

---

# *Introduction to Program Evaluation in Gifted Education*

Carolyn M. Callahan

*University of Virginia*

Program evaluation has been considered an important, but neglected, component by experts in the field of gifted education for at least the last three decades (Gallagher, 1979; Renzulli & Ward, 1969), but yet, the guidance given to the field about this essential area has been limited. *Gifted Child Quarterly* has been one vehicle used to provide important direction to evaluators and practitioners.

## **CATEGORIES OF EVALUATION**

The literature on evaluation, and in particular the literature on the evaluation of gifted programs, may be divided into four categories. Included in the first of those categories are the manuscripts that provide theory and/or practical guidelines. These guidelines sometimes include particular suggestions for the evaluation of gifted programs in general or for the evaluation of specific components of a program such as evaluation of staff development and sometimes provide examples of instrumentation and/or suggestions for instrument development. These articles are represented in this volume by Callahan (1986), Carter and Hamilton (1985), Kulieke (1986), Lundsteen (1987), and Reis and Renzulli (1991). The second category of evaluation articles describes or reports on

## xxiv Program Evaluation in Gifted Education

specific program evaluations. Such *Gifted Child Quarterly* (GCQ) articles as the VanTassel-Baska, Willis, and Meyer (1989) article summarize evaluations of the effectiveness of particular programs such as a self-contained program for gifted. The report by Avery, VanTassel-Baska, and O'Neill (1997) describes the evaluation of a suburban gifted program using multiple models, and Landrum (2001) documents the evaluation of a catalyst program using a consultation/collaboration approach to providing services for gifted students. The third category of program evaluation articles provides stimuli for the discussion of issues surrounding the evaluation process. For example, in this collection authors from outside the field of gifted education have published provocative ideas in GCQ that stimulate our thinking about alternatives to the evaluation of student performance. Baker and Schacter (1996) and Wiggins (1996) offer new ideas for the assessment of gifted students using expert performance as the basis for setting standards. The last category of evaluation articles, research on the evaluation process, is sadly missing from *Gifted Child Quarterly* and nearly absent from the literature on gifted program evaluation. Callahan, Tomlinson, Hunsaker, Bland, and Moon (1995) have conducted research on the factors that increase the likelihood that recommendations from program evaluations will be implemented. Hunsaker and Callahan (1993) examined the degree to which current evaluation practices in the field of gifted education utilized the multiple methodologies, sources, analysis techniques, and reporting formats recommended in the *Standards for the Evaluation of Educational Programs, Projects and Materials* (Joint Committee on Standards for Educational Evaluation, 1981). But these publications represent nearly the entire body of research on evaluation practice.

The articles that represent these four categories from GCQ in this volume have offered important guidance to the field because they raise significant issues. But the glaring omissions within each category and across categories suggest our work is not done.

### Category I: Guidelines for Evaluating Programs and Products

A set of common and useful principles cut across the articles in this category, with individual authors each offering unique variations on the themes. The importance of formative is stressed by Carter and Hamilton (1985) based on assumptions about the importance of gifted programming, but Callahan stresses the importance of addressing all questions of audiences served by the evaluation including summative evaluation questions. Commonalities or agreed upon principles of program evaluation include: carefully planning to ensure that the evaluation addresses the concerns of critical players in the program; narrowing the evaluation questions to those of most critical importance; ensuring that data collection is systematic; recognizing and acting on the importance of creating a collaborative relationship between evaluator and client; ensuring that student product outcomes be considered as one focus of the evaluation plan; ensuring that data collection is systematic, appropriate to the question

posed, and carefully weighed; and finally, that communication is clear and timely (Callahan, 1986; Carter & Hamilton, 1985).

One of the particular suggestions that deserves attention is that evaluations be designed to address not just “easy to answer” evaluation questions with easy to construct instruments, but that evaluations address the most important questions for good decision making, and that evaluation data collection use the most direct and valid tools. Using staff development as her example, Kulieke (1986) notes that while questionnaires will provide data on teachers’ perceptions of their needs and competencies, she also stresses the importance of direct observation in ascertaining the actual teaching behaviors and the quality of implementation of teaching strategies appropriate for gifted learners.

Kulieke has also made a significant contribution by providing an example of the use of the evaluation process as a valuable tool in needs assessment. While all evaluators talk of the use of evaluation data in the development of recommendations for program improvement, her discussion of needs assessment is a concrete example of the way an evaluation process can directly inform program decisions. Her outline of the ways in which evaluation data collected about particular types of inservice programs illustrates the range of purposes that evaluation can serve, as is outlined by Callahan and Caldwell (1986, 1995). Kulieke’s approach also exemplifies ways in which evaluation can provide information on the degree to which the program design meets identified needs of the target audience, the degree to which a program is implemented as designed, and the degree to which behaviors are changed or learning occurs as a result of program implementation.

Callahan (1986) and Carter and Hamilton (1985) both stress the importance of including key decision makers in the selection of foci for an evaluation. However, Carter and Hamilton define the decision makers as the administrators of the program and only suggest going beyond that group in the case where administrators do not have clear goals for the evaluation, while Callahan expands the initial concept of decision makers to include both internal audiences and external audiences who have a stake in the program’s effectiveness.

Another commonality of the Callahan (1986) and Carter and Hamilton (1985) approaches is the identification of essential components of gifted programs that should be considered as targets of the evaluation process. Both of these lists of essential components resemble the Key Features of Program for the Gifted that were identified by Renzulli (1975), but should now be updated to reflect the standards for excellence in gifted programs that have been elucidated by the National Association for Gifted Children to include: Program Design, Program Administration and Management, Curriculum and Instruction, Student Identification, Professional Development, Social and Emotional Guidance and Counseling, and Program Evaluation (Landrum, Callahan, & Shaklee, 2001).

As the process of selecting the more narrow questions within each of these categories evolves, the literature suggests questions of concern to internal and external audiences, questions relating to the central functioning of the program,

## xxvi Program Evaluation in Gifted Education

questions that history suggests represent potential problems that may inhibit good program functioning, and questions where information is needed soon for decision making should take priority (Callahan, 1986).

These authors also present the recurrent theme of the importance of selecting or creating reliable and valid instruments and strategies for collecting data for the decision-making processes of evaluation—whether the evaluation question focuses on process or student outcomes. Carter and Hamilton (1985) note the importance of considering the type of evaluation question and matching both the strategy and the particular data collection instrument selected to the evaluation questions. Like Wiggins (1996), Callahan (1986) and Reis and Renzulli (1991) point out the weaknesses inherent in using only paper-and-pencil, standardized tests to assess student outcomes. But, as this collection illustrates so vividly, little has been done to expand the available collection of alternative instruments that have been subjected to the rigors of psychometric examination. Kulieke (1986) offers an instrument of classroom observation based on the *Martinson-Weiner Rating Scale of Behaviors in Teachers of the Gifted* (Martinson, 1976), but does not provide reliability or validity data. The only alternative instrument offered with the expected, adequate psychometric properties is the *Student Product Assessment Form* (SPAF) (Reis & Renzulli, 1991). The review of the related literature in this article illustrates the field's lack of validated instruments to assess students' creative products, but provides good news about one instrument with excellent statistical properties that could be widely used.

### Category II: Descriptions of Specific Program Evaluations

The evaluation summaries of Avery, VanTassel-Baska, and O'Neill (1997), Landrum (2001), and VanTassel-Baska, Willis, and Meyer (1989) are valuable in illustrating how the general principles presented in the articles discussed in the first category might be applied, but they also illustrate the continued lack of development of instruments to assess student outcome goals. In presenting the results of an evaluation of a local district program, Avery, VanTassel-Baska, and O'Neill (1997) illustrate the ways in which the questions of classroom observations may be used to determine whether or not the curriculum and instructional strategies are differentiated appropriately. This evaluation also provides guidance in outlining the ways in which reporting and communicating results through multiple vehicles enhances utilization of results.

However, as Avery, VanTassel-Baska, and O'Neill (1997) point out, "the traditional linch [sic] pin of student performance" is not addressed in this report, (p. 124) nor is it addressed in the report of Landrum. However, Landrum (2001) has provided a creative example of using alternative data sources as a means of judging the effectiveness of implementing a new program model and using subjects as their own control over a period of time. By collecting data on: changes in use of classroom instructional strategies; the frequency of opportunity for gifted students to engage in activities that might be characterized as more rigorous; and standardized test performance, Landrum confirmed the

effects of the implementation of a consultation model on classroom behaviors of teachers.

The evaluation of the full-time, self-contained classroom (VanTassel-Baska, Willis, & Meyer, 1989) illustrates one strategy for finding a comparison group to use as a basis for attributing the changes in student test scores to the program intervention. One additional important consideration that is suggested by the analysis in this article is the need to present and discuss effect sizes. With such large samples we may find significant, but meaningful effects.

But these evaluations also point to the difficulty in establishing adequate comparison groups and the difficulty in identifying adequate instruments for measuring student outcome variables.

### Category III: Issues in Evaluation

A careful presentation of arguments for and against setting expert performance as the standard in assessing student performance in programs for the gifted is provided by Baker and Schacter (1996) and Wiggins (1996). Baker and Schacter also offer the possibility of using performances of teachers with content expertise to establish the levels of expertise we would set as goals for student achievement. A third possibility they offer is to use the performances of identified gifted students. Their discussion of these possibilities raises issues of developmentally appropriate considerations, but importantly, suggests areas of potentially valuable research in the area of evaluation of student outcomes. Wiggins acknowledges the potential resistance to setting standards using superlative performance as the goal, but argues, "faculties should always calibrate their local standards to such exemplars, anchoring the highest point in their scoring system with such examples of excellent performance. It is the only way for teachers as well as students to have valid, compelling, stable targets at which to aim" (p. 66). He further cautions against the overemphasis on process, form, and content in student products while stressing the importance of evaluating the effect of products if we wish to preserve the development of creativity—the degree to which the products "persuade an audience, satisfy a client request, or solve a problem" (p. 67). His discussion of expectations versus standards is a compelling argument for continued exploration and debate about the use of this approach.

## DEFICITS IN OUR LITERATURE BASE IN EVALUATION

One of the characteristics of a collection of articles from journal such as *GCQ*, or any other research-oriented journal, is that the manuscripts are reviewed using stringent criteria for traditional experimental design (although more recently, the qualitative paradigm has received increasing attention). Evaluations are not research studies with expectations of strictly controlled experimental and control groups, and they are not designed with expectations for generalizability to other settings or with the expectation of expanding the general knowledge

## xxviii Program Evaluation in Gifted Education

base. Because evaluations do not and should not/cannot be expected to adhere to these rigid criteria, they are not likely to be published. Further, most evaluations are carried out with expectations on the part of the program and school district that the information gathered has the purpose of informing the program only of the strengths, weaknesses, and accomplishments of a particular school district's efforts in meeting the needs of the gifted. Hence, the practical examples of good evaluation are limited in the research base. Only authors who are able to gain the permission of the districts with which they work to "expose" the findings are able to translate the particular evaluation report into a publishable manuscript. And only those authors who are able to frame the evaluation study as exemplars of evaluation or who have been fortunate enough to be able to create some semblance of a comparison group have been successful in bringing the knowledge learned from evaluation studies to the public domain.

### Examples of Lessons Learned

The limitations discussed above have deprived our field of important information that could be used to guide others in the formulation of plans for the oft-neglected, but critical, aspect of program planning and revision. Evaluators who are actively engaged in the important process of gathering data on program process and products should be giving consideration to using the lessons that they learn through the evaluation process to inform others who either execute evaluation themselves or seek to engage outside, independent evaluators. Particularly, there is a need to explicate ways in which the difficult question of program impact can be addressed by careful, creative, and systematic data collection. Informing others of the existence of useful assessment tools and providing data on the reliability and validity of new, innovative assessment tools as Reis and Renzulli (1991) have done.

### Longitudinal Evaluation

One of the most neglected aspects of the evaluation of gifted programs has been longitudinal or long-term impact assessment. The evaluation studies in this compendium (Avery, VanTassel-Baska, & O'Neill, 1997; Landrum, 2001; VanTassel-Baska, Willis, & Meyer, 1989) report on student outcomes that represent the impact across one- to two-year assessment periods. The literature fails to carefully track or provide a model for the evaluation of the impact of school programs across the span of the child's career in a gifted program—particularly as the program spans the elementary, middle, and high school years. Hertzog (2003) lamented the lack of evaluation data that provides program decision-makers with evidence of lasting or cumulative effects of programming strategies or curricular modifications. Like longitudinal research, longitudinal evaluation studies require unique planning and investments of resources—both time and energy—that go beyond the ordinary or evaluations that are in response to "crises" in the gifted program. Gifted program evaluation that is implemented "at the moment" is hampered in the assessment of long-term

effects by several factors. First, without anticipatory planning, student databases are not created to ensure that tracking of students is possible, that careful plans are designed and executed to ensure appropriate assessment intervals with careful selection or creation of outcome measures, that students who leave the program or school district are either monitored or even noted, etc.

Longitudinal evaluation, as all evaluation, requires that school personnel can answer the essential question, "How will these students be different—what will they know, understand, be able to do?" What dispositions will they have when they graduate from your high schools (or finish middle school, or leave elementary schools) than they would have been if the gifted program had not existed or they had not participated? Answers to these questions are foundational to determining whether a gifted program is effective in achieving its goals, yet often neither articulated nor evaluated.

### **Evaluating Programs for Special Populations of Gifted Learners**

With the increased attention to groups of learners who have typically been underrepresented in programs for the gifted, and the numerous projects targeting these students through the Javits Gifted and Talented Program, issues surrounding the identification of appropriate program and student goals as well as issues of appropriate techniques for evaluating these programs have emerged. Descriptions of these programs have been provided in such publications as *Contexts for Promise* (Callahan, Tomlinson, & Pizzat, n.d.). As House and Lapan (1994) note, evaluation of these programs presents the same issues as presented in evaluation of all gifted programs, but with an additional layer of issues that emanate from the particular problems in evaluation of any programs that focus on at-risk populations. These include issues of assessment surrounding disadvantaged and limited-English-speaking populations, selection of appropriate indicators (drop-out rates, percentage of students going to college, etc.), and the tenuous nature of gain scores as indicators of success.

### **Uniquely Applicable Models**

Despite the common sense and useful guidelines provided in these articles and in the other parallel articles, book chapters, and monographs, it is notable that all of these principles and models represent adaptations of existing general program evaluation models. The emergence of a new and unique model has not yet occurred. Whether such an evolution is necessary cannot be answered definitively, but we might ask whether such a model might be the stimulus for more widespread acceptance and use of evaluation as an integral part of program planning in gifted education.

### **Models That Integrate Qualitative and Quantitative Approaches**

It is also noticeable that the field has yet to present a clear and useable set of guidelines for the integration of quantitative and qualitative data collection

### xxx Program Evaluation in Gifted Education

and analysis. Barnette (1983) described naturalistic approaches to gifted and talented program evaluation and Lundsteen (1987) presents a research model with an ethnographic base as a potential model. Lundsteen has articulated the importance of qualitative approaches in helping understand the patterns and interactions within a program, which affect program efficiency and potency. Yet, the field has not built upon this work, nor has it provided careful guidance and illustrations in implementation of the qualitative approach, which can easily be misused, and subsequently, disregarded as too subjective if not correctly implemented.

Nor has the field generated a range of solutions to the issues that arise in determining whether program interventions have had an appreciable effect on student growth and development. Carter (1992) offered one model to address the problems in creating a true experimental approach to gifted program evaluation; however the models currently offered represent adaptations of general models rather than models specifically designed to address the many unique issues that have been raised about evaluating programs for the gifted.

## CONCLUSION

It appears that the work of evaluation is not yet finished and many challenges still remain to be addressed by researchers and evaluators. The paucity of application of the knowledge we have and the holes in the research base create the opportunity for experts to make valuable contributions to increased effectiveness and efficiency of programs for the gifted.

## REFERENCES

- Avery, L. D., VanTassel-Baska, J., & O'Neill, B. (1997). Making evaluation work: One school district's experience. *Gifted Child Quarterly*, 41(4), 124-132. [See Vol. 11, p. 61.]
- Baker, E. L., & Schacter, J. (1996). Expert benchmarks for student academic performance: The case for gifted children. *Gifted Child Quarterly*, 40(2), 61-65. [See Vol. 11, p. 109.]
- Barnette, J. J. (1983). Naturalistic approaches to gifted and talented program evaluation. *Journal for the Education of the Gifted*, 7(1), 26-37.
- Callahan, C. M. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly*, 30(1), 38-42. [See Vol. 11, p. 1.]
- Callahan, C. M., & Caldwell, M. S. (1995). *A practitioner's guide to evaluating programs for the gifted*. Washington, DC: National Association for Gifted Children.
- Callahan, C. M., & Caldwell, M. S. (1986). Defensible evaluations of programs for the gifted. In C. J. Maker (Ed.), *Critical issues in gifted education* (pp 277-296). Rockville, MD: Aspen.
- Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). Instruments and evaluation design used in gifted programs. (RM 95132). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.
- Callahan, C. M., Tomlinson, C. A., & Pizzat, P.M. (n.d.). *Context for promise: Practices and innovations in the identification of gifted students*. Charlottesville, VA: University of Virginia.



- Carter, K. (1992). A model for evaluating programs for the gifted under non-experimental conditions. *Journal for the Education of the Gifted*, 15(3), 266-283.
- Carter, K. R., & Hamilton, W. (1985). Formative evaluation of gifted programs: A process and model. *Gifted Child Quarterly*, 29(1), 5-11. [See Vol. 11, p. 13.]
- Gallagher, J. J. (1979). Issues in education for the gifted. In Passow, A. H. (Ed.), *The gifted and talented: Their education and development. The seventy-eighth yearbook of the National Society for the Study of Education* (pp. 28-44). Chicago: University of Chicago.
- Hertzog, N. B. (2003). Impact of gifted programs from the students' perspectives. *Gifted Child Quarterly*, 47(2), 131-143.
- House, E. R., & Lapan, S. (1994). Evaluation of programs for disadvantaged gifted students. *Journal for the Education of the Gifted*, 17(4), 441-446.
- Hunsaker, S. L., & Callahan, C. M. (1993). Evaluation of gifted programs: Current practices. *Journal for the Education of the Gifted*, 16(2), 190-200.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs*. New York: McGraw-Hill.
- Kulieke, M. J. (1986). The role of evaluation in inservice and staff development for educators of the gifted. *Gifted Child Quarterly*, 30(3), 140-144. [See Vol. 11, p. 29.]
- Landrum, M. S. (2001). An evaluation of the catalyst program: Consultation and collaboration in gifted education. *Gifted Child Quarterly*, 45(2), 139-151. [See Vol. 11, p. 77.]
- Landrum, M. S., Callahan, C. M., & Shaklee, B. D. (2001). *Aiming for excellence: Gifted program standards*. Waco, TX: Prufrock.
- Lundsteen, S. W. (1987). Qualitative assessment of gifted education. *Gifted Child Quarterly*, 31(1), 25-29. [See Vol. 11, p. 119.]
- Martinson, R. A. (1976). *A guide toward better teaching for the gifted*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Reis, S. M., & Renzulli, J. S. (1991). The assessment of creative products in programs for gifted and talented students. *Gifted Child Quarterly*, 35(3), 128-134. [See Vol. 11, p. 47.]
- Renzulli, J. S. (1975). *A guidebook for evaluating programs for the gifted and talented*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Renzulli, J.S., & Ward, V.S. (1969). *Diagnostic and evaluative scales for differential education of the gifted*. Unpublished manuscript. University of Virginia.
- VanTassel-Baska, J., Willis, G. B., & Meyer, D. (1989). Evaluation of a full-time self-contained class for gifted students. *Gifted Child Quarterly*, 33(1), 7-10. [See Vol. 11, p. 101.]
- Wiggins, G. (1996). Anchoring assessment with exemplars: Why students and teachers need models. *Gifted Child Quarterly*, 40(2), 66-69. [See Vol. 11, p. 39.]