classroom teachers recommended a resource room program and the school chose to deliver instruction in a cluster group arrangement, the teachers are likely to be disgruntled and perhaps less cooperative.)
- Lack of understanding of rationales or goals (If a teacher does not understand the goals of program, he or she will not be effective in implementing the curriculum.)
- Bias based on a negative personal experience with a given institution (Just one example of a student who was not successful in an accelerated school program can seriously tarnish the image of acceleration programs in general.)
- Potentially harmful side effects (Does being in the gifted program result in a negative influence on social interactions?)

An additional factor which should influence the selection of evaluation questions is the degree to which the question

is "answerable." Unfortunately, in the field of evaluation of gifted programs we usually find two, equally unacceptable, ways of dealing with this factor. On the one hand, we find ourselves answering only very simple, easy to answer but relatively insignificant or narrow questions because there are available instruments or because data can be collected easily. For example, we often find evaluation studies focusing only on student, parent and teacher perceptions of programs or curriculum as indicated on questionnaires. The narrowness of the evaluation questions posed in these studies results in a tendency to ignore halo effects, failure to gather in-depth information on what influences the perceptions of the individuals responding, and failure to verify actual changes or outcomes of program activities.

This is not to suggest that perceptions are unimportant; rather we must consider whether perceptions in and of themselves are sufficient evidence for good decision making. At the other extreme, the evaluator of gifted programs sometimes asks questions at either such a complex or undefined level that it is virtually impossible to find strategies to evaluate the program because the questions do not lend themselves to assessment or documentation.

Evaluation questions have also been limited by the inability of evaluators to look at alternative data sources such as objective and subjective product assessment (performance rating scales, essay ratings, independent project ratings, etc.) The tradition of assessing impacts on the basis of multiple-choice test items severely limits the kinds of evaluation evidence we can gather and transmit to interested parties. Renzulli and Reis (1985) and Callahan (1983) have offered product rating scales which provide alternatives for gifted and talented student product ratings.

One final extension of evaluation questions that should be considered is into the realm of cost effectiveness/cost benefit analysis (CEA/CBA). It is important that we raise questions about the misleading conclusions that can be drawn by simply stating that programs produced statistically significant changes in the students. In times of limited funding for special programs, it is important that we consider the factors of cost relative to the benefit to students in evaluating programs.

A Direction for Further Research

One final note on the issue of evaluation questions as it relates to this special issue of *Gifted Child Quarterly*. This brief piece does not address the issues of research in the area of program evaluation. In fact, the topic of evaluation research (as opposed to the practice of program evaluation) is a distinct topic and practically non-existent in the field of gifted education. The paradigms for evaluation which have been described and implemented in the literature in the education of the gifted and talented are drawn largely from the more general field of program evaluation and have not been subject to rigorous tests of usefulness, generalizability, impact, etc. that characterize the field of research. There is,

therefore, a need to more closely and systematically examine the kinds of evaluation questions asked in terms of their relevance (as perceived by the audiences), their usefulness (as evidenced by actual decision-making based on evaluation findings), and their importance (as determined by the judgments of experts, the program and school administration, the staff of the program, etc.).

The topic of evaluation questions is not and cannot be an entirely independent issue. One must consider the evaluation question in light of the design or paradigm which has been selected for the evaluation. And this is a "Which comes first—the chicken or egg?" question in evaluation. Do we formulate questions and then choose a design or do we choose a design and allow the questions to emerge from our design approach? It is most important to remember the mutual dependencies and collect and report data which acknowledges the interrelationships.

References

- Callahan, C. M. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly*, 27(1), 3-7.
- Callahan, C. M., & Caldwell, M. S. (In press). Defensible evaluations of programs for the gifted. In C. J. Maker (Ed.), *Critical Issues in Gifted Education*. Rockville, MD: Aspen.
- Dinham, S. M., & Udall, A. S. (In press). Evaluation for gifted education: Synthesis and discussion. In C. J. Maker (Ed.), *Critical Issues in Gifted Education*. Rockville, MD: Aspen.
- Elliot Eisner (1979). *The Educational Imagination*. New York: MacMillan.
- Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. *CSE Monograph Series in Evaluation*, 8.
- Parlett, M., & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovative programs. In G. V. Glass (Ed.), *Evaluation Studies Review Annual (Volume 1)*. Beverly Hills, CA: Sage Publications.
- Renzulli, J. S. (1975). *Working draft: A guidebook for evaluating programs for the gifted and talented*. Ventura, California: Office of the Ventura County Superintendent of Schools.
- Renzulli, J. S. (1977). *The enrichment triad model: A guide for developing defensible programs for the gifted and talented*. Mansfield Center, Connecticut: Creative Learning Press.
- Renzulli, J. S., & Reis, S. M. (1985). *The schoolwide enrichment model*. Mansfield Center, CT: Creative Learning Press.
- Renzulli, J. S., & Smith, L. H. (1979). *A guidebook for developing individualized educational programs for gifted and talented students*. Mansfield Center, CT: Creative Learning Press.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum Evaluation (American Educational Research Association monograph series on evaluation, No. 1)*. Chicago: Rand McNally.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523-540.
- Stake, R. E. (1975). Program evaluation, particularly responsive evaluation (Occasional Paper No. 5). The Evaluation Center, Western Michigan University.
- Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision-making*. Columbus, Ohio: Evaluation Center, Ohio State University.
- Talmdage, H. (1975). Evaluation of local school community problems: A transactional evaluation approach. *Journal of Research and Development in Education*, 8, 32-41.
- Wolf, R. L. (1975). Trial by jury: A new evaluation method. *Phi Delta Kappan*, 1975, 57.
- Yavorsky, D. K. (1984). *Discrepancy evaluation: A practitioner's guide*. Charlottesville, VA: University of Virginia, Curry School of Education Evaluation Research Center.