

Evaluating Short-Term Training Programs

A Practical Approach

Kent J. Sheets, Ph.D.

*Department of Family Practice
University of Michigan Medical School
Ann Arbor, Michigan 48109*

Abstract. Short-term training programs are frequently used to implement faculty development activities in post-secondary education. However, rarely are these programs systematically evaluated to assess their impact on program participants. This article describes an evaluation framework that was designed, developed, and implemented as an evaluation study of an existing faculty development program for family physicians. The study demonstrated the effectiveness of the faculty development program and a metaevaluation of the evaluation study identified the strength and weaknesses of the evaluation framework. The original framework was modified according to the metaevaluation results and presented for possible adoption by other instructional developers and evaluation.

Introduction

Short-term training programs are conducted regularly by schools, corporations, hospitals, businesses, churches, the military, and others. In post-secondary education, short-term training programs are frequently used in faculty development programs directed toward the improvement of instruction and teaching. This article describes an evaluation framework designed for use with short-term training programs.

Considerable time, effort, and resources have been expended on the design and implementation of short-term training programs, but little is known about their impact because they rarely are systematically evaluated (Centra, 1976; Gaff, 1979; and Littlefield, Hendricson, Kleffner and Burns, 1979). A review of the research on improving college teaching reported more studies than expected, but of lower

quality than hoped (Levinson and Menges, 1979). A subsequent review reported virtually no adequate studies of workshop impact (Menges and Levinson-Rose, 1980).

If indeed the major objective of faculty development programs is to change instructor behavior (Davis, 1979), the evidence supporting successful change is mostly based on participant satisfaction measures (Caldwell, 1981; Donnelly, Ware, Wolkon and Naftulin, 1972; and Stephens, 1981). The problem is that faculty give a generally positive assessment of their experiences, but do not teach differently (Gaff and Morstain, 1978). In addition, most evaluation efforts stop when the program ends and there are few attempts at follow-up observation of participants' behavior in actual practice.

In light of these reports, serious attempts should be undertaken to evaluate the impact of short-term training used within faculty development programs. These efforts should be designed to provide more rigorous data than the mere tabulation of participant opinions. Such an evaluation study was conducted during 1981 and 1982 at Michigan State University as part of the author's doctoral dissertation. This attempt to rigorously evaluate the impact of a faculty development program is detailed in this paper.

The evaluation approach used during the study, its strengths and weaknesses, and its applicability to other faculty development programs are presented and discussed. Although the approach was used to evaluate a single program, a number of issues are discussed based on that single evaluation study, including recommendations for further use.

The Program

The Family Medicine Faculty Development Program (FMFDP) was conducted by the Office of Medical Education Research and Development

(OMERAD) at Michigan State University (MSU). The FMFDP began operation in July 1978 with support from the Bureau of Health Manpower, United States Public Health Service. It had two major objectives. One was to identify and train new physician teaching faculty for family medicine training programs; the other was to help current family medicine faculty develop or refine their teaching skills. One component of the FMFDP was a teaching fellowship offered to allopathic (M.D.) and osteopathic (D.O.) physicians who had completed or were near completion of a family medicine residency program and to family medicine physicians with one year or less of academic teaching experience. The fellowship experience began in September 1981 as a two week session at MSU. The fellows returned in January, March, and May for three additional one week sessions. The subject matter presented in these sessions focused on teaching principles and techniques used in medical schools and residency training programs. The content was presented in a variety of instructional formats including workshops, seminars, and simulations. Sample topics presented during the sessions included: elements of group development, presentation skills, principles of learning and motivation, and clinical teaching.

Fourteen fellows (thirteen males and one female) participated in the program; four were D.O.'s and ten were M.D.'s. Each fellow received a stipend to defray program participation costs.

An Evaluation Framework

The evaluation framework was designed and developed to provide a mechanism for evaluating the impact of short-term training programs on program participants. It is a set of conceptual components and guidelines based on review of the literature. Factors considered during the review and the

Table 1
Evaluation Questions Asked During
Evaluation of the FMFDP

1. How satisfied were participants in the September session with the content, instructors, and activities of that session?	7. How much additional study of the content of the session was undertaken by the participants between the end of the session and the administration of the delayed posttest?	13. How did the participants rate their expected performance in a repeat of the series of three simulations or presentations?
2. How satisfied were the program directors with the content, participants, and activities of the September session?	8. How did the participants perceive their won expertise in each of the content areas of the September session both before and after the session?	14. How well did the participants utilize the content of the session related specifically to presentation skills in two different presentations?
3. How satisfied with the program were the supervisors of the participants?	9. What types of skills or techniques did the participants use following the completion of the September session?	15. How did the program directors perceive the participants' presentation skills?
4. What was the participants' level of cognitive knowledge of the September session at the beginning of the session?	10. What types of skills or techniques did the participants expect to use in the next six months?	16. Did the supervisors perceive any change in the participants' teaching due to the session?
5. Was there a significant retention of the cognitive knowledge of the content of the session by the participants six months following the completion of the session?	11. How did the supervisors perceive the participants' ability to apply the content of the session?	17. Did the participants perceive any change in their own role or function in their home institutions due to the session?
6. Was there a significant retention of the cognitive knowledge of the content of the session by the participants six months following the completion of the session?	12. How did the participants rate their own performance in a series of three completed simulations or presentations?	18. Did the supervisors perceive any change in the participants' role or functions in their home institutions due to the session?

framework's design included the importance of assessing the program impact, the focus on providing information to decision makers, and the need for flexibility. One assumption was that evaluators would not have access to participants before training starts and, therefore, data could be gathered only during and after the program.

There are five major components of the evaluation framework: type of data collected, who or what is assessed, data sources, data gathering methods, and questions.

Type of Data Gathered

The three types of data to be collected when using the framework are:

1. Reaction (satisfaction) data.
2. Cognitive (learning) data.
3. Behavioral (performance) data.

Most research addresses only reaction data with some additional self-reporting or cognitive or behavioral change, but objective measures of cognitive or behavioral change are seldom used.

Who or What is Assessed

One of the purposes of the evaluation framework is to assess the impact of a short-term training program on its participants. This component is concerned with identifying who or what is assessed in order to determine program impact. For example, the program content is assessed, as are the participants and the program faculty.

Sources of Data

There are numerous potential data sources and, ideally, all those in a position to comment on participant changes should be considered. The participants, program faculty, and the participants' supervisors should at least be consulted. Other possible data sources include the participants' subordinates, peers, students, clients, family members, or others with whom the participants interact. If the program teaches the creation of certain products or materials, it is also possible to use those developed by the participants as data sources. Multiple sources of information are recommended when conducting evaluation studies (Cronbach, et.al., 1980; Patton, 1980).

Methods of Gathering Data

Data gathering methods may be quantitative, qualitative, or both and need not be the same for each data source. Appropriate methods have been suggested by several authors (Baron & Baron, 1980; Bryk, 1978; Cronbach, et. al., 1980; Patton, 1980; and Posavac & Carey, 1980) and include, but are not limited to, the following: interviews, questionnaires, tests, direct observation (live, videotapes, films, audiotapes), and participant and staff self-reports.

Evaluation Questions

General evaluation questions are part of the framework as they are in other

evaluation approaches (e.g., Grotelueschen, 1980). The eighteen evaluation questions developed for the evaluation of the FMFDP are presented in Table 1.

The five components of the evaluation framework form a grid, Figure 1, designed for use as a guide when developing specific evaluation plans. Information that is placed in the cells of the grid becomes the basis for evaluation plans like the one used in this study.

The FMFDP and the fourteen participants were assessed during the study. Data sources included the participants, their supervisors, and the directors of the FMFDP. Data gathering methods included questionnaires, interviews, debriefings, tests, and videotapes. The methods addressed the evaluation questions in Table 1.

Instruments

Several different instruments were used during the evaluation study. Some of these instruments were already being used by the FMFDP staff. Other instruments were developed and used for the first time during the study. The five types of instruments included: End-of-Week Evaluation Forms, Pretest, Posttest, Delayed Posttest, Videotape Rating Scale, Interview Protocols, and Final Debriefing Questionnaire.

End-of-Week Evaluation Forms

These were designed and administered

Figure 1
Evaluation Framework

Type of Data	Assessment of Data	Source of Data	Method of Gathering Data	Evaluation Questions
Reaction (Satisfaction)				
Cognitive (Learning)				
Behavioral (Performance)				

by program directors at the end of both the first and second weeks. They focused on logistics, instructors, facilities, curriculum, and other programmatic elements and measured participant satisfaction and their self-reports of competence on selected topics taught during the session. Likert-scale and open-ended items were included on these instruments.

Pretest, Posttest, Delayed Posttest

A pretest, posttest, and delayed posttest were administered to assess entry level cognitive knowledge, changes in this knowledge after the session, and its retention six months later. The tests were developed using items submitted by FMFDP faculty. Only the session segments directly concerned with skills, techniques, and theories related to teaching and learning were tested. The same 40-item short-answer essay test was administered as a pretest, posttest, and delayed posttest. All fourteen fellows completed the tests.

All three sets of tests were scored independently by trained raters. The scores showed sufficient inter-rater reliability (pretest: .77; posttest: .83; and delayed posttest: .81) that ratings could be averaged into a single score for each participant.

Videotape Rating Scale

The participants were required to give several presentations to the program directors and fellows on skills taught during the session. These presentations were videotaped allowing participants to evaluate themselves. A 16-item rating scale was developed for two of the presentations given by each participant. Each item was rated using a Likert-type scale.

Two trained raters scored the tapes according to criteria established during the training session. Due to unequal

length, only the first ten minutes of each presentation were rated. The inter-rater reliability coefficients were not as high as those coefficients obtained from the written tests (January presentation: .48; May presentation: .59), but were still of sufficient reliability to be averaged.

Interview Protocols

To gather additional information a series of interview protocols was developed. Participants, their supervisors, and two program directors were interviewed. Open-ended and Likert-type items were used on the protocols.

Final Debriefing Questionnaire

As part of the regular evaluation system of the FMFDP, a final debriefing was conducted by the program directors. In addition to the open-ended written questions included on the questionnaire administered during this debriefing session, a question concerning the evaluation activities was inserted to ascertain the fellows' reactions to the various evaluation procedures. After the fellows completed the questionnaire, a discussion was conducted by the program directors.

Analysis Procedures

Quantitative data analysis was required for the test data, rating of presentations, and quantitative items on the interview protocols and end-of-week evaluations. The rest of the data was qualitative, notably the open-ended comments made during interviews, end-of-week evaluations, and the final debriefing.

For those items which were quantitative, frequencies were determined and where appropriate, descriptive statistics were calculated. For the ratings of the two videotaped presentations, descriptive statistics were computed for the individual items, each rater, and the

average of the two raters. Descriptive statistics were computed for the three tests and a univariate analysis of variance (ANOVA) was conducted comparing the results of both the pretest and posttest with those of the delayed posttest.

Qualitative analysis of the open-ended comments was conducted by reading and grouping similar comments for each question and instrument. Comments made two or more times were reported.

Metaevaluation of the Evaluation Study

A metaevaluation was designed and conducted to answer the following questions about the evaluation:

1. What specific problems were encountered during the evaluation study?
2. Was the evaluation framework practical in its use of resources?
3. Was the evaluation framework useful in providing information to the decision makers?
4. Were the methods and instruments used during the evaluation study technically adequate?
5. Were the methods and instruments used during the evaluation study conducted in an ethical manner?

The metaevaluation examined the quality of the process and product of the evaluation study. Three different metaevaluation activities were conducted, self-report, program director interview, and analysis of evaluation procedures.

The evaluator prepared a self-report that addressed the problems encountered, including difficulties related to the development, administration, scoring, and analysis of evaluation instruments. The self-report provided a problem overview and answered the first question.

The second stage of the metaevaluation was approached in two steps. The first step was to identify evaluation standards related to each of the metaevaluation questions. The source of the standards used was the 1981 publication, *Standards for Evaluations of Educational Programs, Projects, and Materials*. After one or more standards were identified for each metaevaluation question, specific questions were formulated to address the concepts within each of the standards. The specific questions were then arranged as an open-ended questionnaire.

Two directors of the FMFDP were in-

interviewed by telephone to answer the final four questions. Additional questions not specifically related to the questions, yet concerned with the effectiveness of the evaluation, were also asked during the interview.

The final stage of the metaevaluation consisted of an analysis of the evaluation procedures used. With the assistance of the program directors and another individual experienced in conducting and evaluating short-term training programs, the evaluator identified five effectiveness factors (Table 2).

Each of the procedures was rated by the evaluator and program directors independently in relation to each of the five factors. A scale from one to three was used to rate each procedure, with one meaning low, two medium, and three high. The procedures were rated according to their actual performance during the study.

Table 2
Evaluation Factors

Factor	Description
1. Direct utility of information for decision making	Ease of applying information to decision making.
2. Time efficiency	Time required for developing, administering, scoring, and analyzing procedures and data.
3. Resource efficiency	Personnel, materials, and equipment required for developing, administering, scoring, and analyzing procedures and data.
4. Credibility	Reliability and validity of procedures and data.
5. Data manageability	Ease of data representation, summarizing, and analysis.

Results

The results of the evaluation study were extremely encouraging and indicated that the two-week session was successful and had an impact. The strengths and weaknesses of the content, instructors, and activities were identified and mostly agreed upon by participants and program directors. The supervisor reactions were also favorable although their comments were limited. Moderate to high satisfaction was reported by the fellows, program directors, and supervisors.

Among the results of the cognitive measures, there was an apparent relationship between participant test performance, self-report of expertise, and self-report of handout use. There was also evidence that suggested this relationship extended to the degree of satisfaction with the topics.

There was a meaningful change in participant cognitive knowledge. The ANOVA results indicated that this change occurred during the first two weeks and was stable six months later.

Behavioral data indicated the participants were using many skills and techniques taught. Support for this finding was provided by data originating from all three sources of information and from the different data gathering methods.

Across the three types of data there were some notable trends that deserve identification. The practical, concrete, skill-oriented presentations were most

highly enjoyed and the content most used. In contrast, a session on learning theory was viewed as theoretical and irrelevant.

In most circumstances, the information provided by participants was consistent to the point of redundancy. There was sufficient reason to believe the data provided by the participants on the reaction measure and self-reports were reliable because they were supported by the supervisors' and program directors' data. The data provided by the participants, program directors, and supervisors were also in agreement in most circumstances.

There was an apparent relationship between the performance of the participants on the three cognitive tests and the videotape ratings. Those individuals ranked at the top of the test results were likely to be highly ranked for the videotaped presentations. Similar results were found for those individuals who scored poorly on the cognitive tests. Scores and ranks for each of the fellows are presented in Table 3.

Results of Metaevaluation

Table 4 contains the five metaevaluation questions and a brief answer to each.

The additional data gathered during the metaevaluation helped identify those procedures and data that were most useful to the program directors and the three preferred procedures were the end-of-week evaluations, final debriefings,

and videotape ratings. The cognitive tests were least preferred due to their low validity and poor ability to discriminate. The interviews with the fellows and the supervisors were also rated low, primarily due to the time and cost associated with conducting the interviews.

Program directors identified behavioral data as being more useful to them than cognitive and reaction data. However, the program directors indicated that a strength of the evaluation study was that reaction, cognitive, and behavioral data were all collected using several methods. Much of the information collected was redundant and, at times, overwhelming.

The redundancy of information was noteworthy in that the reliability and validity of the data gathered during the end-of-week evaluations were strengthened by the data collected by the evaluator during later interviews and were also notable in terms of their relationship with other study outcomes. There was a relationship between participant reactions to topics and presentations and their subsequent behavior and performance.

The terms "relevant" and "applicable" were frequently used by the fellows in their comments. The presentations the fellows enjoyed and perceived as most relevant or applicable to their present or future activities were also those on which they had the highest cognitive test scores. The reported use of handouts

Table 3
Composite Scores and Ranks on Tests and Videotapes

Fellow	Pretest*	Rank	Posttest 1*	Rank	Posttest 2*	Rank	Video 1**	Rank	Video 2**	Rank
1	23.5	14	56.5	11	34.0	14	21.0	13	40.0	10
2	41.0	10	46.5	14	46.0	13	41.0	8	52.0	1
3	48.0	5	69.5	8	61.5	7	48.5	2	41.5	8
4	39.0	13	73.0	6	75.0	1	41.5	6	34.5	12
5	52.0	2	81.5	2	73.0	2	41.5	6	52.0	1
6	39.5	12	79.5	3	63.0	6	32.5	12	25.0	14
7	45.5	8	57.0	12	56.5	11	39.5	10	29.5	13
8	51.5	3	78.5	4	61.0	8	36.0	11	42.5	7
9	46.0	7	72.0	7	66.0	4	48.0	3	44.0	6
10	41.0	10	68.0	9	61.0	8	45.5	4	41.0	9
11	59.0	1	83.5	1	65.5	5	51.0	1	47.0	4
12	44.0	9	59.5	10	55.0	12	***	***	40.0	10
13	46.5	6	57.0	12	67.5	3	41.0	8	48.0	3
14	51.5	3	74.5	5	57.5	10	42.5	5	47.0	4

*120 points possible

**80 points possible

***did not participate

from these presentations and the reported use of skills and techniques taught were also highest.

Conclusions

The strengths and weaknesses of the evaluation framework for short-term training programs are summarized in Table 5.

Since a major focus of the evaluation framework is on determining impact, much data collection must be conducted after program completion. Data were collected as late as eight months follow-

ing the session. Although no timelines are prescribed by the framework, follow-up has to be delayed to allow participants time to integrate new skills and techniques into their daily routines. Final evaluation results and their use for making programmatic changes are delayed accordingly.

There is no conclusive evidence concerning the necessity of collecting all three types of data. While the reaction and behavioral data were easily used by the program directors, it was difficult for them to use the cognitive test data

because of its low validity and poor ability to discriminate, and its utility could not be determined.

The lack of explicit criteria for the cognitive test and videotape rating results constitutes a weakness in the evaluation when results are examined by individuals external to the program. Without explicit criteria and cutoff scores, percentage scores have little or no meaning to other individuals examining the results of the evaluation; a weakness in the procedures used, not in the evaluation framework.

Table 4
Summary of Responses to Metaevaluation

Metaevaluation Question	Response
1. What specific problems were encountered during the evaluation?	The major problems related to the collection of behavioral data and to the development, administration, and validation of the cognitive tests.
2. Was the evaluation framework practical in its use of resources?	Yes, the program directors felt justified in committing the resources required to conduct the field test.
3. Was the evaluation framework useful in providing information to the decision-makers?	Yes, the data were comprehensive and confirmed the program directors' subjective assessments of the program's quality and impact.
4. Were the methods and instruments used during the field test of the evaluation framework technically adequate?	Yes, with the exception of the cognitive tests. There were reasons to question the validity of the cognitive test results.
5. Were the methods and instruments used during the field test of the evaluation framework conducted in an ethical manner?	Yes, the methods and instruments were conducted in an ethical manner. The evaluator was candid in his interactions with people during the field test.

Table 5
Strengths and Weaknesses of the Evaluation Framework

Strength	Weakness
Holistic, systems approach.	Time required from implementation of evaluation framework to final results.
Focus on outcomes and impact.	No immediate formative evaluation feedback possible.
Decision-oriented approach.	Uncertainty whether all three types of data must be collected at all times (procedure).
Prescriptive, but options allowed when selecting information sources, data gathering methods, and evaluation questions.	Excessive amount of information collected (procedure).
May be used for formative or summative evaluation purposes.	No explicit criteria delineated; left to discretion of program directors (procedure).
Allows use of qualitative and quantitative methods.	Questionable validity of cognitive test (procedure).
Collects three types of data to provide comprehensive information to decision makers.	
Collects redundant information (procedure).	

The framework was designed specifically for one program type, short-term training, and was used with a program for training family physicians interested in academic medicine. To fully assess the value of the evaluation framework, it should be used with short-term training programs of varying length, content, and audience.

The evaluation framework might be as effective or more effective with programs that are not short-term. No conclusions can be made based solely on the results of the evaluation and metaevaluation.

Implications for Educational Practice

The ultimate value of this study is based on the products, processes, and procedures which can be used by evaluators, educators, trainers, administrators, and other individuals responsible for short-term training. The procedures and instruments used during this study could be adopted and modified for use in a number of situations. The conceptual approach to short-term training presented throughout the paper could be useful to practitioners. The primary purpose of the study was to develop an evaluation approach to assess the impact of short-term training programs. The evaluation framework was designed, developed, and field tested to serve that purpose. One particular short-term training program, the

September 1981 session of the FMFDP, was evaluated using an evaluation design based on the evaluation framework. The results of the evaluation and metaevaluation were presented with the expectation that the results would be of interest and value to potential users of the framework. Several conditions related to the FMFDP require discussion because these conditions limit the general applicability of the results of the evaluation. First, the group of fourteen fellows were paid participants using release time to participate in the fellowship. It is more often the case that the participants in short-term training programs or the participants' organizations pay to attend such programs. Frequently, participants must use weekends, evening hours, or vacation time to attend.

Second, the two week segment evaluated during the study was part of a longer, continuing relationship between the fellows and the program directors. This continuing relationship facilitated certain data collection activities that might have been more difficult otherwise.

The framework used to collect the evaluation data was designed eclectically, an approach recommended by Baron & Baron (1980), Patton (1980), and Steele (1973). The implication of their suggestions to follow an eclectic approach to evaluation was that the existing models were not functioning

satisfactorily. The framework was operationalized and the evaluation was successful. The question of whether the evaluation framework is superior to other approaches remains unanswered and will remain unanswered until the evaluation framework is used by others.

To fully assess the value of the evaluation framework, additional short-term training programs in other settings with different content, length, and populations must be evaluated. Only then can more definite conclusions be drawn. A revised matrix of the evaluation framework is outlined in Table 6. The revised matrix incorporates the major findings of this study into its design and offers a prescriptive version of the evaluation framework.

References

- Baron, J. B., & Baron, R. M. (1980). In search of standards. In R. Perloff & E. Perloff (Eds.), *Values, ethics, and standards in evaluation (New directions for program evaluation, Number 7)*. San Francisco: Jossey-Bass.
- Bryk, A. S. (1978). Evaluating program impact: A time to cast away stones, a time to gather stones together. In S. B. Anderson & C. D. Coles (Eds.), *New directions in program evaluation, Number 1*. San Francisco: Jossey-Bass.
- Caldwell, P. A. (1981). Preservice training for instructors of adults. In S. M. Grabowski & Associates (Eds.), *Preparing educators of adults*. San Francisco: Jossey-Bass.
- Centra, J. A. (1976). *Faculty development practices in U.S. Colleges and Universities*. Princeton: Educational Testing Service.

Table 6
Revised Matrix of the Evaluation Framework

Type of Data	Assessment of Data	Source of Data	Method of Gathering Data	Evaluation Questions
Reaction (Satisfaction)	STP	P	EOW, FD	How satisfied with the program were the participants?
Cognitive (Learning)	P	P	TESTS	How much of the content of the program did the participants learn and retain?
	P	P	S-R	How did the participants perceive their own learning and retention of the content of the program?
Behavioral (Performance)	P	P	DO, VT	How well did the participants apply to the content of the program in simulated or actual performance settings?
	P	P	S-R	How did the participants perceive their ability to apply the content of the program in simulated or actual performance settings?

Key:

- STP—Short-term training program
- EOW—End-of-week evaluations
- S-R—Self-reports
- P—Participants
- FD—Final debriefing
- DO—Direct observation
- VT—Videotape of participant performance in simulated or actual setting

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

Davis, R. H. (1979). A behavioral change model with implications for faculty development. *Higher Education, 8*, 123-140.

Donnelly, F. A., Ware, J. E., Wolkon, G. H., & Naftulin, D. H. (1972, March). Evaluation of weekend seminars for physicians. *Journal of Medical Education, 47*, 184-187.

Gaff, J. G. (1979). The United States of America—Toward the improvement of teaching. In D. C. B. Teather (Ed.), *Staff development in higher education: An international review and bibliography*. New York: Nichols Publishing.

Gaff, J. G., & Morstain, B. R. (1978). Evaluating the outcomes. In J. G. Gaff (Ed.), *Institutional renewal through the improvement of teaching: New directions for higher education, 6*, 77-83.

Groteleuschen, A. D. (1980). Program evaluation. In A. B. Knox & Associates (Eds.), *Developing, administering, and evaluating adult education*. San Francisco: Jossey-Bass.

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

Levinson, J. L., & Menges, R. J. (1979). *Improving college teaching: A critical review of research* (Occasional paper No. 13). Evanston, IL: Northwestern University, The Center for the Teaching Professions.

Littlefield, J. H., Hendricson, B., Kleffner, J. H., & Burns, G. (1979, April). Improving instructional skills of health science educators. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Menges, R. J., & Levinson-Rose, J. L. (1980). Why research tells us so little about interventions to improve college teaching. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills: Sage Publications.

Posavac, E. J., & Carey, R. G. (1980). *Program evaluation: Methods and case studies*. Englewood Cliffs, NJ: Prentice Hall.

Steele, S. M. (1973). *Contemporary approaches to program evaluation and their implications for disadvantaged adults*. Syracuse, NY: ERIC Clearinghouse on Adult Education.

Stephens, B. (1981). *A manual for clinical supervision and facilitative feedback*. Madison, WI: University of Wisconsin, Department of Family Medicine and Practice.