DOCUMENT RESUME

ED 167 929 CG 013 363

AUTHOR Mitchell, Anita M.
TITLE Counselor Program Effectiveness: Gathering, Using and Reporting the Evidence.
INSTITUTION ERIC Clearinghouse on Counseling and Personnel Services, Ann Arbor, Mich.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 79
GRANT 400-78-0005
NOTE 64p.
AVAILABLE FROM ERIC/CAPS, 2108 School of Education, The University of Michigan, Ann Arbor MI 48109 ($5.00)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Behavior Change; *Counseling Effectiveness; *Counseling Programs; Counselors; Data Collection; *Evaluation Methods; *Guidance Objectives; Management by Objectives; *Performance Criteria; *Program Evaluation
IDENTIFIERS Information Analysis Products

ABSTRACT

A procedure for determining the effectiveness of counseling programs is suggested in this monograph. Since the procedure is systematic and two-pronged--process and outcome evaluation--it provides information as to the degree of program effectiveness and/or ineffectiveness. A self-correcting system, the procedure's progressive steps illuminate possible errors or ambiguities in previous steps, making it possible to modify the procedures while the evaluation is in process and thus avoid costly problems. The paper also describes evaluation problems and suggests what to do when they arise, and presents a section of recommendations for counselors, supervisors, counselor educators, program developers, and educational researchers. (Author)

The work presented herein was performed pursuant to a contract from the
National Institute of Education
Department of Health, Education and Welfare
Under the direction of
ERIC COUNSELING AND PERSONNEL SERVICES CLEARINGHOUSE
Garry R. Walz, *Director*
Libby Benjamin, *Associate Director*
School of Education
The University of Michigan
Ann Arbor, Michigan

1979

COUNSELOR PROGRAM EFFECTIVENESS:

Gathering, Using and Reporting the Evidence

by
Anita M. Mitchell

Series Editors: Garry R. Walz and Libby Benjamin

Developed by
ERIC Counseling and Personnel Services Clearinghouse
The University of Michigan
Ann Arbor, Michigan

This publication was prepared pursuant to a contract with the National Institute of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent National Institute of Education position or policy.

ABOUT THE AUTHOR

Anita M. Mitchell is a senior research scientist, American Institutes for Research in the Behavioral Sciences; senior research scientist and member of the Board of Directors, Canadian Institute for Research; President, ALTERNATIVES, educational consultants; and associate, ALTERNATIVES ASSOCIATES, educational philosophers and consultants. She received her B.A. with honors from Maryville College, M.A. from St. Louis University, and Ed.D. in Psychology from the University of Southern California.

She is a Past President of the Association for Measurement and Evaluation in Guidance, of the California Association for Measurement and Evaluation in Guidance, and of the California Personnel and Guidance Association.

Most recent publications include *Social Learning and Career Decision Making*, with G.B. Jones and J.D. Krumboltz; *Career Development Needs of Seventeen-Year-Olds: How to Improve Career Development Programs*; "Facilitating Full Employment of Women Through Career Education," in *American Women Workers in a Full Employment Economy*; and "The Design, Development, and Evaluation of Systematic Guidance Programs," in *New Imperatives for Guidance*, the most recent publication of the ERIC Clearinghouse on Counseling and Personnel Services.

ABOUT THIS MONOGRAPH

This paper suggests a procedure for determining the effectiveness of counseling programs. Since the procedure is systematic and two-pronged--process and outcome evaluation--it provides information as to the degree of program effectiveness and/or ineffectiveness. A self-correcting system, the procedure's progressive steps illuminate possible errors or ambiguities in previous steps, making it possible to modify the procedures while the evaluation is in process and thus avoid costly problems.

The paper also describes evaluation problems and suggests what to do when they arise, and presents a section of recommendations for counselors, supervisors, counselor educators, program developers, and educational researchers.

TABLE OF CONTENTS

COUNSELOR PROGRAM EFFECTIVENESS: GATHERING, USING AND REPORTING THE EVIDENCE

by

Anita M. Mitchell

Why another publication on the effectiveness of counselor programs? Since Leon Lessinger popularized the term "Accountability," the literature has bulged with reports of the effectiveness of guidance and counseling programs, with models for determining program effectiveness and a plethora of evaluation designs and instruments. Categorizing these documents is difficult since no generally accepted description of the domain of counselor programs exists. As a result, we have an abundance of unrelated information, and we don't know what to do with it.

Most of the documents accessed by this author (both in a formal literature search and in many years of experience in the field) focus on program evaluation in general, or on guidance evaluation. The former are generally too broad to be applied easily to counseling programs, and the latter tend to dwell on single components of an unspecified guidance domain. The intention of this document is to reduce the guidance universe to a manageable size, and to present practical, useful guidelines for practitioners to follow as they seek to pursue that elusive construct known as counselor program effectiveness.

This document differs from most other publications in that it (1) explores the question of whether or not to evaluate counselor programs, showing that evaluation is not always necessary or even desirable; and (2) addresses the changing emphases in counselor programs, and the need for different types of evaluation designs, instruments, and procedures.

Still another dimension that may be somewhat new to the reader is an emphasis on evaluation as being within the range of counselor competencies, as being nonthreatening, and as being motivating and ego-reinforcing. Evaluating the effectiveness of counselor programs requires a two-pronged evaluation strategy--process and outcome evaluation. Outcome evaluation provides evidence of the degree of program effectiveness; process evaluation provides information about why certain procedures and materials are or are not effective.

Counseling Programs and Program Components

Definitions

Perhaps a glossary would do the job, but frequently readers fail to read the glossary first or refer to it to clarify the meaning of a term. A few definitions here should establish perspectives consistent with the purposes of this paper.

Counselor. The term counselor may be left broad, as the measures of effectiveness of programs may be applied by anyone who is performing

counseling functions, regardless of degrees, credentials, rank or training. The counselor may be a dean, a head counselor, a counselor assigned to a specific block of students, an on-call counselor, a peer counselor, or a paraprofessional.

Guidance. The term "guidance" is used to include all legitimate tasks assigned to/performed by counselors, including testing, parent conferencing, consultation with teachers, staff development, analysis of learning environments, and individual and group counseling. Record-keeping and information-giving are not included, as these are not (in this author's opinion) tasks requiring certification.

The first section delimits the guidance domain. If you are sensitive to writer inconsistencies, you will have noticed that the "counselor program" designation in the title changed, by the second paragraph, to "counseling program." This is not a careless error; it is deliberate. The term "counseling program" will be used in this document to define the domain to which this discussion is limited: the effectiveness of programs in which the counselor is engaged in either individual or group counseling, or both. There are two reasons for this limitation:

1. After a deluge of admonitions to counselors that they need to expand their roles and become managers, models, and consultants

(which is true), there is finally recognition that the counseling function will always perdure, regardless of the changing emphasis of the counselor programs;

2. There is precious little written to help counselors focus on the effectiveness of their counseling efforts.

However, the guidelines presented, while addressing counseling programs, are easily modified to apply to other counseling functions such as parent conferencing, staff development, and teacher consultation.

Program. Another term that needs to be defined for purposes of this paper is "program." The discussion of approaches to measuring effectiveness will apply to a single activity (a one-time counseling session or successive sessions all aimed at the same outcome for the same individual or group); to a group of independent activities all having the same desired student outcome(s), e.g., all of a counselor s counseling sessions on underachievement; to all group counseling sessions; to all individual counseling sessions; to counseling engaged in by a single counselor; or to all counseling engaged in by a total counseling staff. In other words, the total school or agency counseling program or any segment of that program can be considered a program when applying measures of effectiveness. The reason for this is that the need for evaluation may be limited to a single incident, to a series of counseling sessions, or to the cumulative effects of all counselors' counseling

efforts. Deciding which program or program component to evaluate is addressed in another section.

Effectiveness vs. Evaluation. "Effectiveness" and "evaluation" also need to be defined. Effectiveness is a measure of the degree to which a program accomplishes what it intended to accomplish. This means that intention must be made visible--a condition that frequently does not exist in counseling programs. The section on practical purposes of evaluation explores this concept further. Evaluation is the process of defining the target behavior (outcomes desired from the counseling), identifying behavioral indicators from which achievement of the outcomes can be inferred, selecting/administering instruments for gathering data about the achievement of the outcomes, and analyzing and reporting the data. In other words, evaluation is the process followed to establish the effectiveness of the program. A sound evaluation strategy addresses both process and outcome, so the evaluator can judge what caused the success or nonsuccess (effectiveness or noneffectiveness) of the program.

Traditional Counseling Program Emphases and Components

Traditionally, counseling has been largely problem-centered. During the last several years school counselors in workshops throughout the country have identified the major content areas

of their counseling. These can be categorized into 11 areas:

- Classroom behavior
- Failure to complete classwork and/or home assignments
- Failure to achieve
- Problems on campus
- Problems with peers
- Problems with parents
- Personal problems (other than drug abuse)
- Problems with drugs
- Help with educational plans
- Help with career plans
- Placement in special programs

Only two of these emphases--help with educational plans and help with career plans--are developmental rather than problem-oriented. Most of the counselors surveyed approached each counseling session without structure, and with desired outcomes only vaguely defined in terms of helping their students cope. Frequently the tools for coping were explained to the students, and it was assumed that they could master and use these tools without guided practice. This is contrary to what we know about learning, and about application of learning in performance.

Since the desired outcomes were not defined, counselors had no basis for determining the effectiveness of their counseling, except through the counselees' voluntary feedback that they felt better and/or were satisfied with the counseling session(s). Evaluation

of counseling programs was characterized by a preponderance of impressionistic data, tallies of counselor-client contacts, and staff/activity lists. These constituted summaries of activities but provided no evidence of effectiveness. Because there was no way to know what was done well, there was no way to refine techniques.

The 11 counseling areas listed above relate to school counseling, and apply to both elementary and secondary levels. Counseling in institutions of higher education generally has focused on educational and career planning, financial aid, housing, and personal and family problems. Counseling in employment and rehabilitation agencies typically has been limited to similar concerns--helping with educational and career planning, and teaching clients how to cope with problems they encounter. Evaluation frequently has been in terms of numbers of clients seen, although the agencies have been more likely than the schools to assess the impact of their services by determining how many of their counselees were successfully placed in jobs, succeeded in jobs, retained jobs, and the like. Little has been done, however, to determine the proportion that counseling contributed to the measured success.

Shift in Evaluation Emphasis

Two things which happened almost simultaneously contributed to a shift in evaluation strategies--a shift away from tallies of activities and toward measures of effectiveness--and presaged the

types of evaluation approaches that are featured in this paper. These simultaneous happenings were: (1) reduced revenues and increased costs, which forced agencies to search for programs/personnel/materials/services that could be eliminated without damaging the educational program; and (2) the involvement of lay community members, many of whom were business- and industry-oriented, in the operations of educational and other governmental agency programs. The struggle for survival and the frenetic activities of some specialists threatened with loss of their jobs created a fertile field for the advocates of accountability, management by objectives, results systems management, PPBS (Planning, Programming, Budgeting Systems), and various evaluation models. Although this period was characterized by confusion, anger, disenchantment, fear, and trial by blunder, it did sensitize specialists (particularly counselors) to the fact that there is need for every employee to be able to display the effects of his/her employment.

The shift in evaluation emphasis forced counselors to specify what changes clients could be expected to make as a result of their counseling, which, in turn, caused them to take a hard look at why they were doing certain things--at why some methods worked and others didn't. This resulted in a shift in program emphasis. Whereas the same 11 content areas continued to dominate most counseling sessions, emphasis changed from tabulation of numbers of contacts to documentation of actual changes in client behavior.

An example of this change in emphasis may be found in an expansion of the 11 counseling content areas listed earlier. Basic student performance objectives for current, ongoing counseling programs have been adopted by counselors in many school districts as a way of defining what they hoped to accomplish with and for students, and of assessing the degree to which they succeeded. Listed below are some examples of student performance objectives. These are not intended as models to be followed or as conditions to be imposed in any counseling session dealing with the specific behavior. Rather, these illustrations should be viewed as a kind of mental template for counselors as they enter a counseling session. In no case should the counselor decide ahead of time that these are the behaviors that will result from the counseling session(s); the counselor has no right to make that determination in advance. Only the broadest kinds of directions should be set in advance--to display to the client some possible actions, and to help the client select from alternative approaches the most desirable means of solving the problem or obtaining the help sought. The examples follow:

- Pupils referred because of classroom behavior will:
  a. identify at least two ways in which they contributed to the problem.
  b. identify at least one interaction variable between themselves and another person(s) in the class (peer or teacher) that maintains the problem.

c. develop a plan for alleviating the problem, specifying at least three possible incremental actions.

d. select the single action they might take which would be most likely to reduce the problem.

- Pupils referred for failure to complete classwork and/or home assignments will:

  a. identify the principal reasons for their failure to participate at the level of teacher expectation.

  b. develop a plan for handling each reason (e.g., Reason: too tired to complete homework. Plan: design a realistic school/work/study/play schedule).

  c. decide which of the plans identified in (b) would be most likely to help them complete classwork and/or home assignments.

- Pupils referred for failure to achieve will:

  a. identify skills needed for achievement in the class.

  b. identify the level(s) of their own skills.

  c. decide what should be their performance level in the class.

  d. if level identified in (c) is greater than that currently being achieved, develop a plan for reaching the expected level of achievement.

  e. specify the part(s) of the plan they are willing to implement immediately.

- Pupils referred for problems on campus will:

  a., b., c., d. as in first example above.

- Pupils referred for problems with peers will:

  a., b., c., and d. as in first example above.

- Pupils referred for problems with parents will:

  a. identify conflict areas.

  b. specify differences between their values and their parents' values in these conflict areas.

  c. develop a plan for resolving value differences (e.g., compromise, informed acquiescence, mutual respect).

  d. decide which conflict area is most amenable to change.

  e. select plan for resolving conflict area identified in (d).

- Pupils referred for personal problems (other than drug abuse) will:

  a. identify the implication of the problem for themselves and for others (e.g., Pregnancy. How does it affect me? Who else does it affect? How?).

  b. identify all reasonable alternative strategies for solving the problem.

  c. develop a plan for implementing those strategies most apt to be effective in solving or lessening the problem for all persons specified in (a).

- Pupils referred for problems with drugs will:

  a. identify their reasons for using/pushing drugs.

b. specify all possible consequences (good and bad) of using particular drugs.

c. identify reasons why the "good consequences" are important to them.

d. for each "good consequence" ("It makes me feel good," "It gives me power," etc.) specify at least one other means of accomplishing that consequence.

e. develop a plan for progressive elimination of their drug problem.

- Pupils referred for help with educational plans will:

a. identify their strengths and weaknesses in abilities, aptitudes, achievement, interests, and values; and relate this information to occupational fields and levels.

b., c., d., and e. as in example above.

- Pupils referred for placement in special programs will:

a. identify their particular needs.

b. specify which of these needs cannot be met in the regular classroom.

c. identify structure/components/content of the special program that will help meet these needs.

d. develop a plan for making best use of special class placement.

This expansion of objectives for the 11 traditional content areas of counseling students is intended to suggest an approach to

counseling which focuses on outcomes. It can be adapted to any content area in any client setting. Importance lies in the fact that a series of possible outcomes is presented to the client by the counselor (unless the counselee is already able to specify his/her desired outcomes), and together the client and counselor determine which objective(s) the client feels will help him/her and which s/he is willing to work toward. The counseling session is a collaborative venture, with both counselor and client openly exploring possibilities for solution of the problem. Once counselor and client have agreed upon the objectives of the counseling session(s), it is easy to verify whether those objectives have been met; this verification is the measure of the effectiveness of the counseling.

Besides the new outcomes-oriented emphasis reflected in the above examples, a twelfth focus is emerging: assisting clients to master the various developmental tasks with which they are confronted--physical, intellectual, emotional/social. Counselors must, of course, continue to help those who have experienced failure in one or more developmental tasks, but it is important to work toward preventing such failures by assisting all clients to approach and master each task. This new emphasis on self-development is reflected in the following example of expanded objectives in a relatively new content area:

- Pupils referred for help with self-development will:
  a. identify and explain their current level of ego development.
  b. identify personal and/or environmental factors inhibiting their self-development.
  c. develop a plan for eliminating the inhibiting factors.

This approach to specifying with the client the expectations from the counseling session(s) obviously facilitates evaluation. In an individual counseling session, attainment of each or all of the objectives can be verified through observation of the client's participation in the discussion and planning, through analysis of the client's oral or written account of his/her perception of the problem and the plan, and through followup of the client's actions after the counseling session. Probably the objectives will be attained over a series of sessions; assessment of results is straightforward.

Perhaps a counselor or a counseling staff will wish to determine the effectiveness of counseling efforts across all drug abuse cases, or all cases of underachievement, or all cases of peer problems. By gathering data on the results achieved in each case, or in each of a random sample of cases, they can establish the effectiveness of their total efforts in the specific area.

Most measures of effectiveness will compare clients' behavior before and after counseling. Some will also compare behavior of

counseled vs. noncounseled populations. For example, in an investigation of the effects of a university student counseling service, the preferences and expectations of 100 student clients before their initial intake session were compared with their post-counseling preferences and expectations (Dreman, 1977). Results were compared with those of an earlier study on nonclients. As predicted, preferences and expectations were more congruent in counseled clients.

In another study Banathy (1974) evaluated the San Jose State University (California) career counseling training project. The evaluation was post-facto and summative (outcome-oriented) based on information derived from logical analysis and empirical data (questionnaire and interview responses) developed during the evaluation. Besides providing evidence that the stated objectives of the project were accomplished with high quality, that the investment was well placed, and that subsequent additional development and dissemination were justified, the results showed that paraprofessionals can perform very effectively as an integral part of an educational career guidance program. This is an example of how evaluation can furnish data for decisions (to expand and disseminate the program), as well as information about variables (effect of including paraprofessionals) which will affect further decisions about program modifications and/or extensions.

The change in emphasis in evaluating programs is further reflected in two additional publications: Alkin and Kosekoff (1973) and Mitchell (1978). Both documents stress the concept of evaluation

as furnishing data for decisions. Alkin and Kosekoff state that one of the initial elements in performing an evaluation is framing the decision context. Dimensions of this context are identification of the decision makers, explicit decision-making purposes, implicit decision-maker motives, developmental stage of programs, and the program's socio-political setting. The Mitchell document presents six models for evaluating different types of career education activities: supplementary activities, curriculum infusion, facilities, product development, staff development, and indirect interventions. In each case, typical decisions are suggested, and procedures for gathering the information necessary to make these decisions are addressed.

Practical Purposes of Evaluating Counseling Programs

There is no point in committing resources to an evaluation effort unless that effort (1) will be pursued with sufficient rigor to make the results trustworthy, and (2) will result in the findings being used for one or more important purposes. Important purposes include measuring program impact to judge effectiveness of one program or of alternate or competing programs; gathering data for decisions: whether to change, improve, abandon, expand, or proceed to more advanced goals; collecting evidence of the relative efficacy of specific procedures and/or materials; collecting evidence to rally support for a program; and identification of materials and

procedures in need of revision.

In order to determine which of these purposes should be served in determining the effectiveness of a program, it is necessary to identify and understand the program's various publics. These publics will always include the clients and all staff members who are affected by the program (e.g., a group counseling program which pulls students out of classes once a week affects all teachers whose classes are missed). Other publics are the agency supervisors and administrators, governing boards, business and industry, education or training institutions into which the clients feed, parents, the general public. Once the relevant publics have been identified, it is wise to establish a panel of advisors, members of which are truly representative of the philosophy/interests/concerns of their constituent groups. These advisors will help determine whether or not a program should be evaluated, and if it should, which purposes will be served by the evaluation. Discussions with such a group may reveal that no evaluation is necessary. For instance, if the advisory group members are satisfied that the counseling program is effective, if they have no questions concerning the program, and if they cannot identify any purpose for which evaluation evidence is needed, it would be foolish to commit resources to an evaluation effort. Such effort should be reserved for other counseling programs in the institution which do not enjoy such complete support. The task will then be to identify those programs that may have been under attack by various factions,

those that do not receive enthusiastic response from clients or from staff, those that appear to be losing their effectiveness, those that are particularly costly, those that are particularly time consuming, those that just don't "feel" right. Such programs need proof of their effectiveness. The next step will be to decide on the purpose(s) of the evaluation, and proceed! Unfortunately, many counselors and counseling staffs don't want to rock the boat, so they never subject their programs to such scrutiny. As a result, when budget cuts are pending, they are empty-handed, with no evidence of the effectiveness of their efforts.

A new publication, The Profession and Practice of Program Evaluation (Anderson & Ball, 1978), presents a comprehensive discussion of evaluation practices, ethics and values in evaluation, and the future of program evaluation. The reader is referred to this volume for further pursuit of some of the ideas presented in this paper.

A few examples of evaluations that address specific purposes are presented here to clarify the issue. In an analysis of patient views toward very brief interventions (counseling consisting of three or fewer sessions), Dorosin (Dorosin, et al., 1976) found that the original expectations of the patient are met in a relatively brief contact, and that counseling termination should not be viewed as a failure or patient dropout. Had Dorosin not conducted this evaluation, he would not have known the causes of termination of counseling, and might unnecessarily have extended his counseling efforts.

The purpose of another study (Hamachek) was to determine the effectiveness of a one-year counseling program designed to counteract the negative and unfavorable parental appraisals of low-achieving junior high school students which reinforce their negative self-concepts and impair their ability to achieve. Evidence provided by the evaluation showed that at the end of the year the counseled students had lower self-concepts than the control students. Such an outcome suggested that the counseling placed more pressure on the students to achieve and thus unwittingly added to the detrimental effects of parental disapproval. This is another example of determining the purpose of an evaluation and designing the evaluation to serve that purpose. The answers may not always be those that were expected or hoped for, but they do furnish data for decisions.

In a paper titled "A Model for Evaluation: Design for a Rape Counseling Program," Bennett (1977) argues that emphasis on the method of evaluation prior to the actual beginning of a program can prove of great value to an administrator in shaping the program and in explaining and defending it to boards and funding bodies. Always, the purpose of the evaluation will suggest the type of evaluation to be conducted.

Organizing Reality:
How to Prove That What You're Doing is Effective

We have discussed the fact that evaluation may or may not be needed, and we have suggested that the purpose of the evaluation (an important and timely purpose) and the plans for using the evaluation data must be established before intiating evaluation processes. The decision whether to evaluate the counseling program of a single counselor, a group of counselors, a total counseling staff, or counseling staffs from two or more agencies, will depend on the purposes to be served by the evaluation. If only one counselor intends to use the results, there is no point in asking all counselors to participate. Muro (1970) demonstrated that evaluation of a comprehensive program provided by a single counselor in a rural area was practical. The total program was broken down into manageable components, one of which was counseling, and effectiveness of each component was reported separately. Gold (1969) reported on the evaluation of the effectiveness of the total Student Counseling Assistant Program at Los Angeles City College. Data gathered from all counseled students and all counselors were grouped, summarized, and analyzed. This example of evaluation of staff effectiveness in one counseling program area provided information that was useful to each counselor and to all counselors, as well as to the program manager. In Oakland, California (Oakland Interagency Project, 1964) an interagency counseling program which included

elementary and junior high schools was evaluated as a total program, and yielded results useful to all decision makers involved in the project.

Let us now assume that you have completed the preliminary steps and have identified a counseling program that you want to evaluate. To illustrate the process of providing evidence of the effectiveness of the program, an example common to most agencies will be used: an individual counseling program for clients with problems in interpersonal relations. Let us say that the reality of the work setting and work load is that the counselors are counseling with so many clients across so many content areas that it is difficult for them to know which of their sessions are effective and which are ineffective. Until they know this, they cannot decide which of their techniques are working and which are not, so they have no basis for refining their approach. Suppose the total counseling staff decide to organize the reality of their services by focusing on one counseling program at a time--they agree that counseling clients about interpersonal relations problems is common and frequent enough to warrant scrutiny.

The counseling staff feel that since this counseling activity consumes so much time and effort, they need to know whether it is an effective program. This, then, is the purpose of their evaluation. It is not to prove anything to administrators; it is not to justify the program to the governing board. It is simply to give

themselves feedback on the efficacy of a large segment of their efforts. They want to evaluate total staff effectiveness, but they agree to preserve data about each individual counselor's program effectiveness in order to give that counselor some information s/he might use to modify priorities. Thus, the decision is to evaluate the effectiveness of the counseling program related to interpersonal relation problems, and to include all counseling staff members in the study. Focusing on results, they hope to determine the impact of this counseling program on the clients by undertaking the following steps:

Step 1. Translating activities into expected results

It is not possible to determine the effectiveness of a program until you have carefully stated the client outcomes expected to result from the program. In this example, the counselors might ask themselves, "Why are we counseling these clients about their interpersonal relation problems? How do we expect them to change as a result of our intervention? What do we want them to be able to know or feel or do now?" This questioning process facilitates translation of the activity (individual counseling) into outcome(s). Answers to the questions will suggest desired client outcomes. In the example, outcomes might be stated as follows: "Clients participating in a series of six individual counseling sessions will demonstrate mastery of at least one interpersonal skill they lacked when they entered the counseling program." Or, "Clients whose

interpersonal behavior reflects lack of respect for the opinions of peers will demonstrate ability to communicate respect in at least one simulated discussion with peers." There is no limit to the numbers and varieties of outcomes that can be specified. Therefore it is important for you to delimit the scope of your inquiry by reference to the purpose(s) of your evaluation. This leads you to the next step.

Step 2. Formulating decision questions

We said that the purpose of the evaluation of this counseling program is to determine its effectiveness, since so much time and effort go into it. That's rather broad and ambiguous. How can the counselors bring some precision to the investigation? First, they should identify the decision makers. Who is involved in the program? Certainly the counselors themselves and their clients--but is anyone else affected by the program? Perhaps the supervisor and/or a governing board might have some concerns. Once all decision makers are identified, they should be informed of the intention to evaluate the program, and asked to identify decisions they might need to make concerning the program. A supervisor/governing board might want to make decisions about whether or not this program is worth the resources expended on it. Counselor decisions might be whether to expand or reduce the numbers and/or types of clients included in the program, whether to transfer responsibility for the program from the total staff to an individual counselor, whether the counseling

technique is more effective than another. Client decisions might include whether the sessions are worth the investment of their time, whether learned skills carry over into real life situations.

Once the decisions are identified, they need to be examined for importance, i.e., "What will happen if we do or do not make this decision? What effect will it have on clients? On staff? On the budget?" Questions like this help narrow the field of inquiry. Some decisions would be nice but are not necessary; some might seem important, but gathering factual information to support them would not be feasible. For purposes of our example, let's say that the decisions to be addressed are the counselor decision whether to expand or reduce the numbers and types of clients served by the program, and the supervisor's decision as to whether or not the program is worth the resources expended on it.

Step 3. Determining questions to be answered and indicators that will best present evidence of success

Since in our example the counselors and the supervisor are the ones who will be making decisions based on the results of the evaluation, they must specify what information they need to make those decisions. What questions do they want to ask? What indicators will they accept as evidence that the program is successful? For instance, the counselor decision revolves around differences in effectiveness of the program with clients with varying interpersonal difficulties. They cannot decide whether to expand or reduce

the numbers and types of clients served unless they get differential information about program effectiveness. Their questions might be: "Which type of problem(s) is most readily ameliorated by this counseling program?" "What is the average time required for successful amelioration of each type of problem?" Supervisors might want answers to these same questions, but because their decision relates to efficient use of resources, they might also want to ask, "Is one counseling technique more efficient than others in use of resources (personnel time, materials)?" The questions agreed upon become the focus of the evaluation.

Next, the decision makers must agree upon what indicators (what behavioral evidence) they will accept as answers to their questions, answers upon which their decisions will then be based. Counselors might specify the five or six types of interpersonal relations problems that most frequently characterize their clients, and then agree on a demonstration task related to each type of problem which the client might perform to display the learned skill. They would also establish success criteria for determining whether the level of skill displayed was satisfactory. For instance, for the problem of lack of respect for opinions of peers, the demonstration task which would serve as an indicator that the skill had been learned might be for the client to role-play a simulated situation with another client in which marked differences of opinion are expressed. The success criterion would be the client's ability to address the

difference without denigrating the other client's point of view.

Supervisors might want indicators such as relationship of cost to success for each technique used with clients with a specific type of problem. They might set standards in advance, such as, "For the program to be considered successful, 70% of the clients must demonstrate ability to use the learned behavior." This success standard would be applied to each technique before cost comparisons were made.

Step 4. Selecting an evaluation design

This is probably the step most frequently neglected or abused by practitioners evaluating their own programs. It is not uncommon to read a report that presents glowing accounts of what program participants are able to do at the end of a program, but gives no evidence that they were not able to do these same things before their participation in the program. Without a pre-specified evaluation design, no comparison standard will have been determined, and it will be impossible to ascribe client outcomes to the effectiveness of the program. Many types of evaluation designs are possible, some much more sophisticated than others, but all provide comparison standards or criteria by which evaluation results can be judged. These standards or criteria must be specified in advance in order to sharpen the focus of the search for information. Because many available references treat evaluation designs in depth, no detailed explanations are provided here. Rather, brief examples of the use of some of the designs are presented.

Experimental design. The evaluation method most commonly known is the experimental design. Because in its most precise form the experimental design depends on random assignment of clients or units of concern (e.g., a neighborhood center or a classroom) to experimental and control groups, and because practitioners tend to characterize this type of design as sophisticated or scientific, it is sometimes rejected without serious consideration of its feasibility. Moreover, since many appear to think this is the only way to evaluate a program, they reject the whole idea of evaluation as being impractical. Among the most common of experimental designs are the pretest-posttest control group design, the posttest-only control group design, and the Solomon-Four group design. See Anderson and Ball (1978) or any good evaluation textbook for further discussion of such designs.

Quasi-experimental design. When the evaluator lacks control over when clients are exposed to a program or which clients are exposed to it, a quasi-experimental design may be employed. Examples of this type of design are the time-series design (in which measurements taken over a period of time before introduction of the new program are compared with periodic measurements after program implementation) and the pretest-posttest nonequivalent group design. Although these designs are generally practical in social system settings, they should not be undertaken without access to the services of an evaluator who is keenly aware of the inherent problems

in each, and of ways to reduce the effects of these problems. See Anderson and Ball (1978) for further discussion.

Correlational studies. These studies are quite common. They set out to show, for instance, that boys can name more professional occupations than can girls of the same age. Such a study may demonstrate that there is a correlation between sex and knowledge of professional occupations, but it does not address the causes of these differences. Without knowing causes, it is difficult to justify making program decisions based on the data--unless the decisions have to do with the correlates, e.g., whether girls need more instruction than boys in order to bring them up to the same level of knowledge of occupations. The problem with correlational studies is that the results frequently are used as though the studies were cause-effect studies.

Surveys. One of the most common evaluation designs used by counselors is the survey. Unfortunately, it is seldom used with enough precision and rigor to ensure that the results are trustworthy. Failure to specify the kinds of information to be gathered, the nature of the population to be sampled, the reliability and validity of the instrument, the motivation of the respondents, the knowledge/experience background of the respondents (and many more concomitants), renders many survey studies useless. A large number of survey instruments--particularly those locally constructed--seek opinions of the respondents, even though the items may appear to be tapping their knowledge. Opinions are of little value unless they

are informed opinions, based on knowledge of the program as it relates to each item. Even when the survey instrument is carefully constructed, the population sample is appropriately selected, and procedures are clearly defined for ensuring full and motivated participation, failure to specify in advance what will be the criteria for judging the meaning and implications of the data collected can abort an otherwise sound study.

Two studies (Ogston, 1970; Tallon, 1973) illustrate the use of the survey design in evaluating counseling programs. Ogston sought to determine student satisfaction with services and any differences in counseling experiences between those who felt they had benefited from counseling and those who did not. He found that somewhat more than half of the students were not satisfied (which may be typical of most counseling centers). He determined that there was a need to re-examine the type of service offered. For the decision he needed to make, i.e., whether or not to re-examine the type of service offered in the counseling center, an opinion survey was adequate. It did not provide information as to why some students were satisfied and some were not, but it did furnish information as to whether it was worthwhile to put additional resources into a more definitive study.

Tallon conducted a survey to evaluate the student counseling services at Niagara College. A 23-item questionnaire was developed to obtain information about respondent characteristics, use of

counseling services, satisfaction with counseling, perception of counseling effectiveness, and preference for certain counseling styles. Counseled respondents were found to be very satisfied with the counseling at Niagara. However, the author found that although most students used counseling services primarily for educational concerns, they would consult an instructor before going to a counselor with academic problems. In the case of personal difficulties, most would consult a close friend before going to a counselor. This study serves to illustrate some possible errors in interpretation. For instance, since the students indicated that they would first go to a friend rather than to a counselor with personal problems, the counselors might infer that this was an area in which they could reduce their services. However, they had not determined why students would rather go to a friend with these problems; perhaps it was because the counselors were not doing a good job in this area, or because they were perceived as rejecting referrals in this area, or because their image was so educational-success oriented that they were not even thought about when personal problems arose. It is imperative that evaluators not stretch the findings beyond the facts that they reveal. No inferences can be drawn unless a true cause-and-effect relationship has been established.

Personnel vs. client assessment. Another evaluation design looks at the congruence between personnel and client assessment. A self-study of the counseling department in a public school system

(Suffield Board of Education, 1974) attempted to clarify the role of counselors as perceived by various members of the school system (e.g., administrators, teachers); the congruity between counselors and others in perception of counselors; and the satisfaction and dissatisfaction of all concerned with the counseling services. Analysis of congruence provided excellent information for program modification, communication, and dialogue. Hoss (1968) reported a study of counseling services and development of a survey instrument at Miami-Dade Junior College. The study sought to answer questions about the impact of counseling services through a questionnaire administered to the teaching faculty and administrative staff about their understanding of the services, the purposes, and the effect on student behavior. This was another attempt to measure congruence between perceptions of different groups affected by a counseling program.

Systematic expert judgment. Another type of evaluation design is systematic expert judgment. This format involves submitting to a panel of experts information about the objectives, programs, and personnel of a counseling program in relation to established needs, and asking them to assess the logic/adequacy/appropriateness of the counseling efforts. One approach to this type of evaluation is the Delphi Technique, a method of developing and refining group consensus, which is frequently adopted in situations where convergence of opinions is desirable. It is particularly useful in process

evaluation. The technique usually consists of a series of questionnaires mailed to a specific group of experts concerning some important problem or question. Responses from each round of questions are summarized and recirculated, and this process is continued until consensus is reached. Obviously this type of evaluation has its limitations; although the consensus of the experts may reinforce a program's direction (or find it woefully inadequate), the results are judgmental rather than factual, and therefore appropriate for use for only limited decisions.

Case study. Still another type of evaluation is the case study. Case studies have been around for a long time, but they are used as an evaluation design too rarely. Actually, careful analysis of a random sample of cases involved in a program frequently can yield more meaningful data for decisions than a more superficial study of a broader population.

With all these types of evaluation designs available, how does one decide which to use? The answer can be found in an examination of the purpose(s) of the evaluation, the decisions to be made, the questions to be answered in order to make those decisions, and the indicators that will furnish answers to the questions. In the case of our example, since the counselors are looking at relative effectiveness of different techniques, relative effectiveness across different facets of a broad client problem, and costs related to effectiveness, they need evidence of client change and objective information about

time, costs, and the like. Therefore, the case study method, if broad enough to furnish adequate coverage of types of problems, would be cumbersome and costly. The systematic expert judgment would not get at the facts. The personnel and client assessment or the survey method would be more likely to yield opinions than facts. The correlational approach would give them comparative data by problem types, but it would not relate these data to specific techniques. Only the experimental or quasi-experimental designs hold promise of providing the data needed for their decisions.

This will not always be the case. In many cases needs may be so straightforward that one of the simpler methods can be employed with reasonable assurance of producing valid results; but in the case of our example, the counselors will need to choose a design that will provide them with data that can be generalized--data from which they can infer that if the same counseling program were repeated, the same results would obtain. Probably their best choice will be the pretest-posttest experimental-control group design.

Step 5. Selecting the evaluation sample

Once the design is selected, the evaluation samples must be chosen. If a large number of clients are involved in the counseling program being evaluated, it is desirable to limit the evaluation to a random sample of the clients. Since the counselors in our example are using an experimental-control group design, they need also to identify a similar population not involved in the counseling program,

and to select a random sample from that population also. Characteristics of the two samples must be comparable.

Contrary to popular belief, random sampling is not achieved by the toss of a coin, by pulling numbers out of a hat, or by choosing every fifth or ninth or tenth name from a list. Statistical randomization gives every member of the population an equal chance of being selected each time a subject is chosen. Tables of random numbers are available in most books dealing with statistics. One should arrange the population in alphabetical or any other predetermined order, number them, and then apply the table of random numbers to selection of the evaluation sample.

If only a relatively small number of clients participate in the counseling program, it will be advisable to include the total group in the evaluation.

Step 6. Selecting/developing data collection instruments and procedures.

Step 3 specified decision questions, and determined the indicators that would provide answers to those questions. Step 3 places constraints on Step 6. Selection or development of instruments cannot be an independent activity; it must be related to the indicators determined in Step 3. In that step it was decided that one acceptable indicator would be clients' successful performance in a role playing situation. The task now is to search for an instrument that provides such a simulated situation. Because this is a unique

program, it is unlikely that an existing instrument will be found; it probably will be necessary to develop the instrument. In many cases, however, the specified indicators are addressed in existing instruments, and a minimal search will reveal them.

Other instruments needed for our example are forms for recording client characteristics and problem categories; forms for recording expenditures of resources (time and materials) in counseling sessions on the various types of problems; and forms for recording comparative data on the effectiveness of various techniques.

Step 7. Establishing baseline data

It is impossible to know how far you have come unless you know where you started from. To measure improvement in interpersonal skills, the counselors must know the status of the clients' interpersonal skills before treatment. Since the term "interpersonal skills" refers to a broad domain, they limited their domain by specifying the five or six types of problems most frequently encountered in counseling sessions, and it is these that will be the foci of their inquiry.

Part of their inquiry pursues the relative effectiveness of the program with the different types of problems. Therefore they will need to identify the central problem for each client in the evaluation sample, and for each person in the control sample. Identification of the problem can be accomplished by interview in most cases, in other cases by observation. Once the problem has been

identified, the person's behavior related to that problem must be sampled in order to establish baseline--starting point--data. This is where the pretest is applied. Usually the pretest will be the same as the posttest. In some cases it may differ slightly, but it will have been determined to be equivalent.

Baseline performance of the experimental and control group evaluation samples must be obtained and recorded. Posttest data will be compared with these data to determine program effectiveness.

Although our example is one that probably will require development of an instrument, many instruments exist which may be used to evaluate the effectiveness of programs. Types of instruments may be test scores (standardized or norm referenced), observation data, review of records, or unobtrusive measures. If you wish to measure aspects of career development, one of several career development instruments may be used. If you wish to measure attainment of unique behaviors not covered by existing tests, you may wish to develop criterion tests. If the behavior you are seeking to evaluate is situationally bound (such as in our example), you may wish to apply observational techniques. Our example uses observation of a structured demonstration by the client; observation of the client in an unstructured situation may also provide the desired information. If the behavior you are seeking is hard to measure by paper and pencil tests, and hard or impossible to observe directly (e.g., a change in attitude), unobtrusive measures may be used. Examples

of unobtrusive measures are attendance patterns, anecdotal records, choices expressed or carried out, peer group participation, recidivism, and referrals. See Metfessel (1967) for further examples of unobtrusive measures.

Bardo and Cody (1975) list difficulties in evaluation of guidance programs that deal specifically with measurement. Some examples are provided of alternative means to the traditional approaches of measurement related to evaluation in general and guidance in particular. A publication by MacLennan and Levine (undated) describes some nontest methods which can be used in program evaluation. Although the publication addresses evaluation of programs for disadvantaged children, the nontest procedures for data gathering are applicable to other programs as well.

Step 8. Monitoring progress toward results

Early in this paper we stressed the importance of a two-pronged evaluation: process and outcome. The outcome evaluation tells us what was achieved by the program, and the process evaluation tells us why it was achieved (and/or why achievement wasn't greater). The previous seven steps do focus on outcomes (evidence of effectiveness), but we would be remiss if we did not reinforce the concept of process evaluation in order to answer the "why" questions.

Using our example, process evaluation would include the monitoring of the counseling program implementation (e.g., names of clients, types of problems, number of sessions in which each

participated, techniques used in the sessions, names of session counselors, deviations from planned procedures, and level of client participation in counseling sessions). This can best be done by preparing an analysis of the program in advance of its implementation, and by designating who is to perform which specific program activities, with which clients, by which dates. This analysis should be written, preferably in chart form, showing task (counseling session and technique), talent (name of counselor and client), and time (dates by which each session will be completed). This is called a T/T/T/ (task, talent, time) analysis chart. During the implementation of the program it is important for each counselor to maintain meticulous records of counseling sessions as specified by the T/T/T, so that when outcome data are available, it will be possible to trace the implementation steps to determine what went wrong (if a segment of the program was ineffective) or what caused the positive outcomes (for those segments of the program that were effective). These are the data needed to make decisions about program modifications. The mere fact that a segment of the program was effective or ineffective tells us little; we must examine the program implementation carefully to discover the reasons for success or nonsuccess.

Pine (1975), in a journal article on evaluating school counseling programs, states that such evaluations have been characterized by an emphasis on experimental, descriptive, and case study approaches. He states whereas these studies are important in providing output

information, they are limited in developing feedback needed while a program is in progress. The author recommends that process data be collected and used for program improvement.

Part of process evaluation has to do with the effectiveness of materials. Whereas our example does not involve purchase or use of materials, many program evaluations do. Weinrach (1974) synthesizes the 1972 revision of the National Vocational Guidance Association's Guidelines for the Preparation and Evaluation of Career Information Media. Such evaluation is part of process evaluation, as the effectiveness or ineffectiveness of materials will probably have a strong impact on program results.

A study by Lynch (1968) was able to provide information for process decisions because process evaluation was employed. This study used flow charts to trace student involvement with successive program features and found that there was limited difference in the results of three different approaches to pre-enrollment counseling techniques. Progress monitoring makes it possible to modify a program during its implementation (if necessary to ensure success) instead of waiting until the program is completed.

Step 9. Collecting, managing, and analyzing data

Each step is important. The overall important point, however, is that each step be meticulously planned and executed. Problems can arise at any step, but problems are best avoided by careful planning and implementation, and by careful documentation of each

step. The designation of an instrument for data collection is just a beginning. Many important decisions still remain: Who will administer the instrument(s)? Who will train the test administrators? In what procedures? When will the instrument(s) be administered? What are the criteria for scoring? Who will score the instrument(s)? If an observational technique is employed, how will inter-rater reliability be established? How will scores or performances be recorded? By whom? When? How will data be summarized? Reduced? Analyzed? All these questions must be addressed before the evaluation is initiated.

The data collection is fairly straightforward. Persons trained to administer the instrument(s) according to author/publisher directions, and to score and record scores, present raw scores to the evaluator. The evaluator decides, based on evaluation design, how to summarize, convert, and/or reduce the scores. For instance, in our example, the evaluator could not summarize the performance of all the clients on the demonstration task and ignore summarizing scores by each type of problem. In order to answer the questions as to which types of problems are best ameliorated by the counseling program, the evaluator must preserve subgroup as well as total group data.

Data analysis is another matter. Sometimes a simple count is all you will need to prove gains. If, however, you are using an experimental design such as in our example, you will need to show that any differences in performance between counseled and noncounseled

clients are not due to chance factors. You will need to apply a test of statistical significance, such as the t Test.

Data analysis, like evaluation design, is simple to complex in direct proportion to the simplicity or complexity of the questions being investigated. If one of the questions were, "Will the counseling program result in twice as many students participating in community volunteer service?", a simple before- and after-counseling count of numbers of students so involved would give you all the information you needed. When you are wording your questions, you should word them in such a way that they will not require more sophisticated information than you really need. On the other hand, a question such as, "Will more students participate in volunteer community service after counseling than before?" is too nonspecific to provide guidelines for judging the program's impact. What if one more student participated after the counseling? Does that make the program a success?

For more information on data analysis, consult any current text on evaluation and/or statistical analysis. Anderson and Ball (1978) is one of the easier sources to follow and understand, in addition to being thorough and accurate. If you are in doubt as to which type of data analysis to apply, you may need a local consultant. Such consultants are available from institutions of higher learning, from research and development laboratories, from private contracting firms, and from some local public and private

agencies.

Step 10. Answering questions and making decisions based on results.

Step 9 ensures the accurate collection of appropriate data, and analysis and presentation of the data in such a way that they can be applied to answering the questions formulated in Step 2. But remember that the collection and analysis of data are just that--no more. The data presented are neutral data, neither good nor bad. Evaluation means applying values to the data and making judgments about the data based on those values. Therefore, early rigor in deciding what decisions need to be made, what outcomes are considered important by the various publics affected by the program, what questions need to be asked, and what indicators will constitute a basis for making those decisions, will facilitate Step 10 and prevent decision makers from being diverted from the original intentions of the evaluation.

At this point the data, presented in a form that permits their direct use in answering the evaluation questions, are applied to the questions without distortion. You should prepare to be disappointed. Frequently a program that "feels good" fails to be validated by the evaluation. This is not necessarily evidence that the program is not effective. It may not be effective in terms of the outcomes under investigation, but it may be very effective in producing outcomes not being studied; or it may be that the indicators and instruments used to assess program impact were not as valid and functional as they seemed to be. Answers

to your questions should not be accepted blindly. If they are different from what you had good reason to expect, trace your evaluation steps backward and see if you can detect a possible flaw. Once found, such a flaw can be corrected, the succeeding steps repeated, and the results markedly changed. Of course it has been the intention of this paper to caution the evaluator to plan and implement carefully, after thorough investigation of the tools and procedures available, in which case the evaluation results should be trustworthy. However, suspect results bear rechecking.

Applying Step 10 to our example, data would need to be arrayed in such a way as to give objective information about the relative effectiveness of the counseling program in ameliorating different types of interpersonal problems, the average time required for successful amelioration of each type of problem, and the relative effectiveness of each counseling technique. Perhaps the problem of lack of respect for others' opinions, which had been one of the most common problems, was improved very little by the counseling program. Or perhaps one counseling technique emerged as effective with this problem whereas other techniques were relatively ineffective. But perhaps that technique required so much counselor and client time that the cost outweighed the values placed on the change. The data provide the information, but evaluation requires attaching values to the data so that decisions can be made.

With the answers to your questions in hand, you as the decision maker are now ready to apply this information to your decisions. A caution here: Evaluation data should not be used to make decisions other than those for which they were designed. It is easy, administratively, to say, "Let's use this information instead of taking the time to generate other information," when faced with related decisions. For instance, an administrator may use survey data gathered in a self-study of counselors for their own planning to present to an accreditation team, in lieu of a study designed for that purpose. This constitutes misuse of data and frequently will result in decisions being made on faulty information.

Step 11. Reporting results

It may seem simplistic to have a section on reporting results. However, failure to write reports that are understandable and meaningful to the particular audiences for which they are intended, and failure to motivate these audiences to read and respond to the reports once they are disseminated, cause the results of many evaluations to be ignored. If more than one group of decision makers was identified in Step 2, you may need a separate report for each. In the case of our example, a single report will probably be adequate, because both the counselors and the supervisor have similar professional and experiential backgrounds and can understand reports written at a fairly technical level. If, on the other hand,

one of your audiences (decision makers) was the governing board, (made up of lay people), you probably would need to prepare a separate report for them, geared to their knowledge and experience in the area under study.

Preparation of appropriately addressed reports, however, is only one part of reporting. A very important activity that will increase the likelihood of the evaluation report's being read and its results used to make the specified decisions, is development and maintenance of motivation to do so. Such motivation begins with involving decision makers in stating needed decisions, formulating questions, and specifying acceptable indicators before the evaluation is begun. But because many evaluations consume a year (there must be ample time between pretreatment and posttreatment to give the program a chance to make a difference), other concerns intervene and decision makers tend to lose their enthusiasm or even their commitment to making the specified decisions. Therefore, it is important that process and progress evaluation reports be prepared during the implementation period, and communicated to these audiences often enough and in such a mode as to maintain their interest and involvement. Occasional short progress reports presented orally over coffee and doughnuts may serve this maintenance function. The techniques chosen will depend upon the agency climate; they should be those most likely to communicate program evaluation information, maintain interest, and ensure ultimate use of results, without constituting an undue imposition on the time and energies of those you are trying to motivate.

Another way to ensure that evaluation results will be used is to develop in advance a written plan for their use, together with names of personnel who will use the results, decisions to be made, and suggested timelines for making those decisions and for acting on them.

What To Do When Evaluation Goes Wrong

At several points during discussion of the steps in evaluation, possible problems were mentioned. The fact was stressed that the evaluator must be knowledgeable about the alternatives and hazards at each step of the evaluation. There is no way that this document can present all problems that might be encountered in evaluating the effectiveness of counseling programs. The reader is referred to other documents for a more complete picture of these hazards. Publications such as Tallmadge (1977), Bonnet (1977), and Anderson and Ball (1978) will help. We will discuss only a few of the more common problems here, and offer some suggestions as to what to do if they happen. Assuming that you have followed all the steps suggested in this paper, we will not address problems that can occur if you fail to complete one or more of these steps. Instead, we will give our attention to common problems you may encounter even when you are trying to follow directions.

Lack of realism about outcomes

The first problem is the failure to apply logic and realism to the first step: translating activities into expected results. In your zeal to justify an activity, you may tend to create all kinds of unrealistic expected outcomes, without taking a close and logical look at whether the activity really has the potential to produce those outcomes. If you have made this error--and you didn't catch it in time--you may find that the results of your evaluation are negative: They indicate that the program was not effective in producing the desired outcomes. If this occurs, the first thing to do is to ask a panel of experts not associated with the program to take a look at the activity and the expected outcomes, and to see if they believe the activity could logically be expected to produce the outcomes. They may detect some logical fallacies which you missed. If they do, you will need to adjust your activity to fit the outcomes, or adjust your outcomes to fit your activity, and again pursue the steps of the evaluation.

Too much information

A second common problem is the tendency to try to get too much out of the evaluation. Forgetting that you really will be making only one or two decisions as a result of what you learn, you are inclined to identify long lists of questions you would like to have answered. Before beginning the project you should cross off all the questions that are not needed to provide

information for a selected few priority decisions. In spite of doing this, if you still end up with too many kinds of information to be gathered feasibly, you will probably notice this problem by the time you select the evaluation design, or at least when you try to select or develop a data collection instrument. At that point, if it looks as if your design is too complex for your resources and/or your instruments are too long or time consuming to be administered feasibly, you should back up and see if you can reduce the number of decisions and questions. Then proceed with the steps.

Too complex indicators

Another problem has to do with indicators (which also affects how the questions are phrased). In our example, if the only indicator of interpersonal skill that you will accept is direct observation of each client in an unstructured social situation, you will have locked yourself into a data collection problem that cannot be solved within the constraints of your resources. As soon as you recognize this, go back and modify the questions and the indicators so that they still provide information essential to making the specific decisions, but are simple enough to be feasible. This issue was addressed briefly in another section. For example, "Do clients pursuing the counseling program for six weeks improve their interpersonal skills significantly more than clients not involved in the program?" is a question that locks you into an experimental design. Experimental designs are not always feasible in human services

agencies, as it is difficult to randomize populations and difficult to identify and involve comparable control groups. If you want to avoid the complexity of this design, reword the questions and restate the indicators of success which will satisfy you and the other decision makers. However, do not water down your questions and indicators to a point that they are meaningless. If you can be satisfied with a simple numerical gain in incidence of such behaviors as attendance or completion of assignments, and that is all the information the decision makers really want, a straightforward question (including as criterion the numerical gain that will be acceptable evidence) will legitimize a much less complex evaluation design. It is never too late to go back to previous steps to make adjustments when you run into problems, but remember that all steps subsequent to the adjusted step must be repeated in order to keep the evaluation consistent and coherent.

Unreliable measurement instruments

A major problem in local evaluations is the failure of locally-developed data collection instruments to discriminate between those who have or have not mastered the expected skill (outcome), or to produce reliable results. Instrument development is not simple. Many seem to think, for instance, that a teacher-made test is a criterion test; on this premise they think that if they specify the desired behavior and develop an item or similar items to those in teacher-made tests, they have a criterion test that will work.

Unfortunately, the development of criterion measures is no simpler than the development of items for a norm-referenced test. Try to avoid problems by subjecting your locally-developed instrument to a series of tests to establish content validity, construct validity, criterion-related validity, and reliability. If you have done all this and your instrument still fails to give you the information you want, look again at the congruence between your activity and your expectations. Perhaps you are trying to measure something that cannot generally be expected to result from the activity, whereas many other good outcomes are resulting that are not being tapped. As in "Monopoly," if this happens, go back to "Go" and again start working your way forward in the process.

Regardless of the problem you encounter, it is always possible--though it may take some digging--to identify the cause of the problem if you have followed the logical steps suggested in this paper. When you do discover the cause, you can remove it, but you must re-perform all subsequent steps since they are progressive and interactive. In no case can you correct an early step without affecting all succeeding steps.

Building a System Out of All Your Counseling Programs

One of the greatest lacks in most counseling services is a system of linkages between and among the sub-programs and/or

individual counselor activities. Each activity may effect certain desired and desirable changes in clients, but until all activities have been evaluated (simple evaluation wherever possible), and the results of all the evaluations have been analyzed and synthesized, it is impossible to evaluate the total program. It is impossible to know whether services are reaching all clients who need or want the services, whether services are overlapping, which services are effective, whether some techniques are more effective than others, whether the total services are taking care of the developmental needs of the clients or are merely problem-oriented, whether the services constitute psychological bandaids or are really capacitating clients to select and apply their own solutions to problems. In other words, although it is possible and desirable to evaluate sub-programs and even individual activities, the impact of the total counseling services cannot be known until all segments or components have been evaluated and the results analyzed and synthesized. This ultimate goal of establishing comprehensive programs which serve all clients and capacitate them to become self-helpers should motivate each reader to embark on an evaluation program <u>now</u>, even if it is to focus on only one counseling component.

Recommendations

The previous sections contain so many directives, suggestions, and cautions that it seems almost anticlimactic to add a section on recommendations. However, a few are in order, addressed to various audiences.

It is recommended that counselors in all agencies perform Step 1 of our evaluation process for each activity they perform. Take an honest look. Why are you doing what you're doing? What client benefits can you honestly and logically expect to accrue from this activity? Are these benefits important to the clients? (It is not enough that they are important to you.) Having performed this task, have the grace and the courage to abandon those activities that are not contributing to client growth, and examine each of the others to see how it can be refined to produce more surely the intended client outcomes.

Next, counselors are encouraged to work together to plan evaluations of individual counseling components, and ultimately to analyze and synthesize the results of all components. Only thus can a coherent program which reaches all clients and which has a developmental thrust (to capacitate clients) emerge.

Counselor supervisors are asked to consider the following recommendations: Encourage and support counselors in their assessment of the effectiveness of their activities, providing expert consultation in evaluation areas in which they lack

competence. Participate in their evaluation efforts, and make sure that their evaluations are providing data for your decisions as well as theirs.

It is recommended that counselor educators make sure that all counselors, regardless of their discipline or work setting, master the concepts and skills necessary to perform creditable ongoing evaluations of the effectiveness of their programs, and that during the full course of their training they be imbued with the importance of performing such validity checks of their work. It is further recommended that counselor educators provide trainees with guided practice in evaluation by requiring them to perform, as part of their field work experience, a sound evaluation of the effectiveness of at least some of their activities.

It is recommended that program developers, when installing and implementing their programs, include in their dissemination materials validated procedures for performing both process and outcome evaluations. Such materials should also include guidelines for modifying both the program and the evaluation steps to fit the situation.

Educational researchers are encouraged to find effective ways to communicate the results of their research, and to translate these results into operational terms, suggesting practical applications for counselors. Professional organizations might take the lead in this thrust toward useful dissemination of research findings.

Summary

Because many general program evaluation models exist but few sources directly address or are limited to evaluation of counseling, this paper begins by outlining the parameters of such evaluation. Counseling program is defined as a single counseling session, a group of related sessions, counseling performed by a single counselor, or the counseling activities of an entire staff. Instead of focusing on comprehensive program evaluation, an 11-step process is presented for evaluating any component or components of a counseling program. The emphasis is on meticulous and rigorous performance of each step. Problems that can and do arise during evaluations are discussed, and the evaluator is urged to trace back through the steps to find where the error or ambiguity crept in, perform that step adequately to remove the error or ambiguity, and then repeat all succeeding steps, modifying them in relation to the initial step change. Recommendations stress the importance of counselors learning how and when to evaluate, reducing evaluation tasks to their simplest terms in order to keep the task reasonable, and ultimately analyzing and synthesizing the results of evaluations of all program components so that a cohesive and interrelated system which addresses the developmental needs of all clients can emerge. The processes discussed in this paper are applicable to determining the effectiveness of any counseling program, regardless of discipline or work setting.

REFERENCES

Alkin, M. C., & Kosecoff, J. B. A new eclectic model for the redirection of evaluation efforts. Los Angeles: California University, Los Angeles. Center for the Study of Evaluation. March 1973. ERIC Document Reproduction Service No. ED 075 086

Anderson, S. B., & Ball, S. The profession and practice of program evaluation. San Francisco, CA: Jossey-Bass Inc., 1978.

Banathy, B. H. The educational information consultant: Skills in disseminating educational information. ERIC Document Reproduction Service No. ED 071 689

Bardo, H. R., & Cody, J. J. Minimizing measurement concerns in guidance evaluations. Measurement and Evaluation in Guidance. 8(3) 175-178. October 1975.

Bennett, J. R. A model for evaluation: Design for a rape counseling program. Child Welfare, 56(6) 395-400. June 1977.

Bonnet, D. G. The design of career education evaluations in 1975-76. Crawfordsville, IN: New Educational Directions, 1977.

Dorosin, D. Very brief interventions: A pilot evaluation. Journal of Counseling Psychology, 24(5) 191-4, April 1976.

Dreman, S.B. Expectations and preferences of clients for a university counseling service. Journal of Counseling Psychology, 24(5) 459-461, September 1977.

Hamachek, D. E. Characteristics of low-achieving, low-concept junior high school students and the impact of small group and individual counseling on self-concept enhancement and achievement. ERIC Document Reproduction Service No. ED 017 549

Hoss, M. A. A study of counseling services and the development of a survey instrument at Miami-Dade Junior College. Washington, D.C.: American Personnel and Guidance Association, April 1968.

Lynch, T. H., Eichsteadt, A. C., Montaba, R. P., Harmon, L. W., & Kramer, P. H. The selective effects of different pre-enrollment counseling programs on entering freshmen. Program given at the American Personnel and Guidance Association Convention, Detroit, MI: April 7-11, 1968. ERIC Document Reproduction Service No. ED 024 976

MacLennan, B. W., & Levine, M. A. Evaluation issues in programs for disadvantaged children: The use of non-test procedures. ERIC Document Reproduction Service No. ED 014 759

Metfessel, N. S., & Michael, W. B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Educational and Psychological Measurement, 27(4) 1967, 931-943.

Mitchell, A. M. A handbook of evaluation models: Ways to evaluate different types of career education activities. Palo Alto, CA: American Institutes for Research, 1978.

Muro, J. J. Evaluation of a comprehensive counseling and guidance service for rural Maine communities. Camden, ME: Maine School Union 69, 1970. ERIC Document Reproduction Service No. ED 046 561

Oakland Interagency Project. Report of the evaluation of the inter-agency school project's counseling program at elementary and junior high school. Oakland, CA: Oakland Unified School District, 1964. ERIC Document Reproduction Service No. ED 001 595

Ogston, D. Followup: A study of counseling outcomes. Canadian Counselor, 4(4) 255-259, October, 1970

Pine, G. J. Evaluating school counseling programs: Retrospect and prospect. Measurement and Evaluation in Guidance, 8(3) 136-142, October 1975.

Suffield Board of Education. A self-study of the guidance and counseling department in the public schools of Suffield, Connecticut, 1973-74. Suffield, CN: Author, 1974. ERIC Document Reproduction Service No. ED 119 081

Tallmadge, G. J. The Joint Dissemination Review Panel ideabook. Mountain View, CA: RMC Corporation, 1977.

Tallon, R. B. An evaluation of the counseling services at a Canadian community college. New York: Niagara University, 1973. ERIC Document Reproduction Service No. ED 085 053

Weinrach, S. G. How to evaluate career information materials. School Counselor, 22(1) 53, September 1974.