Evaluation of Gifted Programs: Current Practices

Scott L. Hunsaker & Carolyn M. Callahan

In an effort to describe current gifted program evaluation practices, a review of articles, ERIC documents, and dissertations were supplemented by evaluation reports solicited by The National Research Center on the Gifted & Talented at The University of Virginia from public school, private school, and professional sources. Seventy evaluation reports were received. These were coded according to ten variables dealing with evaluation design, methodology, and usefulness. Frequencies and chi squares were computed for each variable.

A major concern brought out by this study is the paucity of evaluation reports/results made available to the NRC G/T. This may be due to a lack of gifted program evaluations, or to dissatisfaction with evaluation designs and results. Other concerns included lack of methodological sophistication, reporting, and utility concerns.

Some promising practices were apparent in the studies reviewed. A large sub-set of the evaluations were done for program improvement and employed multiple methodologies, sources, analysis techniques, and reporting formats with utility practices that produce needed changes. In addition, most evaluations focused on a number of key areas in the gifted program rather than settling for generalized impressions about the program.

Trends in the Evaluation of Gifted Programs

The call for gifted program evaluations is certainly not new. Gallagher, Weiss, Oglesby, and Thomas (1983) indicated that as early as 1960, when accountability and evaluation were identified as important components of educational programs, the call for gifted program evaluation was made. Since that time, efforts have been made to identify issues in gifted program evaluation (Renzulli, 1975; Calla-

han, 1983, 1986) and to provide suggestions for appropriate evaluation practices (Callahan & Caldwell, 1984; Callahan & Caldwell, 1986; Carter & Hamilton, 1985). Still recently, Gallagher (1988) included program evaluation among the priorities he identified as crucial for the continued improvement of gifted education.

Research Question

One aim of the National Research Center on the Gifted and Talented (NRC G/T) is to study the designs of gifted program evaluations in order to ascertain the effectiveness of various models and strategies. The purpose is to find designs, models, and strategies that provide the most accurate, timely, and useful information for policy development and program improvement. This research will lead to systematic efforts to assist schools and evaluators in implementing quality evaluation (i.e., evaluation that is useful). A first step in the process is to describe the current state of the art in gifted program evaluation. Accordingly, this paper will address current evaluation designs, methodologies, audiences, concerns, reports, and utility, as well as the relationships among these areas.

Methodology

Data Gathering

Actual program evaluation reports were sought from three sources. First, a search of educational data-bases was conducted to find reports available in the professional literature. Second, an appeal through journals, newsletters and conferences was made to professionals in evaluation and gifted education for copies of evaluation reports. Finally, a special mailing to state-level gifted coordinators and NRC G/T Collaborative School Districts (schools that indicated an interest in participating in NRC G/T research projects at the proposal stage) and a mass mailing to approximately 5000 local school districts was conducted. These efforts resulted in a few hundred responses relating to evaluation. However, many responses included only statements of policy concerning the requirements for evaluation or copies of evaluation instruments. Thus, the present study is based on only 70 evaluation reports, a list of these is available from the authors.

Scott Hunsaker is a staff member and Carolyn Callahan is an associate director and principal investigator for the National Research Center on the Gifted and Talented (NRC G/T). NRC G/T is funded by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. The opinions expressed in this paper are the authors' and not necessarily those of the NRC G/T or OERI.

Journal for the Education of the Gifted. Vol. 16, No. 2, 1993, pp. 190-200. Copyright © 1993. The Association for the Gifted, Reston, Virginia 22091.

Figure 1

Sample Evaluation Coding Variable & Categories

Variable	Categories
Data-Gathering Methodology	Test
	Questionnaire
	Interview
	Observation
	Document Analysis
	Meeting
Data Analysis Technique	Other
	Descriptive Statistics
	Inferential Statistics
	Other Quantitative
	Content Analysis
Reporting Format	Other Qualitative
	General Report
	Executive Summary/Abstract
	Oral Presentation
	Audiovisual Presentation
	Chart/Table/Graph
	Memorandum
	Press Release
	Other

Data Analysis

The evaluation reports were coded on ten variables by NRC G/T staff trained in: gifted program evaluation; evaluation type; evaluation model; evaluator type; data-gathering methodology; data analysis technique; data sources; intended audiences; reporting format; evaluation concerns; and utility information. Predetermined categories within each variable were based on a thorough review of evaluation and gifted education literature (see Figure 1 for a sample of the coding categories). The definition of evaluation terms essentially followed those given in Worthen and Sanders (1987). Frequencies of the categories within each variable were computed. In order to determine the independence of the variables, chi square analyses were conducted on data from all pairs of combinations for nine of the ten variables. The variable "evaluation concerns" was not included due

to difficulties with inflated N (that is, no meaningful categorization of the data was possible while still maintaining independent observations for each evaluation report). More sophisticated analyses were not conducted due to the non-parametric nature of the data. For a full report on the development of the rating scale and evaluation of its content validity and reliability, see Callahan, Lundberg, and Hunsaker (in press).

Limitations

The data reported here are related only to gifted program evaluation, and, due to non-random sampling, generalizations should be made cautiously.

Further, interpretation of chi square analyses should be conducted cautiously. With the large number (36) of analyses conducted, it is possible that a single analysis could exceed the critical chi square value at the .05 level due to chance alone. However, given the large number of analyses (15) that exceeded the critical values in this study, it is still legitimate to pay attention to these results.

Results*Frequencies*

The frequency results will only be summarized in this article. A more detailed description is available elsewhere (Hunsaker & Callahan, 1991). Essentially summative evaluation to determine program worth was used in the majority of evaluations, although about one-third employed formative evaluation designed to improve the program. A majority of the evaluations focused on the concerns of program administrators, with approximately one-quarter focusing on the goals and objectives of the program. Just over half of the evaluations were conducted by internal evaluators.

In terms of data generation and analysis, over three-quarters of the evaluations used questionnaires. However, most evaluations that used questionnaires combined this technique with other quantitative and qualitative techniques. Students, teachers, and parents were the primary sources of information, while administrators were used as sources in less than half of the evaluations. While simple descriptive techniques were used in nearly two-thirds of the evaluations, multiple data analysis techniques were used in just under half. Other

Table 1

**Frequencies of Evaluation Concerns in Gifted
Program Evaluation Reports**

Evaluation Concern	f	%
Curriculum/Instruction	37	52.9
Identification	31	44.3
Organization	31	44.3
General Impressions	30	42.9
Parent/Community Involvement	30	42.9

N = 70

Note. Mean number of evaluation concerns per report = 4.1. Number of reports addressing only one concern = 12. Frequencies add to more than 70 and percentages to more than 100 due to multiple categorizations.

quantitative and qualitative techniques were used with much less frequency.

The top five specific evaluation concerns are shown in Table 1. Only twelve of the reports focused on one concern. The mean number of concerns per report was 4.1. The most frequent evaluation concerns were curriculum and instruction (52.9%), identification (44.3%), organization (e.g., models, schedules) (44.3%), and parent/community involvement (42.9%). General impressions (e.g., My child is challenged. My child enjoys the program) of the program were a concern in 42.9% of the reports.

Nearly two-thirds of the evaluations were reported through general reports and data tables, while just over half the evaluation reports used multiple reporting formats. However, only about one-quarter of the reports provided an executive summary. Just under half the reports contained recommendations only; while just over one-quarter went beyond recommendations to produce time-lines for implementation, task definitions or policy and goal formulation. About one-third included no utility information at all.

Chi Square Analyses

In order to avoid expected frequencies of less than one, a number of categories within each variable were collapsed in the chi square analyses. Of the 36 chi square analyses conducted, 15 had signifi-

Table 2

**Chi Square Analysis of Data Sources
by Data-Gathering Methodologies**

Data Sources		Data-Gathering Methodology		
		Multiple Methods	Questionnaire	Other
Multiple	observed	37	14	2
	expected	32.6	15.1	5.3
Students	observed	5	2	4
	expected	6.8	3.1	1.1
Others	observed	1	4	1
	expected	3.7	1.7	0.6

N = 70

Degrees of freedom = 4

Critical value ($\alpha = .01$) = 13.277

$\chi^2 = 16.59$

cant results at or above the .05 level. Only analyses related to specific concerns raised in this paper are presented here.

In most cases, evaluations relied on multiple sources including students, parents, teachers, and administrators, however a few used only one source. The data gathering process, also, included multiple methods, in most cases (see Table 2).

When the intended audience for the evaluation was solely administrative, evaluations tended to be summative, though the formative type was also used. When the intended audiences were multiple, formative evaluations tended to be favored. Further, exclusively administrative audiences were more likely to receive results from management-centered evaluations (those focusing on evaluation concerns of program management) than from evaluations conducted with other models (see Table 3).

From Table 3, we also note the primary reporting format, regardless of intended audience, was multiple formats. However, administrative audiences tended to get more reports with tables only.

The methodologies associated with utility practices (those practices that result in evaluation findings having some impact on a program) are reported in Table 4. Essentially, evaluations employing multiple methods were likely to yield reports with utility information; however, a number of reports offered recommendations as a

Table 3

Chi Square Analysis of Intended Audience by:

Evaluation Types		Administrative	Research	Multiple
Summative	observed	24	13	1
	expected	24.4	8.1	5.4
Formative	observed	16	1	6
	expected	14.8	4.9	3.3
Other	observed	5	1	3
	expected	5.8	1.9	1.3

critical value ($\alpha=.01$) = 13.277
 $\chi^2 = 14.73$ (df4)

Evaluation Models		Administrative	Research	Multiple
Management-Centered	observed	33	1	3
	expected	23.8	7.9	5.3
Objectives-Centered	observed	8	6	4
	expected	11.6	3.9	2.6
Other	observed	4	8	3
	expected	9.6	3.2	2.1

critical value ($\alpha=.01$) = 13.277
 $\chi^2 = 24.45$ (df4)

Reporting Formats		Administrative	Research	Multiple
Multiple Formats	observed	22	11	5
	expected	24.4	8.1	5.4
Tables	observed	12	0	1
	expected	8.4	2.8	1.9
General Reports	observed	8	0	3
	expected	7.1	2.4	1.6
Other Formats	observed	3	4	1
	expected	5.1	1.7	1.1

critical value ($\alpha=.05$) = 12.592
 $\chi^2 = 13.80$ (df6)
 N = 70

Table 4

Chi Square Analysis of Utility Practices by Data-Gathering Methodologies

Utility Practice		Data-Gathering Methodology		
		Multiple Methods	Questionnaire	Other
Recommendations	observed	22	5	3
	expected	18.4	8.6	3.0
None	observed	6	13	2
	expected	12.9	6.0	2.1
Beyond Recommendations	observed	15	2	2
	expected	11.7	5.4	1.9

$\chi^2 = 70$
 Degrees of freedom = 4
 critical value ($\alpha = .01$) = 13.277
 $\chi^2 = 17.15$

minimum, and a number of the reports went beyond recommendations. Use of questionnaires as the sole data gathering methodology tended to result in reports with no utility information.

Discussion

These results point to a number of important trends that should be addressed. In this section, some concerns will be raised and then some promising practices will be discussed.

Concerns

Perhaps the major concern brought out by this study was the paucity of evaluation reports/results made available to the NRC G/T. This is likely due to the lack of gifted program evaluations, or to dissatisfaction with evaluation designs and results. These explanations are considered likely in light of the high number of responses received at the NRC G/T with regard to evaluation policies and instruments.

Another concern is that evaluations that are carried on tend to be summative evaluations addressed to administrators dealing with

concerns raised by administrators, with information often gained through questionnaires as a sole methodology. Further, information from these evaluations is often disseminated in the form of simple data tables, with little focus on program outcomes. Such evaluations tend not to be used as information for policy development or program improvement. Seeley (1986) aptly described such evaluation as "academic gymnastics" (p. 268).

Promising Practices

A number of promising practices were noted in the evaluation reports reviewed. First of all, a large sub-set of the evaluations analyzed employed a formative type of evaluation. Their expressed intent was program improvement. Further, most evaluations employed multiple data-gathering methodologies from multiple data sources; many used multiple data analysis techniques; and a number reported results through multiple formats. This is important given the apparent association of the use of multiple methodologies, sources, analysis techniques, and reporting formats with utility practices that produce policy development and program improvement.

For example, in an evaluation of a total district gifted program in Minnesota, information about the program was gathered through questionnaires, observations, and focus group discussions. Administrators, teachers, students, and parents of both identified and non-identified children were used as data sources. Data were analyzed using frequencies, content analysis, and professional standards. The results were reported through memoranda, summaries, a lengthy general report, and at meetings with teachers and administrators. The evaluators produced numerous recommendations in specific areas. These recommendations were used as the basis for the work of committees established to reformulate program goals and practices.

A second promising practice, one in accord with the recommendations of Callahan (1986) and Carter and Hamilton (1985), was that most evaluations focused on a number of key areas in the gifted program rather than settling for generalized impressions about the program. While the evaluation of key program components tended to be subjective in nature, important programming issues were dealt with across multiple audiences.

For example, the evaluation of a junior high gifted program in Pennsylvania dealt with six broad areas of concern (e.g., development of IEPs, provision of information to parents). Within each of these areas a number of specific issues were delineated, such as the review of

IEPs by counselors or notifying parents of newly-identified children, yielding twenty-six issues for the evaluation. These were then narrowed to seventeen focused evaluation questions to be studied. These questions were generated through an extensive process involving administrators, teachers, and parents in discussions and ratings of key features of the program.

A final promising practice, that of making evaluation information useful, appears to be taking root. Most evaluations at least generated recommendations. Recommendations often took the form of specific suggestions for program improvement, but also included the identification of program components needing further study. Many went beyond recommendations to formulate committees, goals, action plans, and policies.

For example, one site in Illinois formulated a committee to prepare a response to the external evaluation of their identification procedure. For each finding and recommendation reported by the evaluator, the committee determined the extent to which they agreed or disagreed with the evaluator, defended their level of agreement, and established a plan of action to implement those recommendations with which they at least partially agreed.

Conclusion

Evaluation is an important component of any gifted program because, when appropriately applied, it provides a systematic means to improve program performance. While there still appears to be a need for continued advocacy for evaluation to be done at all, a small core of programs has had success in implementing worthwhile evaluation practices.

Future NRC G/T evaluation studies will focus on identifying the evaluation methods and strategies that foster the attitude of evaluation utilization for program improvement so that specific guidance can be provided to gifted program evaluators. The efforts of professional evaluators and local school districts described in this paper will go a long way in identifying those methods and strategies and in developing support for evaluation as an important component of gifted programs.

References

- Callahan, C. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly*, 27, 3-7.
- Callahan, C. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly*, 30, 38-42.
- Callahan, C., & Caldwell, M. (1984). Using evaluation results to improve programs for the gifted and talented. *Journal for the Education of the Gifted*, 7, 60-74.
- Callahan, C., & Caldwell, M. (1986). Defensible evaluations of programs for the gifted and talented. In C. Maker (Ed.), *Critical issues in gifted education: Defensible programs for the gifted* (pp. 27-296). Rockville, MD: Aspen.
- Callahan, C., Lundburg, A., Hunsaker, S. (in press). The development of the scale for the evaluation of gifted identification instruments. *Gifted Child Quarterly*.
- Carter, K., & Hamilton, W. (1985). Formative evaluation of gifted programs: A process and model. *Gifted Child Quarterly*, 29, 5-11.
- Gallagher, J. (1988). National agenda for educating gifted students: Statement of priorities. *Exceptional Children*, 55, 107-114.
- Gallagher, J., Weiss, P., Oglesby, K., & Thomas, T. (1983). *The status of gifted/talented education: United States survey of needs, practices, and policies*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Hunsaker, S., & Callahan, C. (1991, November). *Trends in the evaluation of gifted programs*. Paper presented at the meeting of the American Evaluation Association, Chicago.
- Renzulli, J. (1975). *Working draft: A guidebook for evaluating programs for the gifted and talented*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Seeley, K. (1986). Evaluation for defensible programs for the gifted. In C. Maker (Ed.), *Critical issues in gifted education: Defensible programs for the gifted* (pp. 265-276). Rockville, MD: Aspen.
- Worthen, B. & Sanders, J. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.

Establishment of a National Data Bank on Identification and Evaluation Instruments

Carolyn M. Callahan & Michael S. Caldwell

Overview

The establishment of a National Repository for Instruments and Strategies used in the Identification and Evaluation of Gifted Programs is one of the projects of the University of Virginia site of the National Research Center on the Gifted and Talented. The goal of these activities has been to collect and evaluate identification and evaluation instruments, systems and designs. The major research questions posed for the identification aspect of the study include: What are the most commonly used instruments in identifying gifted and talented students? What instruments are used for identifying gifted and talented students according to some specific definition or conception of giftedness? What evidence is there of the reliability and validity of these instruments, and is that evidence sufficient to justify their use with given definitions of giftedness, underserved populations, and other population characteristics?

A similar set of questions was posed in the area of evaluation instruments and designs. What are the most commonly used instruments in evaluating gifted students and programs? What evidence is there of the reliability and validity of these instruments in assessing goals and objectives common to gifted programs? What instruments (especially non-traditional and product-oriented instruments) are used in the evaluation of programs for the gifted and talented?

Carolyn M. Callahan is Professor of Education at the University of Virginia and Associate Director of the National Research Center on the Gifted and Talented. Michael S. Caldwell, Associate Professor of Education in the Curry School of Education at the University of Virginia, is also on the staff of the National Research Center on the Gifted and Talented. This paper was supported by the Office of Educational Research and Improvement, U.S. Department of Education. However, the opinions expressed are the authors' own.

The work reported herein was supported under the Javits Act Program (Grant No. 206R0001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Journal for the Education of the Gifted. Vol. 16, No. 2, 1993, pp. 201-219. Copyright © 1993, The Association for the Gifted, Reston, Virginia 22091.