Promoting Evaluation through Collaboration

Findings from Community-based Programs for Young Children and their Families

R I T A   G.   O ' S U L L I V A N
University of North Carolina – Chapel Hill, USA

A N N E   D ' A G O S T I N O
University of North Carolina – Chapel Hill, USA

Collaborative evaluation engages key program stakeholders actively in the evaluation process. Distance between external evaluators and program staff is often minimal, based on the assumption that a collaborative stance will strengthen evaluation results and increase utilization of evaluation findings. Convincing empirical evidence to support these assumptions is scarce. This study uses evaluation findings from a county-wide, comprehensive, early childhood education initiative to support the contention that a collaborative approach can substantively improve evaluation findings.

K E Y W O R D S :   cluster; collaborative; continuum; networking; utilization

Introduction

Unlike distanced evaluations, collaborative evaluation assumes evaluation results can be strengthened by actively involving program staff and/or stakeholders in the evaluation process. It also assumes that by increasing a program's evaluation skills, there is a greater likelihood that evaluation findings will be used for program improvement. To date, however, convincing empirical evidence to substantiate these assumptions is limited. This study uses findings from a county-wide, comprehensive, early childhood education initiative to demonstrate that a collaborative approach to program evaluation can improve evaluation results in a meaningful way. As a county agent for North Carolina's statewide Smart Start program, one Early Childhood Partnership vision is that 'Every child is important and deserves a healthier, loving, stimulating, and safe environment assured by families and a supportive community.' Objectives to realize these goals are divided into three major areas: education and care, family support, and health/translation/other. Since 1995, each year spending an average of US$6

372

million, the Partnership has arranged annual contracts with approximately 43 governmental and community support agencies to provide about 50 activities across the three areas.

For the past four years, an external evaluation team has been working collaboratively with Partnership staff and contractors to assess program outcomes. This article presents evidence from the second- and third-year evaluations that demonstrates how programs improved their ability to demonstrate effectiveness as a result of this collaborative evaluation approach. Its express purpose is to describe the effects of a collaborative evaluation approach rather than report the program outcomes observed.

Collaborative Evaluation

The term collaborative evaluation often is used interchangeably with participatory and/or empowerment evaluation. The Topical Interest Group (TIG) in the American Evaluation Association is called Collaborative/Participatory/Empowerment Evaluation. Cousins, along with colleagues (1992; 1995; 1996; 1998), has done considerable work in the area of collaborative and participatory evaluation. He defines collaborative evaluation as 'any evaluation in which there is a significant degree of collaboration or cooperation between evaluators and stakeholders in planning and/or conducting the evaluation' (Cousins et al., 1996: 210). He includes a number of evaluation approaches within the general rubric of collaborative evaluation. Conventional stakeholder-based evaluations emphasize the importance of including various stakeholder audiences in the evaluation. Some collaborative approaches focus on enhancing evaluation utilization while others emphasize the importance of empowering participants. Empowerment evaluation assumes that stakeholders will assume leadership of the evaluation process. The collaborative approach used for this evaluation evolved from the circumstances surrounding the evaluation and the authors' belief that collaborative evaluation would yield the most effective results. For the authors, the term 'collaboration' implies that people share responsibility and decision making. When stakeholders are is asked to provide information for an evaluation, technically they are participating in that evaluation, but they are not necessarily collaborators in the evaluation design. Similarly, program participants are usually not program collaborators in determining the content or direction of the program. For the authors, collaborative evaluation, rather than the term participatory evaluation, implies the desired level of involvement. To the extent that they are able, program staff and other stakeholders should be considered part of the evaluation team. This does not relieve the evaluator of the overall responsibility for conducting the evaluation or producing evaluation results. Evaluators are engaged because of the expertise they bring to the endeavor, and leadership for the evaluation resides with that expertise.

Collaborative evaluation is often empowering to participants. They enhance their understanding of evaluation and gain new skills, two empowering events. As such, it is a valuable positive outcome of the process but not an intended one as described by Fetterman (1996) or Burke (1998).

373

Collaborative evaluation may be viewed as a natural progression from responsive evaluation. Not only does the evaluation need to be responsive to the programs' needs, but it also should be responsive to the needs of the stakeholders to find the evaluation useful and the needs of the community to have people who possess expertise in evaluation. Thus, evaluators can improve the general state of evaluation by taking every opportunity to enhance clients' ability to appreciate, understand, and conduct evaluations. This is not just conceptually sound but practically useful as well.

Collaborative evaluation also promotes utilization of evaluation findings. Utilization of evaluation findings continues to be a central problem in the field (Ciarlo, 1981; Patton, 1986; Smith, 1988; Stevens and Dial, 1994). Patton (1997) would probably argue it is *the* problem in the field. Some assert that the evaluator should be the person responsible for promoting evaluation use (Chelimsky, 1986; Cousins et al., 1996; Knott, 1988; Mowbray, 1988; Weiss, 1998). The authors, along with others (Fetterman, 1996; Greene, 1987; Guba and Lincoln, 1989; Levin, 1996; Patton, 1988; Linney and Wandersman, 1996), believe that involving stakeholders in the evaluation process will improve evaluation utilization.

Program staff often ignore evaluation findings, because they do not understand them or have not been involved directly in the planning and implementation of the evaluation process. Distanced evaluators, conducting distanced evaluations, fail to engage program stakeholders in the evaluation and thereby limit the potential for the findings to positively influence the program. Logically, if program personnel are collaboratively involved in the evaluation, then their use and understanding of the findings should increase.

Collaborative evaluation approaches have long been controversial in the field (O'Sullivan, 1995). Traditionally, evaluators have debated appropriateness of roles that involve evaluators directly with program staff, whereas collaborative evaluators question the relevance and effectiveness of distanced evaluations. More recently, Stufflebeam (1994) debated with Fetterman (1995) the value of empowerment evaluation. Scriven (1996a, 1996b) raised objections to collaborative evaluation and the potential co-optation of the evaluator, as familiarity with programs and program staff increases.

Until recently, Scriven has cast collaborative evaluation as evaluation consulting rather than actual evaluation. Yet, in many instances program staff in collaborative evaluation can be considered extensions of the evaluation team. As such, they bring to the evaluation effort different levels of expertise that the team leader recognizes and incorporates into assignments. From a management perspective, strengthening evaluation skills among team members is an extremely desirable practise. Most evaluators who engage in collaborative evaluation believe that the advantages gained in program awareness, staff co-operation, access to information, quality of information gathered, and enhanced receptivity to findings, far outweigh the potential for (not the presence of) biased findings (Fetterman, 1995, 1996; Levin, 1996; Patton, 1997).

While the number and types of collaborative and participatory evaluation approaches expand, the accumulation of evidence supporting the approach grows (King, 1998). Much of the literature focusing on participatory evaluation shares

374

lessons learned from the field (see for example, Coupal and Simoneau, 1998; Gaventa et al., 1998). This expanding body of evidence is encouraging but does little to silence those who fundamentally oppose the approach.

The authors practice collaborative evaluation by designing evaluations that engage clients in the process. The level of engagement varies by program evaluation purpose and client, but generally, the authors seek clients who want to actively participate in a collaborative evaluation. Additionally, in light of usually limited evaluation funds, clients/collaborators contribute substantive resources that allow expanded evaluation activities.

Evaluation Voices: A Type of Collaborative Evaluation

Evaluation of the county Early Childhood Partnership presented a number of challenges to the external evaluation team. With more than 50 contracted activities across three programmatic areas, the scope of the evaluation was broad. Even within the three programmatic areas, a wide array of activities was being funded. The budget for evaluation was quite modest and for the US$6 million annually that supported programs, only US$40,000 was available for evaluation. The special circumstances of the evaluation suggested that the evaluation design needed to be carefully constructed. Evaluation Voices (EV) was selected as the approach that would meet many of the design needs.

EV (O'Sullivan and O'Sullivan, 1994, 1998) is an innovative evaluation approach that combines the elements of cluster evaluation with collaborative consensus models of community development. Cluster evaluation was created (Kellogg Foundation, 1991) to strengthen the evaluation expertise among similar programs, thereby better meeting funders' needs for program evaluation information. The literature supports that many positive gains in evaluation occur as the result of cluster networking activities (Barley and Jenness, 1993; Henry, 1992; Jenness and Barley, 1992; Mansberger, 1993; Pearl and Rubino, 1993; Seefeldt, 1992). Community Voices (Callaway et al., 1993) is a leadership development program process aimed at fostering community problem solving among limited resource communities; its process points the way to active participation among constituents. EV combines these two approaches enabling program personnel to strengthen their ability to critically evaluate program effects.

EV assumes that programs with similar goals can strengthen their evaluation strategies through cluster networking and must build evaluation expertise from within. To accomplish this, evaluators focus on a process that:

1. begins with program staff and stakeholders sharing their vision for the program;
2. recognizes the current level of evaluation skill present within programs;
3. identifies barriers to strengthening evaluation;
4. provides training to overcome the identified barriers to evaluation; and
5. generates an action plan to determine next steps in the evaluation process.

EV currently is being used in the areas of education, non-profit management, and

375

sustainable agriculture. Conceptually, EV assumes that effective evaluation is an ongoing activity that spirals through a consistent process of:

1. perceiving a vision for the program;
2. forming evaluation questions relevant to that vision;
3. designing and implementing an information system to help answer those questions; and
4. summarizing the information collected so that the vision for the program may be revisited.

The county Early Childhood Partnership had multiple purposes with multiple service providers working together to realize the project's goals. It was essential that the evaluation process be viewed as something collaborative, something of use to everyone and not something imposed from the top of the project hierarchy. Only in this manner would the data collected provide useful information upon which to implement positive changes and assess project outcomes. Initially, evaluation networking conferences identified representatives from similar service providers and invited them to cluster meetings of 12–15 people. While attendance was voluntary, almost 80 percent of the 52 eligible programs were represented at the first three cluster meetings. The purpose of these meetings was to discuss the overall evaluation plan and amend it to fit their needs. Subsequently, two EV networking meetings were held with selected service providers to prioritize important outcomes, share information about program implementation strategies and evaluation approaches, learn new evaluation techniques, and identify technical assistance needs. Attendance at these meetings was also voluntary and averaged between 40 percent and 60 percent.

Implementation of the Evaluation

Initially in 1995, the program director and a committee member from the evaluation advisory group visited the evaluator to discuss the possibilities for evaluation. The Partnership was in its first 18 months of operation and only six months into the first program implementation year. The program was, and still is, politically sensitive in the state, which meant that its existence could, in fact, be influenced by evaluation results. The evaluation challenges were impressive: the large number of agencies collaborating to provide services, the large number of programs, the limited evaluation funds, and the political sensitivity of the program to evaluation findings. Also, the fact that the services to be provided would vary greatly between individuals added to these challenges. As part of one activity, a child might receive vision screening and no other services from the program; another child might receive subsidized daycare in a preschool center that was working on quality enhancement supported by the program, and the child's parent(s) might receive home visits from another of the program's support services.

Clearly, given the size of the program and the available resources to conduct the evaluation, no viable evaluation could occur without program contractors becoming active participants in the evaluation. They would have to supply basic

376

information about program services that they had already provided as part of the state's mandatory quarterly reporting requirements. Beyond that, these contractors also would have to collect evaluative evidence about their program accomplishments (outcomes). The external evaluation team would need to spend time working with the contractors, set up data collection systems, design instruments, analyze data, and conduct focused studies on important evaluative outcomes (e.g. client satisfaction, quality care, parent education, etc.). Thus, the external evaluation team members could be seen as senior contributors to the evaluation process, while the programs made important but less sophisticated contributions. The key to the success of the evaluation rested with the ability of the evaluators to engage contractors in a collaborative evaluation process. Figure 1 summarizes facts about the program and lists evaluation activities for the first three years of the evaluation.

The evaluation team had proposed using EV cluster networking activities as the way to begin assessing and strengthening evaluation expertise among the program's contractors. The evaluation team held evaluation cluster meetings in the first year of the evaluation to:

- orient contractors to evaluation;
- share the overall evaluation plan;
- explain state reporting requirements;
- help contractors draft annual evaluation plans; and
- share data gathering strategies.

During these meetings and subsequent individual technical assistance visits, the evaluation team emphasized the importance of finding out what contractors wanted to know about their activities, which almost always coincided with what the program wanted to know overall.

The level of evaluation expertise varied greatly by contractor. A few programs were fairly sophisticated in their evaluation practice, while a corresponding number had never collected service statistics before. Most were struggling through the first year of program implementation with the usual delays in hiring, opening new facilities, launching new programs, etc. The state added to these first year difficulties as it worked through its own program start-up complexities, which included changing the format of their quarterly reports three times. The first year's evaluation report (O'Sullivan et al., 1996) provided overall service statistics from programs, shared success stories, reported the results of a survey to identify quality care standards in the county, and began sharing information about county-wide indicators of importance (e.g. infant mortality, number of day care slots in the community, collaboration, etc.).

Building on the year-one activities, the second year of the evaluation began by transferring the compilation of service statistics to the program office and working to strengthen contractors' evaluation plans. EV cluster networking meetings continued as the way to implement this strategy. Contractors participated in cluster workshops on evaluation planning that were followed by individual technical assistance as required. During these workshops contractors were told that they would be asked to share interim evaluation results at an Evaluation

State Fact Sheet

- Program begun by the Governor in 1994 to promote school success for children under six and their families.
- Four components: education and quality care; parent support; health; and transportation.
- Collaborations of existing county agencies with additional support services created as needed.
- 48/100 active, 52 poised to begin.

County Partnership Fact Sheet

- Program started in 1995, annual budget US$8 million.
- 1997–1998: 63 contracted activities across 52 contractors; annual budget US$6 million.
- Annual external evaluation contract: US$36,000–40,000.

Evaluation Activities

1995–1996	1996–1997	1997–1998
• Develop database system. • Cluster evaluation meetings. • Technical assistance in evaluation. • Delphi of quality care. • Collaboration survey. • Success stories compilation. • Annual evaluation report.	• Transfer database to Partnership. • Individual program evaluation plans. • First Evaluation Fair. • Develop Parent Education measures. • Collaboration survey. • Success stories compilation. • Outcome indicator ID. • Annual evaluation report.	• Redesign database for transfer. • Evaluator at Partnership 10 hrs/wk. • Technical assistance for evaluation plans. • Second Evaluation Fair. • Evaluation Fair results used for allocations. • Collaboration survey. • Outcome indicator ID. • Analysis of Partnership and program collected data. • Design Jump Start evaluation. • Annual evaluation report.

Figure 1.   Smart Start: The Early Childhood Partnership Fact Sheet

Fair to be held mid-year. During the Evaluation Fair contractors were expected to report their results by clusters to their peers. At the same time, they were asked to submit a written report of mid-year accomplishments. The external evaluation team members were available to assist contractors with implementation of their evaluation plans. The external evaluation team also worked with the overall program to develop parent education measures, assess collaboration and continue to report on important outcome measures. The Evaluation Fair was held and interim results summarized. At the end of the second year, interim evaluation results were updated and included as part of the second evaluation report (O'Sullivan et al., 1997).

By the third year of the evaluation, the evaluation processes established during the first two years has taken root and successful patterns continued. Evaluation planning occurred during the beginning of the year, with the Evaluation Fair scheduled once again for mid-year. Demand for external evaluation services was such that evaluation team members spent 10–15 hours each week at the program office, providing technical assistance to contractors and staff. Most contractors saw the external evaluators as collaborators and requests for technical assistance increased. Not surprisingly, the quality of evaluation plans improved, as did the timeliness with which they were submitted. External evaluation team members also were asked to assist with data analysis for contractor- or program-collected data. Additional work continued on the identification of parent education measures and other common instruments.

Method

Anecdotal evidence from the evaluation suggested strongly that the collaborative approach used with contractors was working to strengthen evaluation findings. Reporting by the contractors at the Evaluation Fairs in 1997 and 1998 provided a unique opportunity to empirically test the assertion that the quality of reporting had improved. All contractors participated in the overall evaluation. Thus, they all had the same general reporting requirements. All were supported by the same external evaluation team and had equal opportunities for cluster meeting attendance and technical assistance in evaluation. All three years, contractors presented equivalent, pre-existing differences concerning their evaluation acumen. At the same time, contractors had equal access to evaluation assistance. Thus, from a research design perspective, each year the evaluation treatment was constant. Finally, the sheer number of programs would allow meaningful frequency counts, and differences in activity areas would permit comparisons across education and care, family support, and health/translation/ other.

The procedure followed for each of the two Evaluation Fairs was uniform. Programs developed evaluation plans and were told they were expected to report mid-year. They were asked to make a five-minute presentation by cluster at the Evaluation Fair and submit a written report that followed a prescribed format. Evaluation Fair presentations were then summarized by program area and presented to the Partnership Board. From these summaries the two authors

379

independently reviewed the contents and established criteria for review. General reporting activities (i.e. number of programs not reporting, number presenting at the Evaluation Fair, number completing written reports, and number reporting service units) were examined along with the type of outcome data reported. In addition to activity counts (service units), programs were specifically asked to report activity results (outcomes). For the purpose of coding the report summaries, outcomes were divided into the following six categories.

1. No reporting.
2. Reporting service units only, rather than sharing outcomes.
3. Reporting outcomes that were not supported.
4. Reporting some unsupported outcomes with some data-supported outcomes.
5. Reporting mostly data-supported outcomes.
6. Inclusion of comparative data in the outcomes reported.

These categories represent a continuum of evaluation practice (O'Sullivan, 1989). This continuum begins with programs that do not document program activities; they usually see their charge as implementing the programs that they passionately promote and do not feel compelled to collect information about program services. Next on the continuum are programs that view evaluation as exclusively monitoring (e.g. number of people trained or number of services provided to the targeted audience is all they report). The next level of evaluation practice recognizes the importance of outcomes; however, programs report data-free improvements (e.g. parents increased their knowledge of child development; the quality of day care improved) without specifying any degree of change. At the fourth level, programs begin to report some outcomes that specify amount of improvement/change (e.g. 85% of participants read daily with their children; 5 parents found jobs) but data free outcomes are in the majority. At the fifth level, the majority of outcomes reported are supported by evidence. Finally, at the most advanced level, programs report outcome evidence comparatively (e.g. fathers reported spending, on average, 75% more time playing with their children; 70% of students showed above average developmental scale scores).

Results

Table 1 shows a comparison of 1997 and 1998 frequencies for how contractors reported their activities and outcomes achieved at the Evaluation Fair. Contractors' reporting results were divided by activity area (education/care, family support, and health/translation/other) and then are shown combined. Although there was some fluctuation in contracted activities from 1997 to 1998, approximately 90 percent of the contractors reported in both years.

Under the education/care category of service, the number of programs increased from 18 in 1997 to 29 in 1998 and the number of programs not reporting increased slightly from one in 1997 to three in 1998. The number of programs that included service units in their reports increased from 16 in 1997 to 25 in 1998. In 1998, there were a total of eight new programs, of which one contractor

380

Table 1. Comparison of 1997 and 1998 Smart Start Evaluation Fair Data

Category of service/year	No. of programs	Reporting activities					Outcome data reported			
		Not reporting f (%)	Eval. Fair presentations f (%)	Eval. Fair written reports f (%)	Reporting svc. units f (%)	Reporting svc. units only f (%)	Reporting data-free outcomes f (%)	Reporting some data-free/some data-supported outcomes f (%)	Reporting mostly data-supported outcomes f (%)	Reporting comparative data f (%)
Education/care										
1997	18	1 (6.0)	13 (72.2)	17 (94.4)	16 (94.1)	1 (5.9)	14 (82.3)	2 (11.8)	0 (0.0)	0 (0.0)
1998	29	3 (10.3)	21 (72.4)	26 (89.7)	25 (96.2)	1 (3.8)	10 (38.5)	9 (34.6)	6 (23.1)	2 (7.7)
Family support										
1997	24	10 (41.7)	15 (62.5)	14 (58.3)	10 (71.4)	6 (42.9)	6 (42.9)	0 (0.0)	2 (14.3)	0 (0.0)
1998	19	0 (0.0)	19 (100.0)	19 (100.0)	19 (100.0)	0 (0.0)	6 (31.6)	8 (42.1)	4 (21.0)	1 (5.3)
Health/ translation/other										
1997	10	1 (10.0)	6 (60.0)	9 (90.0)	5 (55.6)	2 (22.2)	4 (44.4)	0 (0.0)	3 (33.3)	0 (0.0)
1998	9	1 (11.1)	6 (66.7)	5 (55.6)	7 (87.5)	1 (12.5)	3 (37.5)	3 (37.5)	1 (12.5)	0 (0.0)
TOTALS										
1997	52	12 (23.1)	34 (65.3)	40 (76.9)	31 (77.5)	9 (22.5)	24 (60.0)	2 (5.0)	5 (12.5)	0 (0.0)
1998	57	5 (8.8)	46 (80.7)	50 (87.7)	51 (98.1)	2 (3.8)	19 (36.5)	20 (38.5)	11 (21.2)	3 (5.8)

Note:
f = frequency

381

submitted separate reports for each of seven activities. Of the eight new programs, three did not submit any evaluation reports.

Regarding outcomes, there was one program reporting service units only, both in 1997 and 1998. The number of programs reporting data-free outcomes fell from 14 to 10. Four of the programs reporting data-free outcomes in 1998 were new. The number of programs reporting at least one or more data-supported outcomes together with data-free outcomes increased from two in 1997 to nine in 1998. The number of programs reporting mostly data-supported outcomes increased from zero in 1997 to six in 1998. The number of programs reporting comparative data increased from zero in 1997 to two in 1998.

Under the category of family support, the number of programs decreased from 24 in 1997 to 19 in 1998. Additionally, there were two new programs in 1998 and the number reporting increased from 14 in 1997 to 19 in 1998. The number of programs that included service units in their reports increased from 10 in 1997 to 19 in 1998.

Regarding outcomes, the number of programs reporting service units only fell from six in 1997 to zero in 1998. The number of programs reporting data-free outcomes remained at six both in 1997 and 1998. Two of the programs reporting data-free outcomes in 1998 were new. The number of programs reporting at least one or more data-supported outcomes with data-free outcomes increased from zero in 1997 to eight in 1998. The number of programs reporting mostly data-supported outcomes increased from two in 1997 to four in 1998. The number of programs reporting comparative data increased from zero in 1997 to one in 1998.

Under the category of health/translation/other, there were three new programs in 1998. In 1997, only one program did not report evaluation results and of the three new programs in 1998, only one did not report any evaluation results. The number of programs that included service units in their results increased from five in 1997 to seven in 1998.

Regarding outcomes, the number of programs reporting service units only decreased from two in 1997 to one in 1998. The number of programs reporting data-free outcomes decreased from four in 1997 to three in 1998. The number of programs reporting at least one or more data-supported outcomes with data-free outcomes increased from zero in 1997 to three in 1998. Unfortunately, in this category, the programs reporting mostly data-supported outcomes fell from three in 1997 to one in 1998. There was no change from zero for those reporting comparative data.

Overall, there were 13 new programs in 1998, with the number not reporting evaluation results falling from 12 (23.1%) in 1997 to five (8.8%) in 1998. Four of the five programs that did not report in 1998 were new. The number of programs that included service units in their evaluation reports increased from 31 (77.5%) in 1997 to 51 (98.1%) in 1998.

Regarding outcomes, the number of programs reporting service units only as their results fell from nine (22.5%) in 1997 to two (3.8%) in 1998, and the number of programs reporting data-free outcomes fell from 24 (60%) to 19 (36.5%). The number of programs reporting at least one or more data-supported outcomes with data-free outcomes increased from two (5%) in 1997 to 20 (38.5%) in 1998.
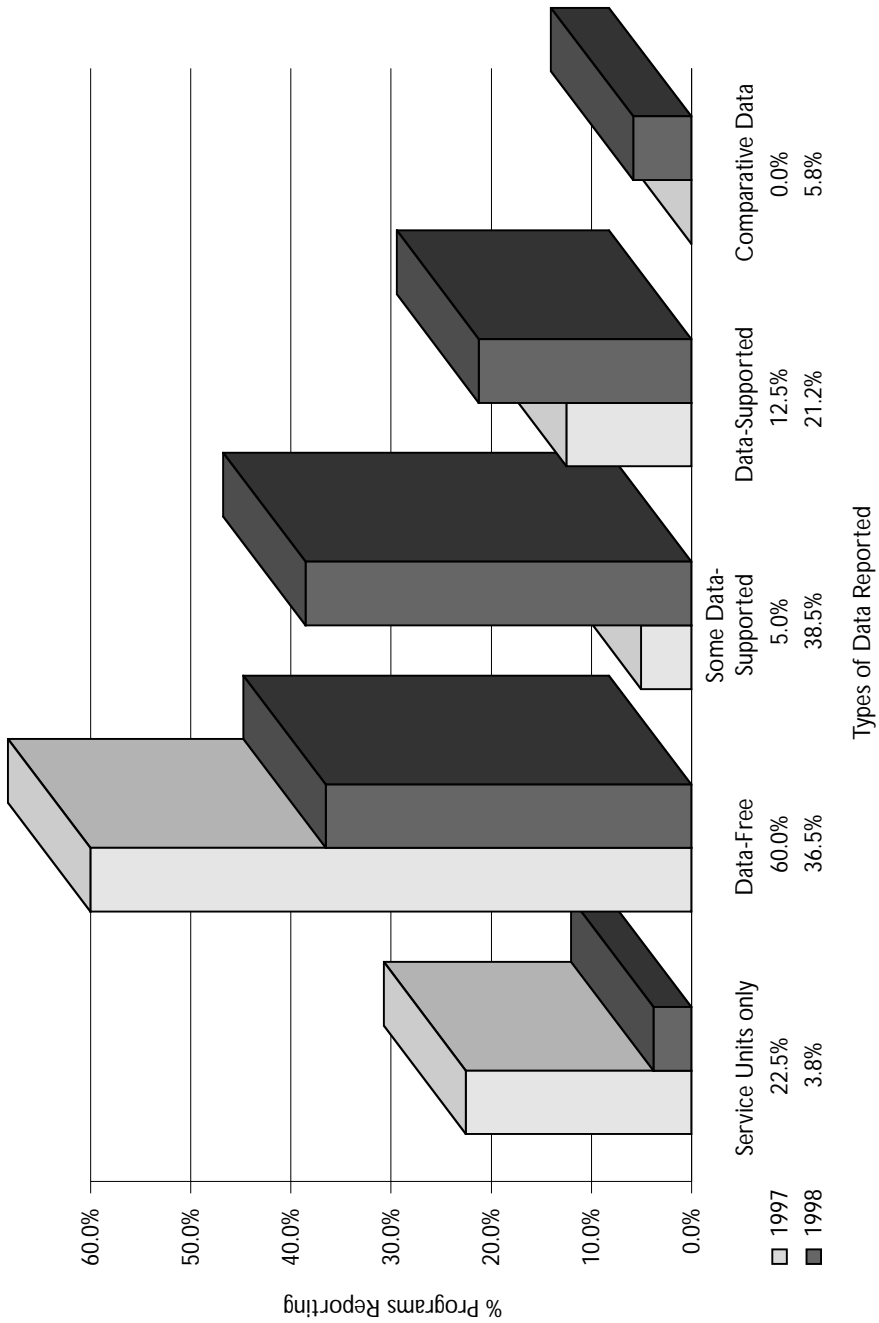
382

Figure 2. 1997 and 1998 Evaluation Fair Results

	Service Units only	Data-Free	Some Data-Supported	Data-Supported	Comparative Data
1997	22.5%	60.0%	5.0%	12.5%	0.0%
1998	3.8%	36.5%	38.5%	21.2%	5.8%

Types of Data Reported

% Programs Reporting

The number of programs reporting mostly data-supported outcomes increased from five (12.5%) in 1997 to 11 (21.2%) in 1998. The number of programs reporting comparative data increased from zero in 1997 to three (5.8%) in 1998.

In terms of strengthening evaluation reporting, the data suggest that contractors improved the quality of their reporting of program outcomes to a remarkable extent. Figure 2 graphically demonstrates positive changes in reporting of outcomes for the top five of six levels on the evaluation continuum. The number of contractors, who reported service units as outcomes, dropped dramatically from 22.5 percent to 3.8 percent. Data-free outcome reports were almost halved, while percent of contractors who used some data to support their outcome reports rose more than sevenfold. Six more contractors joined those who reported outcomes that were mostly supported with evidence. Finally, the 1998 Evaluation Fair saw three contractors report outcome results using some comparative information, whereas in 1997 no contractors used comparative data.

Conclusions

Overall, the evidence supporting the use of collaborative evaluation methods is encouraging. Not only did evaluators have the opportunity to use EV with a large number of programs (43) across different activity areas (education, family support and health/translation/other), but the same evaluation strategy was used for all concerned. Additionally, effects were observed during a two-year period.

Most evaluators support collaborative evaluation, based on the assumption that this approach measurably improves the quality and utilization of evaluation findings. The findings from this study, while still preliminary, lend strong support for the quality improvement supposition of collaborative evaluation. The preliminary nature of these findings also applies to the use of EV components. Clearly stating reporting requirements, cluster networking or the annual presentation requirement might have been sufficient to produce the effect observed. Plans to test the assumption that collaborative evaluation improves utilization are underway.

The Evaluation Fair coupled with evaluation planning has been used successfully by the authors in other places with similar programs and with programs quite different from the one discussed. The opportunity to collect similar evidence has been limited due to potential program confounds or lack of continuity across years. Subsequent Evaluation Fair data should further expand and explore the data trend.

References

Barley, A. Z. and M. Jenness (1993) 'Cluster Evaluation: A Method to Strengthen Evaluation in Smaller Programs with Similar Purposes', *Evaluation Practice* 14(2): 141–7.
Burke, B. (1998) 'Evaluation for a Change: Reflections on Participatory Methodology', *New Directions for Program Evaluation* 80: 43–56.
Callaway, S., M. Arnold and P. Norman (eds) (1993) *Community Voices: Leadership*

384

Development for Community Decision-making, Program Implementation Guide. Prepared with Funding from the W. K. Kellogg Foundation. Greensboro, NC: Cooperative Extension Program, North Carolina A&T State University.

Chelimsky, E. (1986) 'What Have We Learned about the Politics of Program Evaluation', *Evaluation Practice* 8(1): 5–22.

Ciarlo, J. A., ed. (1981) *Utilizing Evaluation.* Thousand Oaks, CA: Sage.

Coupal, F. P. and M. Simoneau (1998) 'A Case Study of Participatory Evaluation in Haiti', *New Directions for Program Evaluation* 80: 69–79.

Cousins, J. B. and L. M. Earl (1992) 'The Case for Participatory Evaluation', *Educational Evaluation and Policy Analysis* 14(4): 397–418.

Cousins, J. B. and L. M. Earl (1995) *Participatory Evaluation in Education: Studies of Evaluation Use and Organizational Learning.* London: Falmer.

Cousins, J. B., J. J. Donohue and G. A. Bloom (1996) 'Collaborative Evaluation in North America: Evaluators' Self-reported Opinions, Practices, and Consequences', *Evaluation Practice* 17(3): 207–26.

Cousins, J. B. and E. Whitmore (1998) 'Framing Participatory Evaluation', *New Directions for Program Evaluation* 80: 5–23.

Fetterman, D. M. (1995) 'In Response to Dr Daniel Stufflebeam's "Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go"', *Evaluation Practice* 16(2): 179–99.

Fetterman, D. M. (1996) 'Empowerment Evaluation: An Introduction to Theory and Practice', in D. M. Fetterman, S. J. Kaftarian and A. Wandersman (eds) *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability.* Thousand Oaks, CA: Sage.

Gaventa, J., V. Creed and J. Morrissey (1998) 'Scaling Up: Participatory Monitoring and Evaluation of a Federal Empowerment Program', *New Directions for Program Evaluation* 80: 81–94.

Greene, J. G. (1987) 'Stakeholder Participation and Utilization in Program Evaluation', *Evaluation Review* 12(2): 91–116.

Guba, E. G. and Y. S. Lincoln (1989) *Fourth Generation Evaluation.* Thousand Oaks, CA: Sage.

Henry, R. (1992) 'A Delicate Balance: The Role of the Cluster Evaluator in Relation to the Foundation Project Directors'. Paper presented at the Annual Meeting of the American Evaluation Association. 4–7 November, Seattle, WA.

Jenness, M. and Z. A. Barley (1992) 'A Constructivist Approach to Cluster Evaluation'. Paper presented at the Annual Meeting of the American Evaluation Association. 4–7 November, Seattle, WA.

Kellogg Foundation (1991) *Information on Cluster Evaluation.* Battlecreek, MI: Author.

King, J. A. (1998) 'Making Sense of Participatory Evaluation Practice', *New Directions for Program Evaluation* 80: 57–67.

Knott, T. D. (1988) 'The Impact of Major Consumer Groups on Health Care Evaluation', *New Directions for Program Evaluation* 39: 33–46.

Levin, H. M. (1996) 'Empowerment Evaluation and Accelerated Schools', in D. M. Fetterman, S. J. Kaftarian and A. Wandersman (eds) *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability.* Thousand Oaks, CA: Sage.

Linney, J. A. and A. Wandersman (1996) 'Empowering Community Groups with Evaluation Skills: The Prevention Plus III Model', in D. M. Fetterman, S. J. Kaftarian and A. Wandersman (eds) *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability.* Thousand Oaks, CA: Sage.

385

Mansberger, N. B. (1993) 'Making Sense Out of a Mixed Bag of Data: The Use of Working Hypotheses'. Paper presented at the Annual Meeting of the American Evaluation Association, Dallas, TX, 3–6 November.

Mowbray, C. T. (1988) 'Getting the System to Respond to Evaluation Findings', *New Directions for Program Evaluation* 39: 47–58.

O'Sullivan, R. G. (1989) 'A Positive Paradigm for Pondering: Wholistic Health and Educational Program Evaluation'. Paper presented at the Annual Meeting of the North Carolina Association for Research in Education: Research Triangle Park, NC.

O'Sullivan, R. G. (1995) 'From Judges to Collaborators: Evaluators' Role in Science Curriculum Reform', *New Directions for Program Evaluation* 65: 19–30.

O'Sullivan, R. G., J. Clinton, H. Schmidt-Davis and S. Wall (1996) *First Year Smart Start Evaluation of the Forsyth Early Childhood Partnership*. Greensboro, NC: Teaching to Diversity, University of North Carolina at Greensboro.

O'Sullivan, R. G., A. D'Agostino, B. Prohm, L. Roche and H. Schmidt-Davis (1997) *1997 Annual Evaluation Report of the Forsyth Early Childhood Partnership*. Greensboro, NC: Teaching to Diversity, University of North Carolina at Greensboro.

O'Sullivan, R. G. and J. M. O'Sullivan (1994) 'Evaluation Voices: Promoting Cluster Evaluation from Within Programs'. Paper presented at the Annual Meeting of the Canadian Evaluation Society: Quebec, Canada.

O'Sullivan, R. G. and J. M. O'Sullivan (1998) 'Evaluation Voices: Promoting Evaluation from Within Programs through Collaboration', *Evaluation and Program Planning* 21(1): 21–9.

Patton, M. Q. (1986) *Utilization-focused Evaluation,* 2nd edn. Thousand Oaks, CA: Sage.

Patton, M. Q. (1988) 'Integrating Evaluation into a Program for Increased Utility and Cost-effectiveness', *New Directions for Program Evaluation* 39: 85–94.

Patton, M. Q. (1997) *Utilization-focused Evaluation,* 3rd edn. Thousand Oaks, CA: Sage.

Pearl, J. B. and A. Rubino (1993) 'The Power of Cluster Evaluation Networking Conferences', Paper presented at the Annual Meeting of the American Evaluation Association, Dallas, TX, 3–6 November.

Scriven, M. (1996a) 'Evaluation: The State of the Art', in *Proceedings of the Australasian Evaluation Society*; extracts in *Evaluation News and Comment* 5(2).

Scriven, M. (1996b) 'Types of Evaluation and Types of Evaluators', *Evaluation Practice* 17(2): 151–61.

Seefeldt, M. (1992) 'Conflicts in Conceptual Framework, Design, Audience Expectation, and Evaluator Style'. Paper presented at the Annual Meeting of the American Evaluation Association. 4–7 November, Seattle, WA.

Smith, M. F. (1988) 'Evaluation Utilization Revisited', *New Directions for Program Evaluation* 39: 7–20.

Stevens, C. T. and M. Dial (1994) 'What Constitutes Misuse?' *New Directions for Program Evaluation* 64: 3–14.

Stufflebeam, D. L. (1994) 'Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where it Needs to Go', *Evaluation Practice* 15(3): 321–38.

Weiss, C. H. (1998) 'Have we Learned Anything New About the Use of Evaluation?' *American Journal of Evaluation* 19(1): 21–33.

RITA O'SULLIVAN is Associate Professor of Education at the University of NC at Chapel Hill and Director of Evaluation, Assessment, and Policy Connections (EvAP). She has provided consultation on dozens of evaluation projects and

386

served as Secretary/Treasurer of the American Evaluation Association from 1992–1997. Please send correspondence to: Evaluation, Assessment, & Policy Connections (EvAP), School of Education, University of NC–Chapel Hill, Chapel Hill, NC 27599, USA. [email: ritao@email.unc.edu]

ANNE D'AGOSTINO is Director of Evaluation Services at Evaluation, Assessment, and Policy Connections located within the School of Education at the University of NC at Chapel Hill. Dr D'Agostino has worked on a variety of evaluation projects, including NC Smart Start, NC Charter Schools, AmeriCorps' Project Star, National Schools Attuned, and W.K. Kellogg's Pre-Birth Through Age Three Initiative. Please send correspondence to: 105A Peabody Hall, School of Education, University of NC–Chapel Hill; CB# 3500, Chapel Hill, NC 27599-3500, USA. [email: anned@unc.edu]

387