DOCUMENT RESUME

ED 275 123 EC 190 905

| | |
|---|---|
| TITLE | Quantitative vs. Qualitative Approaches to Quality Special Education Program Evaluation. |
| INSTITUTION | Council of Administrators of Special Education, Inc. |
| PUB DATE | 85 |
| NOTE | 61p. |
| AVAILABLE FROM | CASE Research Committee, c/o L. C. Burrello, Indiana University, School of Education, Third and Jordan Streets, Bloomington, IN 47405 ($7.00 to CASE members; $14.00 to non-members). |
| PUB TYPE | Reports - Descriptive (141) -- Information Analyses (070) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Administrator Attitudes; *Disabilities; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; *Program Evaluation; *Qualitative Research; *Research Design; Research Methodology; *Special Education; Special Education Teachers; Statistical Analysis; Teacher Attitudes |

ABSTRACT

One in a series of issue papers commissioned by the Council of Administrators of Special Education (CASE), this document presents a comparison of contemporary evaluation approaches for special education programs. The first section describes the two approaches to be compared: (1) traditional scientific inquiry which emphasizes quantitative methods; and (2) the naturalistic paradigm which emphasizes qualitative methods. The second section, written by Lester J. Horvath, is titled "A Quantitative Program Evaluation Approach to Evaluating Quality Special Education Programs," and covers such areas as: the context of the problem (administrative challenge, reasons for program evaluation, contextual issues in special education, and training); assumptions regarding special education program evaluation; the recommended approach (including description, purpose and specific stages of evaluation); and other considerations. Section Three, written by Edith E. Beatty and titled "Qualitative Responsive Approaches to Evaluating Special Education Program Quality," addresses the local context; qualitative responsive evaluation; frameworks; and implementing qualitative responsive evaluations. The fourth section, written by Kenneth R. Olsen, is a comparative analysis of the approaches presented by Horvath and Beatty and discusses areas of agreement between the two approaches (stakeholder involvement, criteria for effective evaluations, emphasis on quality) and areas in which the methods differ (issues and concerns versus decisions, knowing design in advance, need for consensus). The final section outlines administrative implications of the findings and lists questions to help special education administrators in choosing the appropriate method for evaluating the quality of their special education program. (CB)

# Quantitative vs. Qualitative

## Approaches to Quality Special Education Program Evaluation



CASE

# Quantitative vs. Qualitative

## Approaches to Quality Special Education Program Evaluation

Summer, 1985

## For Additional Copies, Contact:

CASE Research Committee
c/o L. C. Burrello
Indiana University
School of Education
Third and Jordan Streets
Bloomington, IN 47405

Price: $7.00 per copy to CASE members
$14.00 per copy to non-members

Prices include postage and handling.

## For CASE Membership Information:

David E. Greenburg
Executive Director, CASE
Indiana University, ES 3108
902 West New York Street
Indianapolis, IN 46223

# Acknowledgments

*On behalf of the CASE Executive Committee I want to thank the CASE Research Committee and the authors of this Commissioned Paper Series for their participation in the production of such a high quality document, Mr. Horvat, Ms. Beatty and Mr. Olsen, CASE is indebted to each of you.*

*My doctoral assistant, Kay Harmless, reworked the Committee's recommendations into the drafts and Leonard Burrello completed the final editing of this publication. The Committee has also launched its second publication entitled "Evolving Organizational Structures" due this Fall.*

David E. Greenburg
Executive Director
CASE, Inc.

## CASE Research Committee

# Table of Contents

**Section One: CASE and Program Evaluation** ..... 1

Description of Approaches ..... 1

Product Overview ..... 2

**Section Two: A Quantitative Program Evaluation Approach to Evaluating Quality Special Education Programs** ..... 3

**Introduction** ..... 3

The Purpose of This Article ..... 4

**Context of the Problem** ..... 4

The Challenge for Administrators ..... 4

Reasons for Program Evaluation ..... 5

Contextual Issues in Special Education ..... 7

Training ..... 9

Terminology ..... 10

State of the Art ..... 10

**Assumptions Regarding Special Education Program Evaluation** ..... 11

General Assumptions ..... 11

Administrative Assumptions ..... 12

**The Recommended Approach** ..... 12

Description and Purpose ..... 12

Specific Stages in the Evaluation ..... 13

**Other Considerations** ..... 17

Validity Issues ..... 17

Reducing Burden on Staff and Intrusion on Programs ..... 18

Evaluation Cost ..... 19

Frequency of Evaluation ..... 19

Ensuring Evaluation Use and Follow-up ..... 19

A Partnership with Parents and Staff ..... 20

Evaluation Standards ..... 20

**Summary** ..... 20

**References** ..... 21

**About the Author** ..... 22

**Section Three: Qualitative Responsive Approaches to Evaluating Special Education Program Quality** ..... 23

**Introduction** ..... 23

**The Local Context** ..... 25
What are the issues ..... 25
The Problem ..... 26

**Qualitative Responsive Evaluation** ..... 27

**Frameworks** ..... 27
Values ..... 30
Relevance ..... 32

**Implementing Qualitative Responsive Evaluations** ..... 33
How Do You Do It? ..... 33

**Summary** ..... 34

**References** ..... 35

**About the Author** ..... 36

**Section Four: A Comparison and Analysis of The Two Approaches to Local Special Education Program Evaluation** ..... 37

**Introduction** ..... 37

**Overview** ..... 38

**Areas of Agreement** ..... 40
Continuum #1 Stakeholder Involvement ..... 40
Continuum #2 Meeting Criteria for "Good" Evaluations ..... 40
Continuum #3 Emphasis on Quality ..... 41
Continuum #4 Numbers and Words ..... 42

**Areas in Which They Differ** ..... 43
Continuum #5 Issues and Concerns vs. Decisions ..... 43
Continuum #6 Knowing Design in Advance ..... 43
Continuum #7 Need for Consensus ..... 44

**Continua About Which There is Insufficient Information in the Papers** ..... 45
Continuum #8 Internal vs. External Evaluators ..... 45
Continuum #9 Using Existing Data ..... 46
Continuum #10 Processes or Outcomes? ..... 47

**Summary** ..... 47

**References** ..... 49

**About the Author** ..... 49

**Section Five: Administ**
**Introduction** .....
**Self Study Guide** . . .
**Implications** ..... 51
..... 51
..... 52

# Section One: CASE and Program Evaluation

In response to the public's call for quality and excellence in education, special educators have come to realize that the previous emphasis on compliance monitoring is necessary but not a sufficient condition in program improvement efforts designed to achieve excellence. CASE has identified evaluation models and practices as a high priority for local special education administrators from the results of our research, conferences, and discussions with local leaders. The papers on program evaluation, teacher, and related staff evaluation have been distributed through the CASE Information and Dissemination service program. A third document highlighting evaluation practices currently being implemented around the United States was published jointly by CASE and the Office of Special Education Programs in the Department of Education.

As part of its continuing commitment to provide information on the most recent developments in quality program evaluation, the CASE Executive Committee charged its Research and Special Projects Committee to continue in this focal area of interest with a document that compares contemporary program evaluation approaches. Contrasting approaches have been a primary issue of discussion at conferences held by CASE over the past two years on the subject of program evaluation.

## Description of Approaches

One approach is based upon traditional scientific inquiry and emphasizes quantitative methods. Horvath defines quantitative program evaluation as a planned attempt to measure the extent to which certain expectations about the program are being met. The second approach is based on a naturalistic paradigm and emphasizes qualitative methods. Beatty defines qualitative program evaluation as an emergent process designed to discover the relevant values and expectations held by constituences who are served by the program. These two approaches can be reviewed along a number of continua. One continuum is that the quantitative or scientific inquiry approach requires a consensus on specific evaluation objectives prior to the conduct of the evaluation itself, while the naturalistic approach uses qualitative methods to discover the values or outcomes held by multiple stakeholders (persons with a stake in the special education program, including the full evaluation committee) in the educational setting. On another continuum, the methods and data collection instruments used in a quantitative approach are specified and linked to the specific outcomes during the evaluation activity. The methods used in a qualitative approach evolve throughout the process of evaluation. Both approaches, however, are credible, useful, feasible, and proper means to conduct quality program evaluation. Given these distinctions, the choice of approach need not be "either/or," but dependent on the leader's perception of the purpose, values, and complexity of the program(s) to be evaluated within the local context. We collected

two top special educators'views of these two processes to illustrate their respective advantages.

## Product Overview

The two authors selected are identified with these respective approaches. They are Lester J. Horvath representing quantitative and Edith E. Beatty representing qualitative. They were asked to draft their papers based upon five organizing questions:

1. What in your view is the current context of special education and the need to consider program evaluation?
2. What assumptions do you make about special education?
3. How do your assumptions affect the conduct of program evaluation?
4. What purpose do you hope to achieve in conducting an evaluation?
5. How do you recommend a program evaluation be designed and carried out?

These two papers are placed in Sections II and III of this document.

When the papers were completed, they were sent out for CASE Research Committee and outside consultant reviews. The drafters, consultants, and Research Committee members read and discussed the papers on September 19, 1984, to determine similarities and differences in approaches and potential uses of these methodologies in local school program quality evaluation efforts. Two consultants were invited to critique the papers: Ken Olsen of the Mid-South Regional Resource Center in Lexington, Kentucky, and Constance Berquist of Evaluation Systems Design, Incorporated, Tallahasee, Florida. Ken Olsen was also asked to lead the discussion of the comparative strengths and weaknesses of each approach. His comments were of such quality, the Committee asked him to capture his suggestions along with suggestions of our Committee. These comments are placed in Section 3 of this report.

The Committee also developed a set of administrative implications that it hoped would assist colleaques in determining what approaches best fit the local context. These statements comprise the fourth and final section of this publication.

We hope this document in the series, will give you a more complete picture of evaluation. This up-close look should help those crafting their own designs and implementation of school improvement efforts.

L.C. Burrello, Research Committee Chair

# Section Two:

## A Quantitative Program Evaluation Approach to Evaluating Quality Special Education Programs

*by Lester J. Horvath*
*Associates in Professional Technologies, Inc.*
*West Hartford, Connecticut*

*The first paper is written on the quantitative Approach. This scientific inquiry approach requires a consensus on specific evaluation outcomes. The goals for all students should be considered as the starting point. The methods and data collection instruments used in a quantitative approach are specified and linked to the specific outcomes during the evaluation activity. In support of this approach, some special education administrators said:*

- *Our district wants data on our program operation. They believe collected facts and figures are the best measure of a program's success.*
- *We selected a quantitative model because it was simpler to administer and could be used right away.*
- *This method required less time and was not as expensive as other models.*
- *Questions were developed from the state of Massachusetts twelve goals for education and were adapted to our unique situation. The wording was altered to make the reading easier for respondents and an observation checklist was made for moderately handicapped students who could not complete a survey.*
- *The quantitative method of evaluation was efficient and low in burden.*

### Introduction

From an administrative viewpoint, good program evaluation must provide practical information which can lead to program improvement and inform the administrator and other decision-makers about the program's progress. In order to do this, the program's processes (procedures and implementation) as well as the program's outcomes (products) must be studied within the context and realities of the specific situation. In special education this means a study of the quality of pro-

cesses such as instruction, student assessment, IEP development, and outcomes such as student learning in basic skills, appropriate student behaviors, and parental satisfaction with the program within the context of local expectations about the program.

### The Purpose of This Article

This paper advocates a specific approach to special edcation program evaluation for special education administrators, and specific strategies for employing the recommended approach. The approach is essentially quantitative, combining several of the most freqently employed design elements as they apply to special education. A major feature of the approach is that various qualitative methodologies can be added to complement the basic model. These alternative methodologies are described in the companion article within this issues paper.

The specific approach advocated in this paper relies on planning the evaluation activities early in the special education program evaluation process, based on the consensus of a committee of program stakeholders. The special education administrator includes parents, professionals, and representatives of all concerned groups in the formation of the stakeholder committee. This committee provides directions for and control of the evaluation throughout the process. The stakeholder committee provides structure through the identification of evaluation questions which define the scope of items on evaluation instruments. The instruments yield quantitative data, such as descriptive statistics for interpretation and the development of a management plan.

## Context of the Problem

### The Challenge For Administrators

Administrators of special education programs are responsible for the management and improvement of a very important component of school district activities. This importance comes from the size of the program, its complexity, and the expectations of the people with a stake in the program.

The size of special education is considerable. In the United States alone, special education serves 4.4 million students at an annual cost of well over ten billion dollars. A special education administrator in a medium size local education agency (LEA) has approximately a one million dollar budget; the scope of large cities and intermediate units is of course, considerably higher. With respect to small LEAs, special education often requires disproportionate administrative attention.

Perhaps no other education program matches special education for complexity of administration. Approximately 50 pages of federal regulations, pertinent state mandates, and a body of expanding case law constitute the legal backdrop. Compounding this procedural complexity is the necessity to provide for an extremely diverse population of students. There is an abundance of state and federal reports to be submitted regarding the program. Professional practices change rapidly, and individual student progress is somewhat difficult to measure. Most special education students spend a substantial proportion of their day in regular education programs, necessitating coordination.

The expectations of parents, students, staff, the community, and the funding agents are varied, but there is a common theme. A summary of that theme is that special education students must:

1. Be provided an education in accordance with the various mandates.
2. Be provided educational experiences in accordance with sound professional practices (i.e., "good services").
3. Benefit from their educational experience in areas such as the basic skills, vocational skills, and social skills (citizenship, physical well-being, and other student outcomes are often included as expectations).

The challenge for administrators, then, is to squarely address the expectations of the various stakeholders through program evaluation, and make constructive use of the evaluation results.

## Reasons for Program Evaluation

From the perspective of the special education administrator, the expectations of various stakeholders (including the administrator) give rise to four reasons for program evaluation:

1. Program Improvement. The evaluation is used as a management tool for improving specific components of the LEA special education program.
2. Policy Analysis. The evaluation provides information to policy makers regarding programmatic issues that require resolution at a policy level.
3. Accountability. The evaluation serves as an accountability report to an administrative, regulatory, oversight, or funding authority.
4. Public Relations. The evaluation is used to provide information which is useful in a public relations or public information effort.

In practice, special education administrators often conduct program evaluation for a combination of the four above reasons. Frequently there are primary and secondary reasons for evaluation. In some cases, additional use is made of evaluation results post hoc. For example, the principal reason for conducting evaluation might be the use of the report as a management tool to identify specific program components (e.g., elementary LD resource rooms) that need attention with respect to specific program expectations (e.g., mathematics achievement). At a later date, the administrator might make use of the same data for policy analysis. The design could compare outcomes across resource rooms that employ different approaches to coordination with the regular education curriculum.

If the evaluation were comprehensive, it could be used for an accountability reporter to the board of education in response to a request for information about the program's effectiveness. Finally, the administrator could include the report as part of the ongoing public relations effort for the program by disseminating the findings to various media. If the findings are positive, they would generate overall support for the program. Negative findings can be used to support requests for additional funding and/or politically difficult programmatic changes.

The following describes each of the four reasons for special education program evaluation with more specificity.

**Program Improvement.** The use of program evaluation by administrators to systematically identify program areas that need administrative attention is an important aspect of administration itself. Administrators continously do ths kind of evaluation informally and without benefit of systematic methodology and "rigor".

Special education is an ongoing program of considerable magnitude, in need of systematic administrative attention. Recent evaluation literature has largely ignored the managerial perspective. This may be because, in the past, the bulk of evaluation was conducted on grant programs that could be terminated or continued on the basis of the evaluation results. The reason for program evaluation in the competitive context of grant programs was a continuation or elimination of the program. Program improvement was a secondary issue at best.

For special education the termination of the program or even the termination of subcomponents is a moot issue because the program and contiuum of services is mandated. Thus, program improvement, change, and restructuring are the focus of the administrator's decision making.

It should be noted that this has important implications when planning a unit of analysis (sometimes called reporting unit) in the evaluation. From the administrative perspective it is important to ask for one's information in useable "chunks". The specific example here would be to structure the evaluation unit of analysis so that it corresponds with administrative planning units such as the LD elementary resource rooms. A poorly planned evaluation might report on too small a unit of analysis for the administrator (e.g., individual students) or too large a unit (e.g., combining the elementary and the high school levels).

**Policy Analysis.** In the past, most policy analysis was done at the federal and state level through grants and contracts to private evaluation firms, universities, and other research groups. As part of the decentralization of educational policy-making, policy analysis is increasingly being conducted at the local school district level. California's program evaluation system is the leading example of this trend within the field of special education. The arguments for allocating resources for policy analysis at the LEA level center around: 1) a better understanding of the contextual issues by LEA professionals; 2) improved technical expertise in many LEAs; and 3) an increased likelihood of use of the findings by other LEAs. Policy analysis evaluation has many similarities to applied research. As program evaluation is the child of research, this is the apple that fell the shortest distance from the tree. Specifically, policy analysis is often asked to "investigate" a proposed treatment, compare the treatment against a control, or compare several treatments.

Some policy analysis problems center around lack of clarification of the question, rather than a need to compare proposed solutions. In such situations, the most valuable section of the policy analysis report is the section that identifies questions for further study. Qualitative methodologies are especially suitable for the identification of issues that lead to development of new hypotheses, due to the rich description of the program and the insightful interpretations provided.

**Accountability.** This reason for program evaluation in special education stems from the need to provide comprehensive and impartial reports about the

program to the major formalized groups of stakeholders. These groups are school boards, local parent groups, the school superintendent, the administrative council (in some large districts), and the state department of education. (From the SEA perspective, these groups would include the state board of education, the state legislature and the U.S. Department of Education.) When an administrator is asked to report on program quality and effectiveness as an accountability mechanism from one or more of these groups, the response should be related to the expectations of the group, as discussed above. In the past, administrators have claimed that evaluation data on student progress is not possible. This response has been considered inadequate by many school boards and parent groups.

When accountability is a significant purpose for conducting program evaluation, a self-evaluation approach directed by the special education administrator stands the likely risk of criticism on the grounds of lack of impartiality. Although this criticism can be countered to some degree if the evaluation uses materials and procedures developed from the outside (i.e., from the SEA or a private vendor), the administrator should be prepared for such criticism. Impartiality and avoidance of subjectivity are critical elements when conducting evaluation for the purpose of accountability.

**Public Relations.** Special education administrators are very familiar with the need for public relations because such efforts have historically been necessary for program survival. Even today, after many years with mandates in place, some districts must defend their programs or face severe cuts that would force noncompliance with the mandates. Evaluation findings can complement the administrator's ongoing public relations efforts.

The professional ethics of the special education administrator and the evaluator(s) are tested when evaluation results are used for public relations purposes. It is not enough to be passive with respect to the issue of ethics and evaluation; the administrator should periodically remind all persons involved with the evaluation that an honest, ethical process is to be followed from conception to dissemination.

If the evaluation is to be a self-evaluation by the LEA, the credibility issues noted above apply here and interact with the ethical issues. A self-evaluation used for public relations purposes amidst a climate of distrust could "backfire", even if the evaluation and reporting were completely honest and even handed.

It has been said that all evaluation has a political public relations effect, even when such effect is not aspired to (Cronbach, 1983). A practical guide for the administrator in preventing a public relations catastrophe is to keep the process as open as possible, and involve representative stakeholders throughout the process on an evaluation committee. It is no coincidence that these very stakeholders are also the core of political support for special education programs. Since the administrator cannot avoid the complicated areas of public relations and politics, it is suggested that a proactive, participatory approach be employed in this aspect of program administration.

## Contextual Issues in Special Education

Evaluators from almost all schools of thought agree that the context (the setting of the program in the broadest sense) is a critical factor in the design and execution of good program evaluation. Special education poses some paradoxical

twists for the uninitiated program evaluator who is not familiar with this particular program. In some ways special education is very much like all of education, but in certain ways the program almost defies understanding from a program evaluation point of view.

**Similarities to Regular Education.** Special education is like all of education when it comes to the basic outcome expectations of the stakeholders. As mentioned above, persons with an interest in the program expect students to profit in basic skills and the other areas commonly listed as goals for all education. With respect to process evaluation, an equivalent theme can also be developed for special education – sound professional practices should be employed in the delivery of services.

When developing a quantitative evaluation system, the goals for all students should be considered as the starting point. There is no need to spend dozens of professional days coming to the same conclusions that the local school board has formulated over the years. It is most revealing to note the similarities between special education and regular education when going through this process. This writer has sat through many committee meetings charged with developing "indicators of quality" (a preordinate evaluation step equivalent to goal setting in goal-based evaluation) only to have the belated discovery that the committee had essentially restated the same goals as the local board of education. Thus, it is recommended that the evaluation committee begin with the regular education goals. Each of these should be reviewed to determine their relevance, with the premise that special education students are entitled to regular education programs and services in accordance with their unique needs. Not only is this more efficient, it emphasizes the philosophical point that special education is not a separate entity from regular education. Special education usually has some additional program goals and objectives, the bulk of which are process oriented.

**Contrasts with Regular Education.** Some contextual issues are very different from regular education. Special education has many activities that are quasi-evaluation in nature or whose terminology is confusing to the evaluator. For example, the individualized education program process has an annual review for each student; this is program evaluation with the student as the unit of analysis. There is, of course, student assessment and evaluation, and comprehensive student evaluation every three years as required by section 300a.534 of the federal regulations. Some states employ a program evaluation mechanism for those program components funded by Part B flow-through funds. Almost all grants awarded through the twenty percent Part B state share funds require a special evaluation. Public Law 89-313 programs have an evaluation requirement. Most school districts employ some form of teacher evaluation which covers special education staff. Finally, there is the process of compliance monitoring.

Compliance monitoring by the state education agency is a form of program evaluation that is limited to a portion of the process issues that can and should be addressed. Basic compliance monitoring addresses the implementation of mandated processes, but it does not address the quality of such implementation. For example, compliance monitoring checks to see that each student's IEP contains objectives. It does not check the degree to which the objectives, taken as a whole, define a good program for the student for the coming year.

Testing of students in special education is an important contextual issue.

Program evaluators who do not understand the purposes for and the limitations of student assessment in special education often attempt quantitative evaluation using student testing as the dependent variable. This has not proved fruitful for several reasons. The variance within the special education population is so great that it masks accurate measurement. The tests themselves are often not validated for application to this heterogeneous group, as contrasted to the assumptions made for Chapter I program evaluation (e.g., using normal curve equivalence scores).

Special education student prevalence varies widely from LEA to LEA. In some cases learning disabilities are identified at a two percent prevalence rate, while some communities identify well over ten percent of their students as learning disabled. This variability in prevalence rates prohibits comparisons and traditional pre-post designs that require objective criteria for program success (e.g., nine months again in mathematics per year). The districts with a higher prevalence rate would tend to show greater testing gains with their LD populations because the population would, as a whole, have less severe problems.

## Training

It was noted earlier that there is considerable turmoil in the program evaluation literature regarding methodology. There are conflicting ideas regarding the fundamentals of research and evaluation design. Although it appears that the debate will not be readily resolved, practicing program evaluators can and do benefit from the exchange of ideas. Strengths and weaknesses of alternative designs can be considered in terms of actual applications for special education program evaluation. Of course, special education administrators do not have the time to be fully immersed in current design issues. Although administrators have shown considerable recent interest in program evaluation, it is not reasonable to expect a full-time administrator to have detailed knowledge of the evaluation profession. (Only the legal profession is entitled to that much of an administrator's time.)

With the developing interest in program evaluation and the recent changes in the field of evaluation, special education administrators will need some sort of training. Even those special education administrators not planning to conduct the studies themselves need to keep abreast of current concepts in research and evaluation. Most special education administration programs do not require a course devoted to program evaluation. One can assume that an administrator who took the standard fare of related courses (e.g., research design, tests and measurements, statistics, experimental psychology) before the mid-1970's was exposed to what was then the state of the art – traditional, quantitative experimental design. During the intervening period, quantitative designs have improved and evolved, and qualitative designs have emerged. The "state of the art" has changed.

As one should know the reason for evaluation before designing the evaluation, one should know the reason for training before selecting inservice. In a workshop of a few days or less, an administrator can learn the overall concepts and become an intelligent consumer of evaluation services. It is not possible to learn how to evaluate one's program after a few days of seminar. There have been numerous attempts at turning administrators into evaluators with workshops over the last ten years; the lack of administrators conducting evaluation is testimony to the failure of that approach.

Prepared materials and ongoing support seem to be the two key ingredients in support of successful training for selfevaluation of programs by administrators. The SEA can provide considerable assistance in this area.

## Terminology

The special education administrator will notice that program evaluation terminology is misleading in some areas. Many terms are value laden. The terms quantitative and qualitative, for example, both describe methods that attempt to measure quality. The distinction rationalisitic/naturalistic is often interchanged with quantitative/qualitative. To appreciate the subtle distinctions in terminology that are at the center of this vigorous professional debate, the reader needs to consider the philosophical assumptions behind the various approaches.

Evaluation professionals have been vigorously contrasting and arguing the merits of methodologies on philosophical and practical levels. There is a considerable debate as to whether the two approaches can be combined in a given study, due to their philosophical contrasts. A review of literature for the last ten years shows increasing interest in qualitative methodologies in research and evaluation.

There is a discrepancy between the professional literature and current practice in terms of methodology use. Miles and Huberman (1984) state that few working researchers are not blending the two approaches. In practice, the vast majority of program evaluation (across program specialties) is conducted using quantitative designs combined with some qualitative methods. In general, practicing program evaluators are not purist regarding the methodologies they employ; rather, they tailor their methods as they approach each situation. This is done to capitalize on the positive aspects of the various specific methodologies.

It has been suggested (Lynch, 1983) that the terms quantitative and qualitative describe types of data rather than methods of evaluation, and that methods fall on a continuum with naturalistic inquiry at one end and experimental inquiry at the other. This type of framework would allow for the definition of a type of evaluation that is naturalistic (i.e., the evaluator has not manipulated the program/treatments) and quantitative (i.e., the evaluator collects numeric data rather than verbal descriptions). Using these definitions, a great deal of present evaluation is naturalisitic inquiry with quantitative data.

Until the field of evaluation agrees on a common terminology, it is suggested that the special education administrator focus on the basics of the evaluation process rather than the labels for approaches.

## State of the Art

**Extended Monitoring Systems.** Many LEAs have experienced, either as a self-evaluation or as an evaluation conducted by the SEA, an activity that extends compliance monitoring beyond mandated issues and into additional areas. These efforts are not comprehensive evaluation, but they have filled a need during a period of special education history when LEAs struggled with making the new special education mandates work.

Such systems typically mix quantitative and qualitative methodologies. For example, the Michigan SEA monitoring model (Michigan State Board of Education, 1981) includes items that inquire as to whether or not special education man-

dates have improved the education of handicapped children. A negative answer from some proportion of respondents would not prove noncompliance. This is a policy evaluation question, not a question on compliance determination. The Michigan manual also includes questions related to process evaluation that go beyond compliance determination. Many states have two compliance systems, basic and extended. In New Jersey there is a basic monitoring system and a Level II system, the latter being an extended monitoring system that formally recognizes the distinction between basic compliance monitoring and their extended monitoring systems.

A review of several of these extended monitoring systems reveals significant problems. The focus of the evaluation tends to be compliance issues and the quality of the compliance process. This is a different focus from an evaluation that measures the quality of the educational process. Extended compliance systems generally do not address outcomes. Such systems also tend to be extremely labor intensive. The system could also cause problems in the relationships between an SEA and LEAs if the distinction between mandates and subjective issues becomes blurred. It should be noted that only a few SEAs have the resources to assure consistent, quality evaluations using an extended monitoring system.

**Two Statewide Efforts.** Approximately ten years ago California and Massachusetts required that their LEAs conduct program evaluation. The early efforts in these two states met with limited success, even though workshops were held to assist LEAs. Despite the workshops, LEA adminstrators did not know how to actually conduct evaluation. To compound the difficulties, compliance issues were often confused with the purpose of program evaluation, and the evaluation report was often viewed as an additional paperwork exercise associated with the LEA Plan. When both SEAs provided a manual for program evaluation along with ongoing support, useful program evaluation becomes a reality for a large number of LEAs (Massachusetts Department of Education MDE, 1981; California State Department of Education, 1982). At this point in time, it can generally be said that LEAs in these two states have conducted successful evaluation through more than one cycle.

# Assumptions Regarding Special Education Program Evaluation

## General Assumptions

The following comprise the writer's general assumptions about special education program evaluation:

1. Special education program evaluation is possible.
2. Initial efforts at program evaluation should be modest in terms of work and complexity.
3. It is better to do a modest program evaluation then to do no program evaluation at all.
4. The focus of the evaluation will become the focus for the program itself (this is called "goal tropism").

5. A specific evaluation design should depend on the reason for the evaluation, the specific local context, and who will conduct the evaluation.
6. Evaluation should address both process and outcomes.
7. Evaluation should be an honest and open process.
8. The people closest to the program are excellent sources of information about the program's effectiveness.

### Administrative Assumptions

The following assumptions focus on the administrative perspective:

1. Program evaluation must not be confused with teacher or staff evaluation, and this must be made clear at the beginning of the evaluation.
2. Most special education administrators do not have the time to manage a program evaluation unless prepared materials/methods and consultative support are provided.
3. The evaluation design and reporting format should not be so complex or esoteric as to "lose" the intended audience.
4. It is critical to involve the stakeholders from the beginning of the evaluation, through a vehicle such as an evaluation task force.
5. The evaluation should burden the staff and other respondents as little as possible; wherever possible, existing data should be used.
6. The evaluation process should be a positive experience for the special education department staff; no special education staff member should be left out of the process.
7. It is important to determine the involvement of regular educators in the evaluation process; in most LEAs all principals and some regular education teachers should be involved in any comprehensive evaluation.
8. Ownership of the evaluation by parents and staff will assist greatly in effecting improvements recommended in the report.
9. Follow-up of a program evaluation requires at least as much work as development of the report.

## The Recommended Approach

### Description and Purpose

The approach recommended in this section stems from the contextual issues and the assumptions described above. It has evolved from reviewing numerous special education evaluation works over recent years from the perspective of evaluator and evaluation materials developer.

**Overview of Stages.** The approach consists of eight stages, which organize the process and allow for the incorporation of various theoretical models at critical points in the process. The eight stages follow:

1. Defining the LEA Context

2. Developing the Design and Instrumentation
3. Collecting the Evaluation Information
4. Analyzing the Data
5. Interpreting the Data
6. Writing and Presenting the Evaluation Report
7. Applying the Management Plan
8. Conducting Follow-up Evaluation

**Theoretical Base.** As stated earlier, practicing evaluators generally use a blend of methods in any given application. This makes it especially difficult to describe an applied model using theoretical labels such as "Discrepancy Evaluation" or "Goal-Free Evaluation." Rather than beg the issue, however , a description of the methodological and theoretical bases follows.

Instead of structuring the evaluation around a specific evaluation methodology, set of program goals, or instrumentation, it is suggested that the process of evaluation itself be the structure. This allows for the incorporation of individual methods that best suit the needs of the individual LEA at each point. A comparable structure, which can accommodate various methods, was employed by the Center for the Study of Evaluation in designing their Program Evaluation Kit (Morris and Fitz-Gibbon, 1978). This concept of evaluation-based organization is similar to Stufflebeam's (1983) Context Input Process Product (CIPP) model, because it emphasizes the administrative perspective and focuses on the decisions to be made. Gable (1982) described a special education program evaluation process through a series of stages based partially on an adaptation of the CIPP decisions.

The recommended approach should be categorized as in the quantitative school because the evaluation design is planned in the early phases of the evaluation, and the focus of information collection is numerical data from a variety of instruments. Regarding the quantitative/qualitative definitions, it should be noted that some of the instruments (e.g., interviews, case study review sheets) used in a typical application of the recommended approach are often considered qualitative. The main distinction in defining the character of the recommended approach is its deliberateness with respect to direction. That is, the evaluation is purposeful in terms of investigating the extent to which certain expectations about the program are being met. It is recommended that these expectations and the direction be set by a broad-based committee representing the various stakeholders during the first phase, as part of defining the LEA evaluation context.

## Specific Stages in the Evaluation

The following paragraphs describe the eight stages of the evaluation process.

**Defining the LEA Context.** This first phase of the evaluation involves discussing the purpose of the evaluation, the resources, program structure, and other aspects of context evaluation (Stufflebeam, 1983) in a broad-based committee format. Special education administrators are by necessity experienced in sharing authority with consumers and staff through the team process. In this setting the open process achieves what House (1980) termed "democratizing evalua-

tion." Using this type of open process to establish evaluation questions allows the best of two models, stakeholder and goal-based evaluation, as advocated within the context of special education program evaluation by Ando (1984).

A working subcommittee needs to be assigned for specific tasks during this phase. After the subcommittee is assigned, it can begin to do the ground work for the evaluation. Program objectives regarding both process and product (outcomes) should be drafted in consideration of the programmatic context of the LEA. Subcommittee members should interview a broad spectrum of stakeholders in the formulation of these program objectives. It is suggested that the subcommittee conduct these interviews using a combination of structured questions and open ended questions. The structured questions should be derived from existing LEA documents that state educational goals for regular education and for special education. For example, the board of education may have passed a resolution stating that all students age 16 and above should be provided an opportunity for vocational skill training. The interview guide in this example would then pose the question, "To what extent do you feel that availability of vocational skill programming is an important program objective for special education in our schools?" Note that this example is a process evaluation objective in that it addresses the delivery of a program rather than a student outcome. If the board resolution had been that students should acquire vocational skills, the objective would be considered a product (outcome) objective. An important discussion among subcommittee members as they develop an interview guide is to determine whether or not an issue is process, product, or both. In this example, the subcommittee could decide to expand the issue to address product as well as process. The implementation of this specific issue is also a point to be determined. It should be stressed that the subcommittee has the latitude to apply common sense and the implicit concerns of the stakeholders in the development of the interview guide and in the interpretation of the results. This should not be viewed as a lock step process, beyond the control of the evaluators. A "pure" and traditional goal based evaluation does not allow for flexibility at this point in the process. The recommended approach in this article stresses the concerns of the stakeholders and is responsive to the context of the program.

After the subcommittee summarizes the results of the interviews, the summary should be presented to the evaluation committee for revision and adoption. A product of this stage of the process is a list of program process and program product evaluation questions as described by Olsen (1979) and Gable (1982), or as discussed in the Massachusetts system (MDE, 1981).

**Developing the Design and Instrumentation.** The next phase of the evaluation consists of the actual design. Put simply, the first order of work here consists of determining how each evaluation question should be answered. The evaluator's judgment is called for when determining how many instruments (sources of information) are sufficient to address each question. Specific steps in this phase include the need to:

1. Determine proper evaluation approaches for each of the program objectives, and reach committee consensus on the suitability of each approach.

2. Develop the specific instrumentation, assignments, and instructions for each of the evaluation approaches.

3. Develop design mechanics, such as the unit of analysis and the sampling plan.

4. Pilot test the instrumentation and make necessary revisions.

5. Proofread and print sufficient quantities of the instrumentation.

**Collecting the Evaluation Information.** Although staff and parents may have been told about the evaluation in a general way when the process was initiated, it is a good idea to notify them at this point and explain that this is a program evaluation (not staff evaluation) and that their cooperation with surveys, interviews, etc. will be greatly appreciated. Some LEAs do this notification through an article in their regular newsletter to the community, while others send a special mailing.

The recommended approach described in this article often involves interviews and folder reviews. Interviewers need to be trained on the interview protocol, and case folder reviewers need to be trained on actual case material to increase reliability and agree on a format for any qualitative data recording.

The specific instruments need to be distributed according to the sampling plan, and a system for following up return of completed instruments should be established. An important point in collecting information is to obtain the maximum rate of return for each instrument. Issues of access and confidentiality of data need to be addressed.

Although the recommended approach is primarily quantitative, a good deal of anecdotal information, planned and incidental, is usually collected during the process. In some cases this information is quite useful, but in other cases it is unrelated to the purpose for evaluation. The person(s) in charge of such data have three responsibilities: logging the data according to some logical schema; and, presenting periodic updates on the information to the evaluation committee to see if areas for additional systematic data collection are warranted.

The final steps in collecting the data are also important. Nonrespondents should again be asked to provide the information to increase rates of return. All files and materials must be returned to their exact place of origin. People who conducted interviews, reviewed cases, etc. should be promptly "debriefed" so that their impressions can be obtained on common themes, incidental findings, and any identified limitations regarding the data that was collected.

**Analyzing the Data.** The data analysis techniques to be employed vary from evaluation to evaluation, but some common points can be put forth as suggested practice. In general, however, descriptive statistics should comprise the bulk of the analysis. Terms such as "percentage satisfactory" are especially useful to the reader of the report. The "percent" is one of the most widely understood concepts.

If the purpose of the evaluation was to statistically compare groups, and the instrumentation and sampling were designed in a manner to support such analysis, the proper difference test would be performed at this stage. In general, such traditional quantitative analysis is not helpful to the evaluation. The excep-

tions are usually larger policy studies, where generalization of principles across LEAs or settings is part of the evaluation purpose. Undue emphasis on statistics is usually somewhat harmful for program improvement studies, because the focus too easily becomes the detail of the numbers rather than how to improve the program. The debate about practical significance vs. statistical significance applies to this situation. For example, increasing the sample size affects statistical analyses in a manner which may not translate to an increase in practical significance.

Computers have been increasingly helpful in taking the burden out of data analysis. The common mainframe statistical packages such as Statistical Analysis System (SAS) and Statistical Package for the Social Sciences (SPSS) provide quick and dependable descriptive and inferential statistics. Evaluation using the recommended approach has been conducted using a microcomputer for all data analysis. It is suggested that a small sample of computations be conducted by hand, to ensure that the computer is doing the intended calculations.

**Interpreting the Data.** The committee as a whole should be involved in data interpretation. Discussion of the results at this stage yields a list of major findings. When important evaluation questions remain unresolved due to conflicting data, additional information should be collected as deemed necessary by the committee. The overall focus of data interpetation should be the process and product findings as they pertain to the purpose of the evaluation.

**Writing and Presenting the Evaluation Report.** Writing the report is an extention of the interpretation activity described above. The focus of the committee at this stage turns from identifying the findings to developing recommendations and plans for action.

The reader of the report should be provided an overview of the logical flow of evaluation processes from the first phase of the evaluation (defining the LEA context) through the development of recommendations and management plan. Clarity of purpose and understanding the perspectives of the audiences are important guidelines in writing. Graphic displays, executive summaries, and a full final report are the essential products.

An important part of the final report is the management plan. Traditionally, evaluation reports have concluded with a recommendations section. It is suggested that this be taken a step further, with a plan to implement the recommendations. A limited number of important findings should be developed into plans with practical approaches, timelines, persons responsible, and resouces needed. A well-stated management plan can assist in securing resources and making politically difficult changes.

**Applying the Management Plan.** Even though the most pervasive debate in evaluation concerns methodology, the biggest failure of evaluation has been administrative follow up for program improvement. Recent examples, however, have demonstrated success in applying program evaluation.

A well-written management plan can serve as a roadmap for the special education administrator. Placing an emphasis on the major points in department meetings, inservice, and written communications can assist making positive changes. The stakeholder group representatives (e.g., principals and parents) who were involved in the evaluation can help considerably in this stage.

**Conducting Follow-Up Evaluation.** This stage of the recommended approach is part of the application of the management plan, but goes a step beyond

in that it is the beginning of the next cycle. There are several advantages to using some of the same evaluation questions (and instrument items) in repeated evaluations. For some of the basic program expectations such as student skill acquisition, LEA baseline data have been established. The success of the points in the management plan can be determined, with new approaches and/or additional resources allocated to those areas that continue to need attention. The burden of instrument development itself is also reduced.

The focus of each successive program evaluation should be evolutionary, keeping important program goals and stakeholder concerns from past evaluations, dropping issues no longer important, and adding new program priorities. Between evaluations, the LEA should devote energy to the follow up of management plans. Thus, program evaluation is a dynamic and integral part of special education administration.

## Other Considerations

The following important points deserve the special education administrator's attention when planning for program evaluation.

### Validity Issues

In general, the source of validity for each evaluation question stems from the fact that multiple sources of information are employed to address the question.

This definition of validity is sometimes difficult to understand by persons who have a background in traditional tests and measurements. Such persons have been trained that extensive validity and reliability statistical measures must be obtained on all instrumentation before use. This is not applicable for evaluation instruments that have been individually designed for application in a specific context. Rather than attempt a series of statistical measures, a far more meaningful approach to validity is achieved by comparing information from multiple sources within the evaluation. For example, if the student records checklist agrees with teacher surveys and parent interviews regarding mathematics achievement by elementary students served in the resource rooms, a high degree of confidence can be placed in the findings.

### Reducing Burden on Staff and Intrusion on Programs

Too often, an otherwise sound activity is not worth pursing because of staff burden and intrusion on operating programs. Evaluation can definitely fall into this trap. The evaluator and the evaluation committee have a great deal of work to do, and that work must be done with a high degree of rigor. The respondents such as teachers, administrators, parents, and related services staff in many LEAs have been deluged with surveys and meetings for a host of projects over recent years. To the extent that past experiences with such interruptions have taken place, indifference or even resentment on the part of the respondent may occur. Thus, evaluation activities have the burden of proof regarding their value.

The first steps in conducting evaluation are extremely important. It is assumed that outside assistance in design and instrumentation will be obtained. If possi-

ble, a evaluation expert familiar with special education and the selected model should be consulted regarding technical aspects. Because of the substantial time commitment required for the evaluation, the special education administrator should plan to set aside approximately five months. A similar commitment of time for the cadre of staff and parents comprising the evaluation committee should also be secured.

Self-evaluation without the help of prepared materials is far more burdensome than with the use of prepared materials or contracting with an external evaluator. One reason for the pilot test of instruments is to determine the length of time for respondents and to consider this factor when revising instrumentation. It is far easier to create burdensome instruments than good, clean instruments which quickly obtain the necessary information. Part of the reason that so many respondents resent surveys is past experience with surveys that appeared to be poorly designed first drafts. The special education administrator should reflect on his or her attitude and experiences regarding surveys (and requests for information and interviews) during the instrument development and sampling plan stages of the project.

## Evaluation Costs.

Closely related to burden, the issue of cost is very much a factor in selecting a specific evaluation strategy. In the case of outside evaluators, the cost alone becomes the deciding factor for some LEAs. Self-evaluation can be conducted for considerably less, although equivalent costs are essentially paid "out of hide" by the special education administrator and staff.

When program evaluation first became standard practice with federal programs, a common standard was to set aside five percent of grant [illegible] for program evaluation purposes. That standard does not apply to ongoing special education programs because such programs do not depend on positive [illegible] results for their annual continuation. The aggregate cost of a five percent [illegible] would be staggering. Although special education program evaluation is sometimes conducted partly for reasons of accountability, it is more often principally used as a management tool by the special education administrator. Using this reason as the focus, cost effective evaluation can be defined as evaluation that improves the program, in a quality sense, more than the expenditure of the same funds would have produced through direct service. For example, would a $12,000 program evaluation every three years improve the program more than $4,000 worth of additional staff for each of those three years? (For clarity, this example ignores interest and inflation.) This general rule is useful in measuring the relative cost-effectiveness of any evaluation.

## Frequency of Evaluation

There is an obvious connection between the above two sections dealing with burden and cost, and the issue of frequency of evaluation. For these reasons, many LEAs decide on comprehensive evaluation on a three- to five-year cycle. During the years between evaluations, these LEAs devote administrative attention to following though with the program improvements specified in their evaluation report's management plan.

Some LEAs choose to evaluate only a portion of their program each year. This approach in effect addresses a cluster of units of analysis and provides relatively easier differentiation of approaches and evaluation questions among program components. To return to the assumption that some evaluation is better than no evaluation, evaluation of program components may be the best decision for some LEAs.

## Ensuring Evaluation Use and Follow-up

It cannot be emphasized strongly enough that evaluation should not be conducted if the report will merely sit on the shelf. If the program evaluation produces a management report section with specific responsibilities for various staff members and reasonable dates by which certain actions must be taken, the use of the evaluation is far more certain.

A sense of ownership of the evaluation report by staff, parents, and other stakeholders is a contributing factor to evaluation use. If the special education administrator dominates the evaluation at the early stages, broad-based ownership is not likely to occur. It is suggested that the administrator foster a healthy diversity of opinion within the evaluation committee, and yield in deference to persons closer to some of the issues during the committee's work.

## A Partnership with Parents and Staff

A natual continuation of the discussion of ownership is the development of a common purpose among the stakeholders in a program. As an anecdotal exampale, the parents in one LEA were very much included in the management plan as a follow-up to the program evaluation. Some of the management plan actions dealt with funding speakers for the local parent group. Other actions of particular interest to parents received prompt attention by the school superintendent. From a long-term perspective, an improved partnership between parents and the school is probably even more important than the specific management plan actions.

Special education has made great advances in structuring parent-school partnerships within various school activities. It is natural that such partnerships will be built into the program evaluation process as well. Special education administrators know well that this partnership is more than an end in itself. It is a cornerstone of the legitimate political process of program improvement.

## Evaluation Standards

Accepted standards, such as those put forth by The Joint Committee on Standards for Educational Evaluation (1981) should be followed in the implementation of the evaluation. These standards can be viewed to ensure that the evaluation is useful, feasible, proper, and accurate.

# Summary

Special education program evaluation is conducted for the purposes of program improvement, policy analysis, accountability, and public relations. Good program evaluation addresses the program's context, the expectations of people with a stake in the program, the program's processes, and products or outcomes of the program.

A practical approach to evaluation can be based on quantitative methodology and structured around the phases of the evaluation process. Specific design elements and instrumentation should be based on contextual issues, the purpose of the evaluation, and stakeholder expectations of the program. Special education administrators can determine the relative cost-effectiveness of conducting evaluation, comparing the advantages of various levels of outside assistance. When an efficient process is employed, evaluation becomes an integral part of administration.

Successful evaluation requires involvement and ownership in the evaluation process by parents and staff. This involvement begins with structuring the evaluation, and follows through to implementing the management plan.

# References


Anderson, S.B., & Ball, S. (1978). *The Profession and Practice of Program Evaluation*. San Francisco: Jossey-Bass.

Ando, L. (1984). *Advocating a Combination of Stakeholder and Goal-Based Evaluation Models. Opinions Bulletin Board, (January 27, 1984), SpecialNet, GTE Telenet*. Washington, D.C.: NSMI.

Bogdan, R.C. & Biklen, S.K. (1982). *Qualitative Research for Education: An Introduction to Theory and Methods*. Boston: Allyn & Bacon.

Brinkerhoff, R.O., Brethower, D.M., Hluchyj, T., & Nowakowski, J.R. (1983). *Program Evaluation: A Practitioner's Guide for Trainers & Educators* (Design Manual). Boston: Kluwer-Nijhoff.

Brinkerhoff, R.O., Brethower, D.M., Hluchyj, T., & Nowakowski, J.R. (1983). *Program Evaluation: A Practioner's Guide for Trainers & Educators* (Sourcebook/Casebook). Boston: Kluwer-Nijhoff.

California State Department of Education. (1982). *Guide to Special Education Program Evaluation*. Sacramento.

Cronbach, L.J. & Associates. (1980). *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.

Cronbach, L.J., (1983). *Ninety-five Theses for Reforming Program Evaluation*. In Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (Eds.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 405-412). Boston: Kluwer-Nijhoff.

Fitz-Gibbon, C.T. & Morris, L.L. (1978). *How to Design a Program Evaluation*. Beverly Hills: Sage.

Gable, R.K., (1982). A Model for Evaluating Special Education Programs: An Application to Determining the Management and Instructional Benefits of Microcomputers. *Journal of Special Education Technology*. 5(2), 35-43.

Guba, E.G., & Lincoln, Y.S. (1981). *Effective Evaluation*. San Francisco: Jossey-Bass.

House, E.R. (1980). *Evaluating With Validity*. Beverly Hills: Sage.

Lynch, K.B., (1983). Qualitative and Quantitative Evaluation: Two Terms in Search of a Meaning. *Educational Evaluation and Policy Analysis*. 5(4), 461-464.

Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (Eds.). (1983). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff.

Massachusetts Department of Education. (1981). *Special Education Program Evaluation: A Management Tool*. Boston.

Massachusetts Department of Education. (1981). *Special Education Program Evaluation: A Management Tool (Training Manual)*. Boston.

Michigan State Board of Education. (1981). *Procedures & Standards for Monitoring Special Education Programs in the State of Michigan*. Lansing.

Miles, M.B., & Huberman, M.A. (1984). Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft. *Educational Researcher, 113*(5), 20-30.

Morris, L. & Fitz-Gibbon, C. (1978). *Evaluator's Handbook*. Beverly Hills: Sage.

New Jersey State Department of Education. (1980). *A Guide for Evaluation of Special Education Programs and Related Pupil Personnel Services*. Trenton.

Olsen, K.R., (1979). Evaluating Effectiveness of Special Education Programs. *Mid-South Bulletin No. 10.* Lexington, KY: University of Kentucky.
Patton, M.E. (1978). *Utilization – Focused Evaluation.* Beverly Hills: Sage.
Patton, M.Q. (1980). *Qualitative Evaluation Methods.* Beverly Hills: Sage.
Patton, M.W. (1982). *Practical Evaluation.* Beverly Hills: Sage.
Renzulli, J.S. (1975). *A Guidebook for Evaluating Programs for the Gifted and Talented.* Los Angeles: National/State Leadership Training Institute.
Rossi, P.H., Freeman, H.E., & Wright, S.R. (1979). *Evaluation: A Systematic Approach.* Beverly Hills: Sage.
Sechrest, L. (Ed.). (1980). *New Directions for Program Evaluation: Training Program Evaluators.* (No. 8). San Francisco: Jossey-Bass.
Spradley, J.P. (1979). *The Ethnographic Interview.* Holt, Rinehart and Winston.
St. Pierre, R.G. (Ed). (1983). *New Directions for Program Evaluation: Management and Program Evaluation.* (No. 18). San Francisco: Jossey-Bass.
Stufflebeam, D.L., & Webster, W.J. (1980). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis. 2*(3), 5-20.
Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M. (1971). *Educational Evaluation and Decision Making.* Itasca, IL: Peacock.
Stufflebeam, D.L. (1983). *The CIPP Model for Program Evaluation.* In Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (Eds.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation.* (pp. 117-141) Boston; Kluwer-Nijhoff.
The Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects, and Materials.* New York: McGraw-Hill.

## About the Author

Lester J. Horvath is a special education program evaluator based in West Hartford, Connecticut. In 1980 he founded Associates in Professional Technologies, Inc., which developed the statewide evaluation systems for Massachusetts and North Carolina. He has also been involved in individual special education evaluation projects with LEAs, graduate follow up studies, and the evaluation of a severe/profound educational program. He has served as a consultant for the Connecticut State Department of Education for seven years.
Prior to his involvement with program evaluation, he was a special education teacher in Ohio and Connecticut, and coauthored a college textbook on the case study approach to learning and behavior problems. He received a Ph.D. in special education administration from the University of Connecticut.

# Section Three:

# Qualitative Responsive Approaches to Evaluating Special Education Program Quality

*by Edith E. Beatty*
*Northeast Regional Resource Center*
*Burlington, Vermont*

*The second paper is written on the qualitative approach. This approach uses natualistic inquiry to discover the values and expectations held by relevant audiences. The methods used in a qualitative approach evolve throughout the process of evaluation. The framework which structures this approach that considers human and political factors are important. Some comments from administrators throughout the country about the qualitative approach are:*

- *We were looking for an evaluation process that allowed us to capture the uniqueness of our community and the people in it. The qualitative method of evaluation provided everyone – directors, teachers, and parents – with an easy way of conceptualizing our program.*
- *This method was selected because it could cover the whole spectrum of services for handicapped students.*
- *The process is ongoing and involves the total staff and community in evaluation.*
- *It was worth it to us to take more time and use this approach since it was a valuable catalyst for participation in our district.*

## Introduction

The story goes that three umpires disagreed about the task of calling balls and strikes. The first one said, "I calls them as they is." The second one said, "I calls them as I sees them." The third and most clever umpire said, "They ain't nothin'till I calls them" (Simons, 1976; p. 29, in Welck, 1979; p. 1). The umpire who correctly asserts "They ain't nothin'till I calls them" rather neatly fingers a key element in (evaluating): the important role that people play in creating the environments that impose on them. (Evaluations), despite their apparent preoc-

> cupation with facts, numbers, objectivity, concreteness, and accountability, are in fact saturated with subjectivity, abstraction, guesses, making do, invention, and arbitrariness . . . just like the rest of us. Much of what troubles (evaluations) is of their own making . . . (Weick, 1979; p. 5).

A local advisory committee for special education decides to make program evaluation the number one priority for the coming school year. Never was a decision more easily reached. No one could remember a time when such an important agenda item breezed by so comfortably and smoothly. Consensus? Everyone in agreement? Probably not. Let's look at what the committee members were celebrating in their minds on their way home that night.

- This will fit in beautifully with our "Excellence in Schools" effort for next year.
- Our staff will learn some good evaluation skills; this will fit right into our Comprehensive System of Personnel Development (CSPD).
- We need some impact data for our school boards and legislators, preferably related to cost-effectiveness; some justification for all the money expended and appropriated.
- This provides a forum to showcase our schools; let everyone see that ours are the best in the state.
- Finally, a chance to collect some hard data as to which teachers are really teaching our kids.
- Let's see if we've really done what we said we were going to do.
- This activity will bring this committee closer together.
- This will improve the special education programs in the district.
- The timeliness of this activity is perfect! My daughter is graduating soon and the high school still hasn't dealt with her transition. It'll show them that secondary programs have to think beyond graduation.
- This evaluation will show the effectiveness of our three to five year old population initiative; we need to prove that we cannot wait any longer to provide programs from birth to five.
- This is a nice logical next step to compliance monitoring. Now we can give the state department, and they, in turn, Washington, some data which really show what we're doing for handicapped kids.
- Great! We'll provide a picture of all the handicapped kids underserved or yet to be served!

These members, like members of many existing committees and task forces at the local, intermediate, state, regional, and national levels throughout the country with similar charges, view program evaluation differently, see multiple purposes and uses for such an effort, and assume diverse directions and approaches to take in order to evaluate the effectiveness of special education programs. Their "celebrations" might be seen as a collection of issues and concerns which are on the minds of stakeholding audiences in or about or affected by the program (Stake, 1975) and must not be ignored in the process of the evaluation. They may even be used

as the "advance organizers" or guideposts for the evaluation. The chairperson of this committee has a tough and challenging job ahead. Planning evaluations requires a great deal of negotiation. Individuals will be tempted to quickly shop for good models, tried and true instruments, and reputable evaluators before considering what they actually want from the effort and what they will do with the data they will gather.

Good evaluation requires that the approach fits the context. In order to achieve such a fit, the evaluation must be contextually relevant and responsive, and driven by the stakeholders.

> A main point for the practitioner is that evaluators may encounter considerable difficulties if their perceptions of the study being undertaken differ from those of their clients and audiences. Typically, clients want a politically advantageous study performed, while the evaluators want to conduct questions-oriented studies, since these allow the evaluator to exploit the methodology in which they were trained. Moreover, the audiences usually want values-oriented studies that will help them determine the relative merits of competing educational goods and services. If evaluators are ignorant of the likely conflict in purpose, the evaluation is probably doomed to failure from the start . . . it is imperative to remember that no one type of study consistently is the best in evaluating education.
>
> It is virtually impossible to assess the true worth of any object. Such an achievement would require omniscience, infallibility, and a singularly unquestioned value base (Stufflebeam and Webster, 1980; p. 17 and 18).

Special educators need to continue to develop and implement more fitting approaches to evaluation efforts and recognize that there are many questions to consider prior to selecting a model, a set of instruments, or an evaluator.

In this paper, the author will describe some responsive evaluation approaches which go beyond the scientific paradigm, ones that are driven by organizers other than goals, and ones which rely more on qualitative methods than on quantitative ones. The issues special educators are currently facing in program evaluation, are discussed and explored. Some newer, more fitting approaches, are presented along with a set of questions and cues for how responsive evaluation might fit the local context and provide for effective evaluation of special education programs.

# The Local Context

## What are the issues?

As program evaluation has become a nationally high priority, many agencies, projects, and professional associations have made it their number one area for program assistance. A national task force, consisting of representatives from the Office of Special Education Programs (OSEP), OSEP supported projects, the six Regional Resource Centers (RRCs), the National Association of State Directors of Special Education (NASDSE), the Council of Administrators of Special Education (CASE), and consumer groups at large, was formed to identify issues and plan ways of using available information to address these issues concerning program development and evaluation.

Staff from CASE and the RRC network have identified a host of issues within the context of program evaluation. The following section discusses some of the issues most relevant to local and intermediate units.

**Roles and Relationships.** Federal, state, and local levels have historically maintained different roles and responsibilities. The push for compliance is strong from the top down, while cost effectiveness and the school improvement priorities have their roots in local autonomy. Potential collaborative efforts have become confused by differences in agency orientation.

**Compliance Monitoring and Quality Evaluation.** Until recently, special education program evaluation has focused on compliance monitoring; federal agency assumptions tell us that an excellent program is a compliant one. At the local level educators recognize continuing responsibilities for compliance, but feel the drive to move beyond thinking only about minimal standards. Special education has been using the same technology to measure compliance and effectiveness, seeing the two along one continuum. It would do well to recognize the differences. It is unlikely that the same evaluation approach will fit both processes.

**Regular and Special Education Partnerships.** Similar considerations apply to the regular and special education relationships in defining quality and effectiveness. What is a "good" program? Are "effective" programs for nonhandicapped students "appropriate" for handicapped students? Many local and state agencies are attempting collaborative efforts to evaluate and improve effectiveness, but again, are finding the need to identify the human and political factors in doing so. Some communities have combined special education program evaluation and school approval/school improvement programs; others have decidedly separated the two.

**Logistical Concerns.** In organizing for such an ambitious task as program evaluation, local developers are quick to point to some logistical issues and concerns regarding the identification and appropriation of resources. Gathering fiscal, human, political, and technical support for evaluation (e.g., negotiating design decisions, training staff to do evaluation and/or identifying outside "experts", using results, paying for evaluations) is a potentially discouraging piece of the process.

If one pays attention to the context of the program to be evaluated and the stakeholding audiences in and around the programs, these four issues and concerns can be addressed.

## The Problem.

With increased attention to program quality evaluation, there is a high level of interest in moving away from rigid and closed-ended methodologies and toward more open and descriptive approaches. "We need more than numbers" and "Traditional evaluations have not provided us with the data we really need" are examples of statements from administrators interested in expanding the current technology. People seem to like the notion of qualitative approaches, but are resistant, perhaps, because they believe that the methodology lacks structure or technology, that it is not for the "real world" and is not the status quo, or that it is more expensive and tends to yield "soft data." Some current models discuss audience identification and qualitative methods, but fall short of implementing the underlying assumptions and embracing the design process of responsive evaluation. Somewhere along the line in evaluation, scientific inquiry was the predominant choice and that evalua-

tion was to focus on how well goals and objectives were met. The methodology additionally would need to be valid, reliable, objective, and perhaps even generalizable. Significance of the findings would depend on percentages worth noting, co isensus, and a singular truth in terms of what is "good" or meritorious.

The world isn't that easy to evaluate. The author's experience at the local level with handicapped kids, their teachers, and administrators indicates that life just "ain't like that"; human beings don't necessarily behave scientifically, predictably, or as we have controlled them. Things simply aren't as rational as they seem and what we need is evaluation findings that are easily gathered, credible, and useful. It's imperative then to look at the very assumptions, paradigms, approaches, and methodology most fitting to address evaluation questions.

The current practice by practice could be expanded by exploring the present and growing state of the art in educational evaluation. In fact, given how far special education has come with regard to program development, ironically, professionals are still employing evaluation methods developed in the early seventies. Educating handicapped children in public schools is far more sophisticated today than it was in the early seventies; so is educational evaluation.

The author pulls together some of the current thinking in educational evaluation and shows that there are paradigms within which to describe programs other than the scientific one and there are advance organizers in addition to goals and objectives. It's important to recognize our culture as having pluralistic values, not consensual ones, that quality is in the eyes of the beholder, and that worth must be a partner to merit. Relevance is at least as important as rigor and one ought to consider the tradeoffs of each prior to assuming the need for one over the other.

## Qualitative Responsive Evaluation

**What is it?** As stated earlier, qualitative responsive evaluation meets two criteria: it fits the context and is driven by the stakeholders'purposes and uses. The author has resisted using a title for this approach because it is not a model, but rather a way of thinking about evaluation. Many existing models do fit this criteria and an eclectic posture can be assumed as long as the fittingness of assumptions and criteria are in place. In describing what qualitative responsive evaluation is, three themes are discussed: the *frameworks* which structure the approach, *values* considerations in the process, and *relevance* or criteria for judging evaluations.

### Frameworks

There are many design issues evaluation developers need to consider in order to achieve the right fit of approach to evaluation effort; all to often many important considerations are left undesigned and consequently the evaluation is less likely to respond to identified purposes and uses. There is neither time in local evaluators'lives nor room in this paper to be theoretically detailed or complete. However, three of what the author believes to be critical pieces in designing program evaluation will be highlighted.

**Paradigms.** One of the very first steps of good design is to explore the basic assumptions and postures individuals make about "getting at truth," or the "paradigm" in which the evaluation will be housed. As will be seen later,

methodologies and even to some extent, advance organizers, can be mixed; paradigms cannot.

Nearly all of special education evaluation has been designed and conducted within the scientific or conventional paradigm. Efforts to evaluate student ability, performance, or IEP's are usually norm-referenced, objectives-based, discrepancy-based, or some combination. Program compliance is usually monitored by some form of the *Discrepancy Evaluation Model* (DEM) (Provus, 1971), measuring discrepancies between legal mandates and program status. Many of the early and currently operational program evaluation models have been designed using some variation of DEM or Stufflebeam, et al's (1971) *Context, Input, Process, Product* (CIPP) model, a decisions-based approach. Evaluations within the scientific paradigm assume a singular reality, objectivity in the evaluator-evaluatee relationship, a preference for quantitative data, paper and pencil instruments, and a preordinate design (Guba & Lincoln, 1981).

In search of improved means of evaluating quality and effectiveness, it is suggested one "try on" the assumptions and postures of the "naturalistic paradigm" (Guba, 1978; Guba & Lincoln, 1981 and others). Within this framework, one assumes that there are multiple realities, that the evaluator and respondents interact and depend on one another, qualitative data are preferred, the human is often the instrument, and the design is emergent and continual (Guba and Lincoln, 1981).

Since Provus' and Stufflebeam's work in 1971, many of the major evaluation models have shifted to the naturalistic paradigm and will be discussed in the next section.

**Advance Organizers.** Another early consideration for the evaluator(s) is the question of what will guide the evaluation, or how the evaluation will be organized in advance.

Most early efforts, including DEM, were organized around goals or objectives (Tyler, 1950; Stake, 1967; Popham, 1975). With the great school improvement movement in the 1960's, conventional evaluation approaches were found to be inadequate to address the evaluation needs of the many ambitious programs. Evaluation changed dramatically and the following are only examples of the many subsequently developed approaches.

In 1963, Lee Cronbach introduced decisions as a guidepost, as the CIPP model later did, focusing the evaluation for decisionmakers.

Michael Scriven, in 1974, still operating within the scientific paradigm, identified effects as a new advance organizer in his *Goal-Free* evaluation model. He questioned the differentiation between intended and unintended results and the assumption that meeting only intended results would warrant success. He and his teams of "professional evaluators" went in to study effects whether they were intended or not against a profile of needs. Scriven also contributed the concepts of "needs assessment" and the "formative and summative" distinction in evaluation.

Eliot Eisner was the first to break out of the scientific paradigm, beginning the challenge to the appropriateness of science for evaluating programs concerned with human behavior. Eisner's *Connoisseurship* model was the first to employ the human being as the instrument using the metaphor of an art critic or an Olympic judge.

Introducing values and acknowledging pluralism in our society was a significant step outside of the scientific paradigm. Robert Stake, who had earlier authored a popular objectives-based evaluation model developed *Responsive Evaluation*. This model was responsive to the issues and concerns of stakeholding audiences or relevant publics, and was not preordinately designed. He recognized the tradeoff of precise measurement in order to increase the usefulness of the findings to stakeholding audiences.

Guba (1978), and Guba and Lincoln (1981), have pulled together a comprehensive and eclectic model, the *Naturalistic Responsive Evaluation Model* (NRE), designed within the Natualistic paradigm. Their model stretches the responsiveness of Stake's design to respond to human and political factors, qualitative or quantitative methods depending on the identified issues, concerns, and reporting needs of the relevant audiences.

The adversarial process similar to that used in an administrative hearing in the law is the design of the *Judicial Model* developed by Robert Wolf in 1974 and is another example of an issues-organized model within the naturalistic paradigm.

The *Illumination Model* proposed by Farlett and Hamilton in 1977 is similar to the responsive model and is designed to illuminate the problems and issues and focus on information gathering rather than decision making.

Guba (1978), Guba and Lincoln (1981), and Stufflebeam and Webster (1980) provide a wealth of information on the various models and their descriptors as well as references for the writings of the models'authors. These brief descriptions have been included to provide the reader with an overview of the continued development and trends in educational evaluation.

**Methodology.** Once the assumptions, posture and advance organizers are designed, other remaining framework considerations are more easily selected and organized. Being responsive to issues and concerns of stakeholders within the naturalistic paradigm means that some combination of quantitative data (demonstrated by numbers) and qualitative data (portrayed in words) will be gathered depending on audiences'reporting needs, with the likelihood of the evaluation being primarily qualitative. Evaluating "quality" or "effectiveness" of special education programs seems an appropriate arena for employing methods such as interviewing and observation to yield case studies, quality indicator vignettes, portrayals, and other thick descriptions.

Evaluators of network television are no longer relying on "head counts" or numbers of viewers to determine the success and future of television shows; they are using qualitative evaluation of viewers'responses to programming with success and now would not return to relying only on the quantitative. Through interviewing, they are finding out not only how many people are watching what, but also who's watching, why, and when. Advertisers used to assume that "if they watch it, they like it." Now, through interviewing viewers, they know more about how people use television and how they feel about the shows they watch (Roberts, 1984).

There is an analogy to be drawn. The author is choosing to expand assumptions that if child counts are proportionate, kids are in programs and achieving and progressing measurably, then there are quality programs. Those criteria are

not usually what make or break a program. Individuals' perceptions of how well the program is working, how the teachers are teaching, and how the kids feel about being in those programs are only examples of the qualitative data needed to describe and judge program effectiveness.

## Values

Qualitative Responsive Evaluation falls within the naturalistic paradigm. Therefore, in "getting at truth" special education evaluators resist the temptation to seek any one "correct reality" that exists if only we could uncover it. Living through any presidential election, the 1960's, a jury trial or the advisory committee meeting described earlier in this paper, provide proof enough that the world is a pluralistic society with multiple values as well as multiple views of quality. Individuals need to respond to the divergent expectations of those commissioning and using evaluations across all levels; human and political factors greatly affect the success or failure of attempts to evaluate program effectiveness.

At the local level, as at the state and federal agencies, there are several audiences which carry with them countless issues and concerns which they want program evaluation to address, only some of which are compatible. Nowhere does it seem more obvious than in the context of special education. It used to be that decision-making boards were comprised of those who "knew best." The concept of who knows best has become quite broader. What is given up is the pursuit of what is best for all kids; there are now advising and governing boards collaboratively designed to invite multiplicity and even interference to meet the needs of handicapped kids who have a host of audiences affected by the programs which serve them. Homogenizing the data is not a helpful pursuit and often responds to no one's needs. Responding to the issues, concerns, claims, and values of identified stakeholders, even ones not of the same persuasion, predicts greater utility for the evaluation results.

**Stakeholders.** Beginning the implementation of the evaluation design by identifying those individuals who have a stake in the evaluation outcomes and garnering what their expectations of the purposes and uses are is simply good practice. Using issues and concerns as advance organizers rather than intended program goals or objectives is also good practice, since goals are likely to have been designed by a singular audience and may or may not be adequate or ambitious.

Addressing stakeholders at the onset is to be responsive to some of the previously identified issues in the context of local program evaluation. If state and local agencies recognize one another as having a stake in the other's efforts, then developers can work collaboratively for mutual responsiveness rather than in competition for orientation or conflict of responsibility. What is needed is a less perfect model and a more flexible approach.

Similarly, issues of defining "improvement" or "effectiveness" for regular and special educators, and fitting evaluation practices to measuring compliance and describing or judging quality, can be addressed by delineating the issues and concerns of stakeholders and responding to them throughout the design.

**Merit and Worth.** The approach to evaluating presented thus far is one of describing or judging, rather than measuring or assessing. Evaluation, or the process of valuing, can be looked at on two planes. We can describe an entity,

or in this case, a special education program, in two ways; this distinction can be made by looking at merit and value (Scriven, 1978) or "merit and worth" (Guba & Lincoln, 1981, pp. 39-52). *Merit* is the intrinsic value a program has on its own, free of contextual considerations or applicaton; *worth* is the utility something has within a context or applied to a specific purpose. The concept of merit and worth is helpful in our consideration of issues like the notion of replicability; and the compliance and quality relationship.

Local control, or autonomy, is critical to communities. Social mores and values differ not only from state to state, but also from city to town, and household to household. Thus, a meritorious program in Palo Alto's schools may not be worth replicating in South Paris, Maine.

The author was challenged on making such a statement at a national meeting by the question, "Do you meant to say, what's good for handicapped children in California is not good for handicapped children in Maine?" "Exactly, she thought, feeling that she had made her point clear. Her colleague continued with his argument to say that that's not using what we know. "That's like penicillin," he said, implying that she was asserting that each state or local agency should wait to discover their own defense against bacteria. "No," she said, "penicillin is like monitoring (or merit); vitamins, apples, or vegetarian diets – deciding what proactive and supplemental measures to take in addition – is what we don't want to standardize. We want to leave it to the judgment of worth."

Beyond the standards set by law and policy, local communities like to maintain control of decisions regarding the search of excellence or improvement relative to local values and educational philosophy.

State staff from Rhode Island, the smallest state in the union, tell it all when they say, "We can leave our office at eight o'clock in the morning, tour each district in the ocean state, and be back for lunch. But, within our state we have rural, urban, suburban, coastal, island, wealthy, poor, and otherwise diverse organizations." Rhode Island developers chose a self study guide which allowed for local flexibility, and predict that no two sites'program evaluations will look alike. If local control and non-generalizability exist in Rhode Island, it's fair to assume that it's the case throughout the nation.

**Standards.** Special educators are pleased that the standards set forth in the Education of the Handicapped Act were developed; unquestionably, the lives of handicapped students are qualitatively better today than they were prior to 1975. They also share a common goal of developing newer and higher standards, or quality indicators, for educating handicapped students in public schools.

Being overly concerned with standardization, however, may be an inappropriate pursuit of generalizability. Within the naturalistic paradigm, there is no one perfect set of standards which is consensual and fitting to every context. Standards are value-laden; the ideal of determning generic standards for local special education programs is valiant, complicated, and time-consuming. Standards are also judgmental; qualitative judgments are made based on values, perceptions, and intuitive impressions. It seems, therefore, what is needed is minimal standards, some recommended standards, and possibly some selected standards. This would discourage aspiring to describe the perfect special education program quality indicators or program evaluation model, or evaluation instruments to fit all contexts.

## Relevance

> Much that is written on evaluation recommends some one's "scientifically rigorous" pla . Evaluations should, however, take many forms, and less rigorous approaches have value in many circumstances. Scientific quality is not the principal standard, and evaluation should aim to be comprehensible, correct and complete, and credible to partisans on all sides (Cronbach, 1981)

In this section, concerns of rigor and relevance are addressed by describing two sets of criteria: 1) the procedure specifically designed to address the "trustworthiness" of naturalistic evaluations, 2) the standards by which all good evaluations can be judged

### Trustworthiness.

> Problems of rigor arise from the inquirer's need to persuade other inquirers or audiences of the authenticity of the information provided and the interpretations that are drawn from it. How can one tell whether the information and interpretations are correct? Whether the information has purely local significance or might have meaning in many situations? Whether it will be found consistently? Whether the interpretations are free from the particular biases of the inquirer? (Guba & Lincoln, 1981, p. 87)

Local evaluations don't often "glow" with procedures to insure validity and reliability; in fact, many admittedly are of the "quick and dirty" variety. Assuming that individuals concerned with evaluating are concerned with correctly evaluating, the scientific and naturalistic description of four aspects of trustworthiness as outlined by Guba (1981; p. 80) are compared

| Aspect | Scientific Term | Naturalistic Term |
|---|---|---|
| truth value | internal validity | credibility |
| applicability | external validity, or generalizability | transferability |
| consistency | reliability | dependability |
| neutrality | objectivity | confirmability |

Each aspect is restated to fit this context

**Credibility** – "How do we insure that the valuation findings are credible to the identified audience members, particularly the users?"

**Fittingness** – "How do we judge the "worth" of other evaluation strategies or special education programs to our context, or our to other contexts?"

**Dependability and Confidentiality** – "How do we "audit" both that we used sound design and procedure and that we gathered data in ways that other accomplished evaluators would do so?"

Several activities for these procedures are detailed in the literature; what is important here is that qualitative data do not need to be "soft" or lacking in the standards of rigor. Trustworthiness criteria are more "relevant" to real world evaluations outside of the laboratory.

**Good Evaluation.** The Joint Committee on Standards for Educational Evaluation was formed representing all of the audiences concerned with educational evaluation to develop *Standards for Evaluations of Educational Programs, Projects, and*

*Materials,* an impressive document now considered a primer for good educational evaluation. "The Joint Committee was guided by the assumption . . . that sound evaluation can promote the understanding and improvement of education, while faulty evaluation can impair it," (Stufflebeam, et al, 1981; p. 5). Evaluations ought to be useful, feasible, proper, and accurate according to the Joint Committee members who spent years piecing together these guiding principles. Demonstrating that qualitative responsive evaluation has grown to fruition, the standards include audience identification, evaluational interpretation, political viability, human interactions, and analysis of qualitative information as examples of criteria by which evaluations can be judged.

# Implementing Qualitative Responsive Evaluations

## How do you do it?

Since qualitative responsive evaluation approaches are just beginning to find their way into the special education arena, and since local implementers' responses to the assumptions and postures are likely to be, "How do we do it?". The introductory vignette is revisited to portray the implementation considerations.

In moving toward fitting evaluation approaches to their local efforts, the local advisory committee should address the following questions and cues in their negotiations:

**1. Why Evaluate?**
- Decide what purposes and uses are feasible to address.
- . . .what they will do with all the findings?

**2. What will be evaluated?**
- Responsible members must put boundaries on the effort.
- Decide scope and limits.

**3. What are the advance organizers?**
- Check the assumptions members are making both about evaluation and about the special education programs.
- Consider organizing the evaluation around the issues and concerns of stakeholding audiences such as those listed in the introduction.

**4. Who are the relevant audiences?**
- Identify those "who care."
- Look at those who "plan," those who "use," those who "provide," those who "benefit" from, and those who "suffer" from the programs.

**5. What values do they hold?**
- Seek pluralistic perceptions
- Invite interference and accept conflict early on; save disappointment or disaster later.

**6. By what standards should programs be judged?**
- Consider merit (inherent value).
- Don't forget worth (contextual importance).

**7. Design and implement when?**

- Continually design.
- Begin immediately; don't stop implementing until the program does or until you need to furnish a report.

**8. Evaluate where?**

- Conduct evaluation in the natural setting of the program.
- . . .in the community or the "local context."

**9. Evaluate how?**

- Respond to the expressed needs of the stakeholders. (If they want statistics – give them statistics; if they request video-tapes – video-tape. . .within the resources of the evaluation)
- Use as many methods as are available to be responsive and relevant (e.g., interviews, observation, content analysis of documents and records).
- Use human instruments!

**10. What to do with the findings?**

- Look for recurring themes.
- Accept non-homogenized data. If 85% of the respondents make one judgment, explore why 15% have a different opinion.

**11. How to report and advertise results?**

- Revisit the issues and concerns and reporting needs of the identified audiences.
- Report feedback continually in various forms and amounts at various times (narrative and oral vignettes, portrayals, case studies, technical reports, executive summaries).

**12. How to achieve trustworthy data?**

- Spot check information with stakeholders.
- Use an evaluation team to "triangulate" both good procedures and correct findings.

## Summary

As local, state, and federal audiences continue their quest for better program evaluations, requesting "more than numbers" and asking "who cares" are two promising steps to finding more contextually fitting approaches.

> Grownups love figures. When you tell them that you have made a new friend, they never ask you any questions about essential matters. They never say to you, "What does his voice sound like? What games does he love best? Does he collect butterflies?" Instead they demand: "How old is he? How many brothers has he? How much does he weigh? How much money does his father make?" Only from these figures do they think they have learned anything about him. If you were to say to the grown-ups: "I saw a beautiful home made of rosy brick, with geraniums on the windows and doves on the roof," they would not be able to get any idea of that house at all. You would have to say to them: "I saw a house that cost $20,000." Then they would exclaim, "Oh, what a petty house that is! (Saint Exupery, 1943).

# References


Brinkerhoff, R.O., Brethower, D.M., Hluchyj, T., & Nowakowski, J.R. (1983). *Program evaluation: A practioner's guide for trainers and educators: Sourcebook and casebook*. Boston: Kluwer-Nijhoff Publishing.

Cronbach, L.J. and Associates. (1981). *Toward reform of program evaluation*. San Francisco: Jossey-Bass Publishers.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries." *Educational Communication and Technology, 29,* 75-91.

Guba, E.G. & Lincoln, Y.S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass Publishers.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw Hill Book Co.

Lynch, K.B. (1983). Qualitative and quantitative evaluation: Two terms in search of a meaning. *Educational Evaluation and Policy Analysis, 5,* 461-464.

Maxwell, G.S. (1984). A rating scale for assessing the quality of responsive/illuminative evaluations. *Educational Evaluation and Policy Analysis, 6,* 131-138.

Scriven, M. (1972). Pros and cons about goal-free evaluation. *The Journal of Educational Evaluation, 3,* 1-4.

Stake, R.E. (1975). *Program evaluation, particularly responsive evaluation*. (Available from Evaluation Center, Kalamazoo, MI).

Stufflebeam, D.L. & Webster, W.J. (1980) An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis, 6,* 5-20.


## Reference Notes


Beatty, E.E. & Blasczyk, J. (1983, October). Expanding typical evaluation to include the right stuff. Paper presented at the Maine Staff Development Conference, Portland.

Conor, J.S. (1983). *Models for evaluation of special education and related services at the local level*. (Available from the Council of Administrators of Special Education, Inc., Indiana University, Bloomington).

Conor, J.S. (1984). *Models for evaluation of special education and related services at the local level: Final report*. Washington, DC: Associated Consultants, Inc.

Farrow, F.A. (1983). *Effective state monitoring policies: Vol. 4. Quality monitoring and monitoring of state operated programs* (Contract No. 300-82-0829). Washington, DC: The Center for the Study of Social Policy.

National Association of State Directors of Special Education. *State responses to 'annual evaluation' requirements for state plans*. Washington, DC: Author.

Olsen, K.R. (1984, March). *State program evaluation in special education*. (Available from Mid-South Regional Resource Center, University of Kentucky, Lexington.)

Roberts, E. (1984). *Who's really watching television?* Cambridge, MA: Cambridge Forum/NPR.

## About The Author

Edith E. Beatty is the Associate Director of the Northeast Regional Resource Center located at Trinity College in Burlington, Vermont and a Doctoral Candidate at Indiana University in Bloomington, Indiana. A member of the Joint Office of Special Education Programs and Regional Resource Centers National Task Force on Special Education Program Development and Evaluation, Ms. Beatty has responsibility for directing RRC program assistance in program evaluation to the eight Northeast states. Before joining the NERRC, she was a local Director of Special Education in Maine and directed local needs assessments and evaluation efforts for the State Department of Education in Indiana.

# Section Four:

# A Comparison and Analysis Of the Two Approaches to Local Special Education Program Evaluation

*by Kenneth R. Olsen*
*Mid-South Regional Resource Center*
*Lexington, Kentucky*

*The special education administrator becomes involved in program evaluation for a variety of reasons. The administrator must make the decision based on the best fit with his or her own orientation, the stakeholder needs, and the LEA program context. The next section is a review of the similarities and differences in the qualitative and quantitative approaches.*

## Introduction

The local special education administrator has no doubt found the previous two papers by Horvath and Beatty of considerable interest. The authors defined their intentions early in their papers. Horvath stated that his approach was eclectic:

> This article advocates a specific approach to special education program evaluation for special education administrators, and specific strategies for employing the recommended approach. The approach is essentially quantitative, combining several of the most frequently employed design elements as they apply to special education. A major feature of the approach is that various qualitative methodologies can be added, to complement the basic model. These alternative methodologies are described in the companion article within this issues paper. (p. 6)

Beatty told the reader that her approach operated on a different set of assumptions than the "scientific paradigm":

> In this paper, I will describe some responsive evaluation approaches which go beyond the scientific paradigm, ones that are driven by organizers other than goals, and ones which rely more on qualitative methods than on quantitative ones. I'll discuss the issues we're currently facing in program evaluation, explore

some newer, more fitting approaches, and will present a set of questions and cues for how responsive evaluation might fit the local context and provide for effective evaluation of special education programs.

The authors have both done an excellent job of addressing their stated purposes. They have clarified their perceptions of the framework in which evaluation takes place and have put forth convincing arguments for their approaches. But the local special education administrator reading these papers probably has additional questions. For example, the administrator may be asking; "What are the utilities of each of the models?" "Are there clear choices between the two?" "Can the assumptions of each be violated in implementing some sort of an eclectic approach combining the two?" This paper is designed to clarify the similarities and differences between these papers and to draw some conclusions and implications for local special education administrators.

## Overview

**Reviewer's Approach.** This review is grounded in the assumption implicit in the Beatty paper and explicit in the Horvath paper that local special education program evaluation is both necessary and feasible. The reviewer's background and experience has focused on pragmatic evaluations and technical assistance to state and local special education agencies. Therefore, this review attempts to avoid ideological distinctions between the papers when the ideologies have no practical implications. The focus is on helping the special education administrator make more informed decisions about program evaluation.

**General Impressions.** Two general impressions immediately strike the reader when comparing these papers. First, it is clear that the authors have addressed their task at slightly different levels. Horvath describes in detail his assumptions about evaluation realities and the steps necessary to implement his approach. Beatty, on the other hand, spends most of her paper defining her general orientation and attempting to clarify her perception of the contexts in which evaluation takes place. This difference in level of detail makes it somewhat difficult to contrast the approaches and requires that the contrasts be made in relation to generic approaches rather than specific steps. The requirement to make generic comparisons may be an advantage because the second impression is that the authors struggled in vain to stay within a single methodology. Originally Horvath had been asked to advocate "quantitative" methods and Beatty had been asked to advocate "qualitative" methods. Thankfully, neither presented a pure method' on the far end of this one continuum. Instead, their papers present to the reader a rich set of alternatives along many continua. The mixture of methods described in each paper is more realistic than a single dimensional approach. As Miles and Huberman (1984), state: "The history of research in many fields shows shifts from 'either-or' to 'both-and' formulations." Both papers represent compromises and eclectic approaches rather than single paradigms, models or theories. We might better speak of their papers as what Stake calls "persuasions" rather than "models" (1981). Because the terms "paradigm", "model", "theory", and perhaps even "persuasion" are so encumbered with non-essential connotations, this review will use the term "approach" to refer to the content of each paper.

**Topics for Comparisons.** Ten continua have been selected along which the two approaches may be compared. The location of each approach of how the authors (Beatty and Horvath) would answer the following ten questions:

1. Who should make decisions about the evaluation?
2. To what extent is it important that the evaluation meet criteria for "good" evaluations?
3. To what extent should the evaluation focus on quality versus compliance?
4. Should the evaluation collect, analyze, and report numbers or words?
5. Should the evaluation lead to specific judgements or to general enlightment on issues and concerns?
6. Can we know in advance how the evaluation will be conducted?
7. Do "stakeholders" have to agree?
8. Who can conduct the evaluation?
9. To what extent can existing data be used?
10. Should the evaluation focus on processes or outcomes?

## Figure 1

**Comparisons of The Two Approaches on Ten Continua**

H = Approach proposed by Horvath
B = Approach proposed by Beatty

| | | | |
|---|---|---|---|
| 1. | Evaluator Driven | ------H--B- | Stakeholder Driven |
| 2. | Must Meet Criteria for "Good" Evaluations | -HB--------- | "Rigor" is not Critical |
| 3. | Emphasis on "Quality" | --HB------- | Emphasis on Compliance |
| 4. | Uses and Reports Numbers | ---H---B--- | Uses and Reports Words |
| 5. | Focuses on Goals and Decisions | --H------B- | Focuses on Issues and Concerns |
| 6. | Design Is Preordinate | ---H------B- | Design Is Emergent |
| 7. | Stakeholder Consensus Needed | --H------B-- | Stakeholder Pluralism Incorporated |
| 8. | Conducted By Internal Staff | ---H?---B?-- | Uses Consultants |
| 9. | Uses Existing Data | -----H?B?---- | Gathers New Data |
| 10. | Focuses on Outcomes | --H?----B?-- | Focuses on Processes |

The remainder of this paper is intended to show how the authors might answer the first three or four questions in similar ways, the next three or four questions in contrasting ways and to predict how the last three questions might be answered based on what appears to be limited information in the papers. Implications for administrators are presented relative to each continuum.

## Areas of Agreement

The two approaches appear to be in agreement regarding at least three of the continua and are not far apart on the fourth. Both approaches require active stakeholder involvement and valid and reliable processes. They both focus on quality rather than compliance issues. Finally, they both use a mixture of approaches including the use of numbers and words.

### Continuum #1: Stakeholder Involvement

The "Stakeholders" in an evaluation are those whom Beatty defines as those who:

- care about the program;
- plan the program;
- use the program;
- provide the program;
- benefit from the program; and
- suffer from the program.

Both authors indicate that a broad-based group of such individuals should make decisions regarding the evaluation. They both indicate that stakeholder involvement is not simply a first step but is an ongoing part of the evaluation process.

**Implications.** The administrator selecting either of these approaches must be aware that ownership has to be shared in order for the evaluation to be effective. Neither the administrator nor any individual using either of these approaches to evaluation (the "evaluator") can maintain control over the entire process and expect the resultant evaluation to be relevant and useful for causing change. The administrator interested in exploring stakeholder-based evaluations in more detail is referred to Bryk (1983).

### Continuum #2: Meeting Criteria for "Good" Evaluations

**Standards for Evaluations.** Beatty devotes considerable space to the issue of judging her evaluation approach whereas Horvath devotes only limited space to this issue. However, both discuss the use of "multiple data sources" and "triangulating" as ways of creating evaluations that are both feasible and credible or valid. Cost is an obvious concern. Horvath spends two paragraphs discussing evaluation costs (p. 34). They both reference *Standards for Evaluations of Educational Programs, Projects and Materials* developed by the Joint Committee on Standards for Educational Evaluation (1981). Both call for evaluations that are useful, feasible, proper, and accurate.

**Cautions.** Both authors raise questions about paying so much attention to factors to ensure that the Standards for Evaluations are met that the intent of the

evaluation is lost. Beatty cautioned the reader that "relevance is at least as important as rigor". Horvath suggests that typical measurement theory would be inappropriately applied to unique local evaluations. Beatty says that evaluation data must be easily gathered, credible, and useful.

**Implications Regarding Trustworthiness.** The administrator cannot select either approach expecting that it is an "easy" way of addressing program evaluation. Both require structured, systematic, and replicable procedures in order that the resultant information is "trustworthy".

**Implications Regarding Costs.** Both are concerned with costs. There are at least three types of costs that must be considered in conducting an evaluation. The first is the amount contracted for or expended out of pocket to conduct the evaluation beyond day to day administrative costs. This would include costs for consultants, task force meetings, questionnaire production, and mailing and report production. Second is the amount for hidden costs such as staff time, overhead, and other costs that would be difficult to track in an internal evaluation. Finally, there may be a "cost" in terms of side effects. Negative side effects could include such things as loss of instruction time, a hostile atmosphere during an evaluation process, public attention to negative evaluation reports about special education programs that may result in budget reductions, etc. The three types of costs should not always be isolated as evaluation costs. If the evaluation is an integral part of the management practice of the administrator, the cost cannot be separated from the decision making process that should be a normal part of each administrator's activities. Administrators that gather and use evaluation information for program improvement do not object to these costs. Those who feel that evaluations are forced upon them and that the evaluation results are not intended for their benefit, will usually find evaluation costs exorbitant and will resent the evaluation process.

## Continuum #3: Emphasis on Quality

**"Quality" is Distinct from "Compliance".** Both authors agree that their approaches emphasize the evaluation of "quality" and imply that their approaches do more than monitor for compliance with federal and state law. Horvath laments the fact that many states have tried to extend their monitoring systems to serve as systems for evaluating quality. He states:

> A review of several of these extended monitoring systems reveals significant problems. The focus of the evaluation tends to be compliance issues and the quality of the compliance process. This a different focus from an evaluation that measures the quality of the educational process. Extended compliance systems generally do not address outcomes (p. 21)

Beatty agrees. She states that "We have been using the same technology to mease compliance and effectiveness, seeing the two along one continuum. We would do well to recognize their differences. It is unlikely that the same evaluation approach will fit both processes."(p. 48)

**Another View.** Olsen has addressed this issue in a number of papers (1979, 1984 a,b,c) and has taken an opposing view. He proposes that "monitoring" is one process of evaluation and that the difference relates to the types of program standard against which a program is being judged. Experienced local special education administrators have observed six years of monitoring relative

to the Education of the Handicapped Act. Local administrators who have had no training in program evaluation may find it difficult to define differences between monitoring and evaluation. In this author's view, what is now perceived to be a standard of "quality" may at sometime become a compliance standard if it is supported in research and in public opinion and then becomes law or regulation. Therefore, the difference relates to timing. What currently appears in federal and state law consists of extensive procedural requirements. If the laws were revised to include outcome statements, then "monitoring for compliance" could be equated with what some people perceive to be evaluation of "quality".

**Implications.** The local special education administrator may be unable to find authoritative sources outside the school district, i.e. state and federal regulations, around which to design the evaluation. Instead, it will be necessary to depend on the stakeholders to define "quality" when using either of these approaches.

## Continuum #4: Numbers and Words

**Initial Differences.** It is on this continuum that the two authors begin to diverge. Although both authors agree that evaluations may involve the use of both numbers and words, they have different levels of emphasis. Beatty suggests that qualitative data are preferred. The Joint Committee on Standards for educational evaluation defines qualitative information as "Facts and claims presented in narrative, not numerical, form". (p. 155). Horvath cautions readers that "(the) terms quantitative and qualitative, for example, both describe methods that attempt to measure quality. The Joint Committee on Standards for Educational Evaluation define quantitative information as 'Facts and claims that are represented by numbers". (p. 155). Horvath's methods clearly lean more toward the use of quantitative information. For example, he refers to analyzing data using "mainframe statistical packages" and to reporting "percents".

**Not Entirely.** The authors are not dogmatic regarding their positions. Horvath cautions that "undue emphasis on statistics is usually somewhat harmful for program improvement studies, because the focus too easily becomes the detail of the numbers rather than how to improve the program" (p. 30). He also suggests that "a good deal of anecdotal information, planned and incidental, is usually collected during the process". Horvath is not at the far end of the quantitative - qualitative continuum. His method is a mixture of quantitative and qualitative methodology. Likewise, Beatty suggests that a variety of methods be used, including the gathering of quantitative data, depending on the audiences'needs. Both authors allude to the difficulties in analyzing qualitative data but neither discusses this difficult task in detail.

**Implications.** Neither approach may be categorized simply by making a decision about whether it uses numbers or words. When conducting a study under either approach one might be using a checklist, a survey form, an interview protocol or other similar methods. Both authors would say that the more important decision relates to how the information will be used and how the evaluation is focused. The newer qualitative methodology requires a different set of skills and an administrator may wish to read the 7 page overview in How To Evaluate Education Programs (January 1985) to get a sense of the process and some additional references.

## Areas in Which They Differ

While the two approaches are very similar in terms of who makes decisions, the extent to which the evaluation must be trustworthy and the emphasis on quality, and the approaches that differ only in terms of degree regarding the use of qualitative and quantitative methods, they differ greatly in three other areas. There are the areas of focus, the extent to which one may know how the evaluation will be conducted in advance and the extent of stakeholder consensus needed. These differences are orientation issues that tend to drive the whole evaluation. Beatty (and Egon Guba) would probably say that one cannot compromise on the three issues listed below, that any compromise would represent hypocrisy and would represent an approach inconsistent with their philosophy. On the other hand, Horvath might be more likely to say that one may be eclectic regarding these issues. The approach described in his paper falls on the far end of each of the three continua but in practice his orientation is quite eclectic.

### Continuum #5: Issues and Concerns vs. Decisions

**A Clear Difference.** The first difference relates to whether or not the administrator should care about having answers to specific questions when the evaluation is done. Beatty's approach is analogous to "turning the light on". The purpose is to illuminate the issues and concerns, to define the various viewpoints, and to describe what is currently happening relative to the issues and concerns. She notes that one may have "minimal," "recommended," and "selected" standards but these standards are not the focus of her evaluation. The Horvath approach, on the other hand, is clearly goal oriented (i.e., standard oriented). He indicates that the stakeholders should define the program objectives and to sort these objectives into process objectives, product objectives and objectives that deal with both process and product. The Horvath approach is to developing a specific management plan and to a follow-up evaluation. The quality approach leads to "revisit(ing) the issues and concerns and reporting ... to identified audiences."

**Implications.** The administrator has to work with the stakeholders to define the purpose of the evaluation. If there are clear standards against which the stakeholders would like the school district special education program judged and the standards have been articulated in goal objectives and/or tasks, the Horvath approach may be a more direct method of comparing the district to these standards. If on the other hand, the administrator and the stakeholder group have some nagging concerns and issues but are not certain what the problems might be, the Beatty approach may be more appropriate. There are considerations related to the other two main differences between the approaches that may have even greater implications for the administrator's choices.

### Continuum #6: Knowing Design in Advance

**Pre-ordinate Design.** The authors differ in how they would answer the question "Can we know, in advance how the evaluation will be conducted?" In the Horvath approach, developing the design and instrumentation is an extensive step early in the evaluation process. He states that the first order of work "consists of determining how each evaluation question should be answered". He

calls for determining proper evaluation approaches for each objective, developing specific instrumentation and assignments, developing design mechanics and analysis procedures and conducting pilot tests and quality control measures prior to collecting any information.

**Emergent Design.** Beatty's approach is to have the design emerge from the data collection process. Design is continual and builds upon itself. Each finding drives the next data collection procedures. In her model the evaluator goes back to the stakeholders with information and requests direction for the next step of "turning the light on". She calls this a "responsive" evaluation. There is a hint of responsive evaluation in the approach proposed by Horvath. His fifth step (p. 16) involves collecting additional information if important evaluation questions remain unresolved due to collecting conflicting data. However, 90% of what will occur under his approach is clear after the first two stages of his evaluation process. Beatty's approach is more incremental. At any point in time, one might know only 20 or 30% of what might occur in the future using the emergent design approach proposed by Beatty.

**Implications.** The inferences of this continuum have implications for the special education program administrator related to funding and looking at unintended effects. The preordinate design proposed by Horvath is more clearly circumscribed and easier to communicate to funding agencies and school boards. It is fairly simple to prepare an evaluation contract that specifies how much information will be collected, on how large a sample, and how a report will be prepared. The contract can specify the goals that will be evaluated and the total cost. The funding agency and the administrator can control the extent to which the evaluation is allowed to deviate from the design. An emergent design is not so easily circumscribed. When one adopts the phlosophy that one cannot know in advance what he/she should be specifically looking for, the evaluation requires extensive trust of those involved in the evaluation. One can exert only limited power and control in such circumstances. The risk of looking at sensitive areas and/or going beyond the concerns of the majority are much greater under the Beatty approach than the Horvath approach. The benefits of using an emergent design would be the ability, by definition, to look at "unintended consequences," "side effects," and other processes and outcomes of a special education program that may or may not be related to the goal structure of that program.

## Continuum #7: Need for Consensus

The third area where the two approaches clearly differ is in the extent to which the stakeholders have to agree regarding the purpose and procedures for the evaluation. This difference relates to focusing on goals and decisions vs. issues and concerns (Continuum #5), and the issue on stakeholder involvement (Continuum #1). Although both approaches have extensive stakeholder involvement, Horvath indicates that they must agree on the direction the evaluation is taking whereas Beatty indicates that agreement is not only unnecessary, it is artificial and leads to arbitrary results.

**The Consensual Approach.** Horvath suggests working with the stakeholders in a subcommittee format to come to consensus on the goals of education and the critical aspects that need to be evaluated. His approach in-

volves a working subcommittee that reports to the total committee. Critical to his approach is "a common purpose among the stakeholders in a progam" (p. 38).

**The Pluralistic Approach.** The Beatty approach begins with the assumption that there are multiple realities and that minority opinions have equal value to the majority opinion. There is no need to come to consensus. She states what is given up is the pursuit of what is best for all kids; there are now advising and governing boards collaboratively designed to invite multiplicity and even interference to meet the needs of handicapped (students)". She states that individuals need to respond to the divergent expectations of those commissioning and using evaluations across all levels" (p. 30). In her approach, the evaluation must be designed to explore why the minority differs and whether or not their perception of reality might add light to the issues or concerns of both the majority and minority.

**Implications.** The local special education administrator must answer the questions regarding the degree of consensus needed carefully. The implications of using the Beatty approach are that there is no "correct answer". Early in the evaluation process the stakeholders can be surveyed regarding whether or not they agree on the purpose, approach, and use of results from the evaluation. If there is consensus, the use of the Horvath approach may be most appropriate. If there is disagreement, the administrator and the stakeholder group will be forced to decide whether they feel they must use a democratic process that results in consensus or whether the pluralism inherent in any group will be accepted and the evaluation will respond to everyone's needs.

## Continua About Which There is Insufficient Information in the Papers

The positions of the two authors on the seven continua described thus far in this review have been fairly clear. There are three other continua along which the approaches may vary but the papers did not contain sufficiently extensive or parallel information to make comparisons and contrasts. These three issues are perhaps not as important as the previous seven but represent decisions that must be made by the local special education administrator.

### Continuum #8: Internal vs. External Evaluators

**Issue Insufficiently Addressed.** Neither of the authors devoted much space to the issue of using internal or external evaluators. Horvath notes that "most special education administrators do not have the time to manage a program evaluation unless prepared materials/methods and consultive support are provided" (p. 12), suggesting that consultants may be necessary in the initial stages of an evaluation but that the management of a program evaluation may be possible by the special education administrator him or herself. Beatty raises a large number of "who", "what", "where", and "why" questions on pages 33 and 34 of her paper and suggests throughout that the appropriate instrument for evaluation is a human intervener. However, her paper never specifies who that individual might be.

**A Question of Trust and Training.** It is this reviewer's experience that evaluations conducted by internal staff involving only qualitative data are considered suspect by audiences that prior to the evaluation had expressed concerns

regarding program operations. Thus, if the special education supervisor's administrator and supervisors or teachers conducted interviews with parents to obtain perceptions regarding processes and results, the school board and/or parent groups may question the results. On the other hand, if there is a high level of trust between the stakeholders and the administrator and the special education staff have the resources and skills to conduct a qualitative evaluation, such an evaluation can meet stakeholder needs. This issue also relates to the extent to which the local special education administrator must be trained in program evaluation practices. Horvath suggests that it is not reasonable to expect full time administrators to have detailed knowledge of the evaluation profession but suggests that they will need some sort of training to learn overall concepts and become intelligent consumers of evaluation services.

**Implications.** Consultant assistance may be needed regardless of the approach selected. Under the Horvath approach, the state CASE organization may wish to meet and come to consensus on a core set of goals and objectives for special education programs. CASE could then request that the state provide consultant assistance to develop some generic evaluation materials relative to those goals and objectives. Local special education administrators could then adapt such a system with minimal consultant assistance relative to local issues. In the Beatty approach, consultant assistance is needed throughout, especially if credibility is in question. Responsive and qualitative evaluations of the type proposed by Beatty are not replicable across school districts and therefore external support is needed in each case. The Illinois State Board of Education included an excellent chapter on "Selecting An Evaluator" in its *Handbook for Evaluation of Special Education Effectiveness* (1982). The administrator may wish to read that information (six pages) which is both entertaining (e.g., titles such as "Why do Some Evaluators Talk Technical Gibberish?") and informative (e.g., Choosing from Four Styles and Competencies of Evaluators).

## Continuum #9: Using Existing Data

The extent to which the two authors would advocate the use of existing data rather than gathering new data is not clear. LEAs are involved in a variety of data collection efforts including state education agency (SEA), compliance visits, accreditation, and school approval processes; regional association accreditations; staff evaluation systems; vocational education evaluations and Part B evaluations. In most cases, these systems result in quantitative or numerical data and even categorical data i.e., yes/no's. Therefore, this review assumes that existing data would be slightly more useful in applying the Horvath approach than in Beatty's responsive evaluation approach.

**Implications.** In both cases, whether the evaluation question is designed in advance or developed at emerging stages of the evaluation, the process could and should involve reviews of available quantitative and qualitative information prior to gathering new data. This continuum does not represent a major distinction between the approaches. However, the Special Education Programs (SEP) of the U.S. Department of Education has proposed that SEAs (and supposedly LEAs) collect less new information and make better use of existing data in conducting evaluations. This continuum is proposed in this paper to raise the

awareness of local administrators that extensive information already exists and should not be ignored in conducting program evaluations.

## Continuum #10: Processes or Outcomes?

Evaluations may be designed to look at what Stake (1967) calls the "Transactions" or processes in a program, may be designed to look at outcomes or impact, or may be designed to look at both. It isn't clear if the authors differ on this continuum.

**Difference in Emphasis?** Horvath suggests that both process and product (outcome) objectives should be drafted by the stakeholders and used in the subsequent evaluation. He laments the undue emphasis on processes in compliance monitoring and suggests that to conduct evaluations of quality, additional emphasis needs to be placed on looking at outcomes. Consistent with the philosophy in her approach, Beatty does not preordain the emphasis of the evaluation but describes how the stakeholders would make such decisions. Because her evaluation approach is not based upon standards one would not expect her to specify her position on the continuum. Qualitative evaluations are particularly well suited for evaluating processes (How to Evaluate Education Programs, 1985) and her approach places more emphasis on qualitative than quantitative data. Implications for Evaluating Outcomes. There has been a general backlash to emphasis on the procedural requirements in federal law. The concern is that the procedures may not have any correlates to outcomes. The public and funding agencies are looking for more emphasis on outcomes and effects. Either of these methods are appropriate for looking at outcomes. Using Horvath's approach, one would most likely produce percents of people satisfied with their program, numbers of students graduating or achieving at certain levels, and numbers of children receiving certain types of services. These quantitative data would be supplemented by comments and interpretive information. Using Beatty's approach, a report would most likely describe a few case studies explaining why or why not a particular student or teacher was satisfied, achieving, or happy. Contrasting examples also would be presented.

**Implications for Evaluating Processes.** If the evaluation is to focus on processes, the user of the Horvath approach might be able to describe the extent to which a school district school or teacher adhered to a particular process defined in a manual or some other reference source. The user of the Beatty approach would more likely describe what was occuring regardless of what an authority might think should be occuring, i.e., without an external standard. Both models could focus on either outcomes or processes. The type of information would be similar but the orientations and presentations would be different.

# Summary

The special education administrator becomes involved in program evaluation for a variety of reasons. Each administrator comes to the task with a different orientation. Some administrators require extensive structure, organization, and advance planning before becoming involved. Others are willing to tolerate extensive ambiguity, to "go with the flow" and are open to change and new directions. Exter-

nal pressures may be unidirectional or conflicting. The components of the special education program to be evaluated may have been in existence for a long period of time or may just be emerging. All of these factors directly affect the choice of an evaluation approach. The administrator must make the decision based on the best fit with his or her own orientation, the stakeholder needs, and the LEA program context.

## Contrasts

This review has pointed out that for the most part, evaluation approaches are eclectic and some compromises are possible. The distinctions between the two approaches were frequently found to be ones of degree rather than being "either or". There are three areas where compromises using the two approaches presented in this document are unlikely. Continua 5, 6 and 7 in this paper were described as areas where first one cannot be focused both on standards or [illegible] and generically on issues or concerns. Secondly, one cannot have both [illegible] and an emergent design. Finally, one cannot both require consensus and accept pluralistic perceptions of reality.

## Conclusion

Beatty and Horvath have presented approaches to local program evaluation that can assist the special education administrator in making decisions about his or her approach. The administrator must answer the questions posed on page 39 of this paper for himself or herself. Careful consideration of these ten issues will increase the likelihood that the resultant evaluation will be "accurate, proper, useful, and feasible". (Joint Committee on Standards for Educational Evaluation, 1981).

# References


Bryk, A. (Ed.) (1983). *New Directions for program evaluation. Vol. 7. Stakeholder-based evaluation.* San Francisco: Jossey-Bass.

Capitol Publications (1985, January). *How to evaluate education programs: A monthly guide to methods and ideas that work: How to use qualitative evaluation methods.* Arlington, VA: Author.

Illinois State Board of Education (1982, July). *Handbook for evaluation of special education effectiveness.* Springfield: Author.

Joint Committee on Standards For Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw Hill Book Co.

Miles, M. and Huberman, A. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher, 13,* 20-30.

Olsen, K.R. (1979). *Bulletin 10: Evaluating effectiveness in special education programs.* (Available from the Mid-South Regional Resource Center, University of Kentucky, Lexington)

Olsen, K.R. (1984, March). *State program evaluation in special education.* (Available from the Mid-South Regional Resource Center, University of Kentucky, Lexington)

Olsen, K.R. (1984). *Topical issue paper: Integration of evaluation and monitoring systems.* (Available from the Mid-South Regional Resource Center, University of Kentucky, Lexington)

Olsen, K.R. (1984). *Topical issue paper: State standards for special education.* (Available from the Mid-South Regional Resource Center, University of Kentucky, Lexington)

Stake, R. (1967). The countenance of educational evaluation. *Teachers College Record, 68,* 523-540.

Stake, R. (1981). Persuasions, not models. *Educational Evaluation and Policy Analysis, 3,* 83-84.


# About the Author


Kenneth R. Olsen is an Associate Director of the Mid-South Regional Resource Center located at the University of Kentucky in Lexington, Kentucky. He received his doctorate in school psychology from Penn State University in 1975. He is responsible for coordinating a variety of regional and national program evaluation efforts within the Regional Resouce Center Program. He has provided technical assistance regarding program evaluation to special educators at local, intermediate unit, state, and national levels.

# Section Five:

## Administrative Implications

*Prepared by: CASE Research Committee*
*Eileen E. Kittelsen*
*Elizabeth J. Lynch*
*William W. Swan*
*Eileen McCarthy*
*Theodore F. Riggen*
*Leonard C. Burrello, Chairperson*

*The Research Committee and the two outside consultants brainstormed a set of implication statements which were re-worked into a series of questions to provide special education administrators with a framework to guide the selection of approaches to quality program evaluation. The implications in this section are listed below:*

1. *Context – why, purpose, what program?*
2. *Philosophy – beliefs – assumptions*
3. *Use of evaluation*
4. *Trust – credibility and probability of the outcome*
5. *Control – "boundedness"*
6. *Evaluation questions or prob[illegible]ien's*
7. *Power*
8. *Internal/external expertise*
9. *Time frames and work scope – doability*
10. *Evaluation of the evaluation*
11. *Budget*

## Introduction

The purpose of this section is to improve the capacity of the special education administrator to make an informed decision regarding the most appropriate type of program evaluation to conduct in the administrator's unique educational setting. It is recommended that the special education administrator consider these key analytical questions regarding his/her agency. The answers will serve as a guide toward choosing either a qualitative or quantitative approach or a combination thereof. For example, if a special education administrator is operating with a local school board that is placing demands on him/her for data-based effectiveness outcomes, it would likely be politically unwise to conduct a qualitative study. On the other hand, if the district special education staff is placing demands on the special

education administrator for increased participative management and representative involvement in planning and decisionmaking, the administrator may want to pursue a qualitative approach. The decision must be based on a comprehensive analysis of the needs of the district and the community. It is hoped that the following questions guide the administrator in this decision-making process.

The questions are organized under the eleven headings noted above. The side headings provide a means to record your own response to each of the questions the Research Committee poses.

# Self Study Guide

## 1. CONTEXT

- Why is the evaluation being conducted? notes
- What is the expected outcome of the evaluation?
- Who are the "key stakeholders" in this evaluation? internal external
- What is the power base from which the program head is operating? notes
- Who are the key power people? Are they different from some of the key stakeholders? notes

## 2. PHILOSOPHY/BELIEFS

- Is there a written problem beliefs and purpose of the evaluation? yes no
- Is there a statement of beliefs for the evaluation approach being contemplated? yes no
- Is that philosophy based on the stated philosophy of the school district or community? yes no
- Is the evaluator's philosophy consistent with that of the stakeholder? yes no somewhat

## 3. USE OF EVALUATION

- Is the evaluation to be used as a planning tool? yes no somewhat
- Who will use the evaluation to further an understanding of the present conditions of the program? notes
- Who will use the evaluation to inform others? notes

## 4. TRUST

- Is the program head trusted by the stakeholders? yes no somewhat
- What factors will increase the probability that the evaluation will be perceived as valuable? notes

- Is the evaluation being conducted to confirm a predetermined outcome or decision? yes no somewhat

## 5. CONTROL/"BOUNDEDNESS"

- How "controlled", "bounded", or "pre-designed" should the evaluation be? minimally very
- Are standards and outcome measures for the program already clearly stated? yes no somewhat
- Are there preexisting contracts that will dictate the evaluation design? yes no somewhat
- How flexible are the stakeholders in accommodating changes in the evaluation plan? minimally very
- Are the evaluation questions already established by an external force such as the board, state requirements, federal guidelines? yes no somewhat

## 6. PROBLEM STATEMENTS

- Are the problem statements clearly stated? yes no somewhat
- Will the answers to the evaluation questions provide the needed information? yes no somewhat
- Is the system open to the program changes? yes no somewhat

## 7. INTERNAL/EXTERNAL EXPERTISE

- Do any staff members have the expertise to conduct the evaluation? yes no somewhat
- Is there an evaluation unit in the district that can provide technical assistance or staff to conduct the evaluation? yes no somewhat
- If the evaluation is conducted internally, will the results have credibility? yes no somewhat
- How will the staff react to an external evaluation? positively negatively
- Will the evaluation be conducted by a member(s) of the staff (internal) or by an external person(s), or by a combination (external, internal)? internal external combo

## 8. TIMEFRAMES

- What timeframes and constraints exist for the evaluation? note
- Are the timeframes affected by budgetary constraints? yes no somewhat

- Does the evaluation process accommodate the school calendar? yes no somewhat
- What logical factors such as "locating stakeholders" should be considered? note
- At each step of the evaluation process, have the training needs of the internal evaluation team been [illegible]? yes no somewhat
- Is there a specific plan to address each training need? yes no somewhat

## 9. EVALUATION OF THE EVALUATION

- Is there a written plan, accepted by stakeholders, whereby the evaluation effort will be monitored? yes no
- Is there a written plan, accepted by stakeholders, by which the evaluation effort (process and outcome) will be evaluated? yes no
- In the overall plan for evaluation, are there specific points which allow reprogramming as necessary? yes no somewhat
- Is the plan understandable and concise? yes no
- Are the recommendations practical? yes no somewhat
- Are recommendations indepenent or interdependent, i.e., Does one depend upon others? yes no somewhat

## 10. BUDGET

- What are the budgetary needs? notes
- What are the budgetary restraints? notes
- Does the evaluation process need to be modified in light of needs/restraints? yes no somewhat