DOCUMENT RESUME

ED 177 371 CE 023 166

AUTHOR Kershner, Keith M.

TITLE RBS Career Education. Evaluation Planning Manual. Education Is Going to Work.
INSTITUTION Research for Better Schools, Inc., Philadelphia, Pa.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO RBS-BK-4
PUB DATE 76
CONTRACT NE-C-004-0011
NOTE 28p.; For related documents see CE 023 015, CE 023 064-067, and CE 023 170-172
AVAILABLE FROM Research for Better Schools, Inc., 444 North Third Street, Philadelphia, PA 19123 ($2.25)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Career Education; Educational Assessment; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Guidelines; *Planning; *Program Evaluation

ABSTRACT

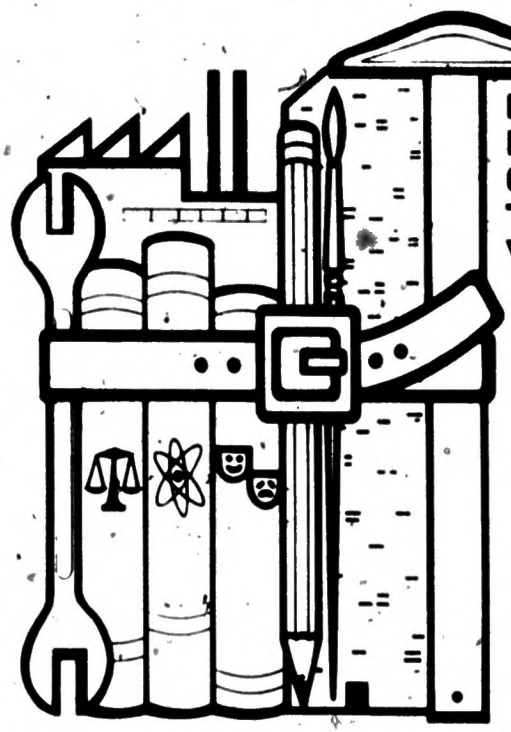
Designed for use with the Research for Better Schools career education program, this evaluation planning manual focuses on procedures and issues central to planning the evaluation of an educational program. Following a statement on the need for evaluation, nine sequential steps for evaluation planning are discussed. The first two steps, program definition and evaluation questions, serve as a guide for developing the intended scope of the program and evaluation. The next five steps review evaluation methodology in terms of (1) the statement of hypotheses, (2) selection of subject groups, (3) selection of instruments, (4) creation of a data system, and (5) design of an analysis plan. The final two steps focus on planning concerns in evaluation reporting and cost projection. (LRA)

RBS CAREER EDUCATION EVALUATION PLANNING MANUAL

by

Keith M. Kershner

Publication No. BK-4



RBS Career Education

EVALUATION PLANNING MANUAL

RESEARCH FOR BETTER SCHOOLS, INC. (RBS), is a private, non-profit educational research laboratory located in Philadelphia, Pennsylvania. The EVALUATION PLANNING MANUAL is part of a series of curriculum and procedural materials developed by the RBS CAREER EDUCATION PROGRAM (Michaelita B. Quinn, Director) for a pilot project in experience-based career education (EBCE). Additional materials in the evaluation series include:

INSTRUMENT SERVICE GUIDE

ANALYSIS SERVICE GUIDE

PROGRAM MONITORING MANUAL

RBS CAREER EDUCATION: EVALUATION PLANNING MANUAL was prepared by Keith M. Kershner.

INTRODUCTION

Evaluation has been a continuing component in the development of RBS Career Education. The evaluation findings have provided a useful source of information in refining the program as well as offering evidence of program effectiveness to participants, sponsors, potential adopters and other members of the educational community.

Since RBS Career Education has become available for adoption by public school districts, a series of materials and services has been prepared to assist adopters in planning, implementing and evaluating the program. This series includes:

Materials		Services
Evaluation Planning Manual	–	Evaluation Technical Assistance
Instrument Service Guide	–	Instrument and Scoring Service
Analysis Service Guide	–	Analysis Service
Program Monitoring Manual	–	Evaluation Technical Assistance

The materials are intended to assist in evaluation planning and design, while the services make available Research for Better Schools' evaluation systems and expertise to support implementation. These materials and services are described more completely in the *RBS Evaluation Package Overview* and can be obtained from Research for Better Schools.

This *RBS Evaluation Planning Manual*, one element in the series, focuses on procedures and issues central to planning the evaluation of an educational program. Evaluation planning is discussed in a framework of major sequential steps. After a statement on the need for evaluation, the intended scope is developed by addressing planning activities in program definition and evaluation questions. Evaluation methodology then is reviewed in terms of hypotheses, subject groups, instruments, data systems and analyses. Concluding sections treat planning concerns in evaluation reporting and cost projection.

The manual has been designed to provide guidance in planning educational program evaluation. It has been developed in the career education context, but the evaluation concerns addressed are common to other programs as well. Assumptions about local conditions have been avoided in the interest of providing broad coverage

of generic topics. Applying planning suggestions to any specific program requires consideration of these local conditions. Research for Better Schools can provide evaluation technical assistance to aid in the formulation of individual designs.

2

NEED FOR EVALUATION

As it has been developed for RBS Career Education, evaluation functions to meet several major information needs, which are categorized as student diagnosis, program planning and demonstration of program effects. The content of each of these categories and the relevance of the evaluation materials to them will be described briefly.

Since RBS Career Education emphasizes the individual treatment of students, it is important to have detailed and accurate information about each student in the program. Such information can be an aid in placing students, planning their experiences and providing personal guidance. A major criterion for selecting the instruments recommended in the evaluation package (see *RBS Instrument Service Guide*) was their ability to yield useful individual data. For instance, the *Self-Directed Search* includes occupations considered, self-estimates of interests and competencies and prescriptive summaries. The *Student Attitude Survey* capsulizes student attitude toward school, work, self and others. The *Comprehensive Tests of Basic Skills* reflect functional levels in reading, mathematics, language and study skills. The *Student Demographic Data Questionnaire* gives basic information on student background variables. These instruments or others which may be selected provide the information for assembling a student profile that becomes a part of each student's record and may be used to chart his or her course. It will indicate interests, strengths, weaknesses and perceptions which are helpful in designing experiences for individuals.

The same data gathered for individual students may serve a program planning function when summarized for all students in the program. At this level, student career interests, for example, help to determine the range of career experiences which should be provided. Group needs in basic skills suggest the nature and extent of academic content and materials which would be appropriate. Affective needs also can be identified, and program elements can be focused to meet them.

Surveys of participant opinions likewise are relevant to program planning. They measure the perceptions of students, parents and community participants regarding the program and are intended to gather opinions about its various aspects and the success of its implementation. Their design allows adaptation to each site's needs and interests. Information gathered in this way is helpful in assessing program conduct from the viewpoint of the people involved, and the results often have planning implications.

Finally, the effects of the program on students may be investigated by administering instruments in a research paradigm and analyzing student development. These procedures yield information about student progress on the selected measures during their program experiences. Changes observed among project participants then may be compared with changes over the same time period among similar students who have not been engaged in the program. These comparative analyses make it possible to draw inferences of relative program impact on student development. Student effects are tested by using statistical analyses in a hypothesis framework representing the desired effects of the program.

Similar kinds of student data thus can be used in several ways if proper evaluation planning is accomplished while program implementation is being designed. Needs in student diagnosis, program planning and monitoring and the demonstration of student effects can be identified through an evaluation component. In this way evaluation activities can serve program operations, development, administration and research.

Once the importance of program evaluation has been endorsed, the next step is to plan an evaluation component which will be effective in meeting the needs. The following sections of this manual address the major steps in the evaluation process from a design perspective. Issues central to planning the evaluation are discussed, and important decision points are indicated. Detailed implementation concerns and problems have not been included because their dependency on individual conditions prevents comprehensive and concise treatment here.

PROGRAM DEFINITION

The first step in evaluation planning is preparing an accurate description of the program which is to be evaluated. Three levels or types of description are necessary for evaluation purposes: a program overview, objectives of the program, and operational strategies.

The **program overview** describes the components which have been planned for implementation, how they are organized and who is to participate. It serves as a broad definition of the scope of the program and the context within which the evaluation is to be conducted.

Program objectives are statements of what the program intends to accomplish. They may be a combination of process objectives and product objectives. Process objectives relate to the completeness or adequacy of implementation, while product objectives are concerned with the outcomes or effects of the program. An example of a process objective would be, "to provide three career exploration experiences for each enrolled student." A product objective might be, "to increase the career maturity of participating students." It is important to develop a list of all objectives which the program is intended to meet and to define them as specifically as possible. Sometimes it is necessary to have some objectives that are more abstract than others, but it should be understood that they will be more difficult to evaluate. The statement of objectives, then, defines the intent of the program in precise terms and establishes expectations of how the program will perform.

Once the objectives have been identified, they can be grouped and assigned priorities. Grouping should be done by objective type: process or product, cognitive or affective, short-term or long-term, etc. The grouping can be done in any manner which results in an understandable framework of objectives that represent the interest areas of program associates and sponsors. Relative priorities then should be assigned to guide the allocation of evaluation resources. Although the evaluation must be addressed to all objectives, some among them may merit differential attention depending on their relative importance to the program, the probability of obtaining conclusive results, the potential decision-making value of findings or other factors. Establishing priorities among objectives is helpful in clarifying any existing hierarchies of objectives and suggesting relative emphases in the evaluation design.

Operational strategies link objectives to those program elements which are designed to accomplish them. Each objective should be described in terms of the operational procedures designed to attain it. Such descriptions serve to assure that

stated objectives actually are associated with specifiable project activities. Objectives which cannot be tied to at least one activity signal a problem which requires a redesign of either the program content or the objectives involved.

The completion of these three descriptive tasks results in a definition of the program in terms of its scope, intent and programmatic process. This combined definition becomes the basis of the evaluation effort in that it circumscribes that which is to be evaluated and also establishes expectations and accountabilities for the program. Defining the program is a process that should include planners, implementers and evaluators. All project personnel should acknowledge and support the resultant statements of definition to ensure that everyone proceeds on a common basis.

EVALUATION QUESTIONS

If the evaluation effort is to meet the needs of implementers, planners and sponsors, it must focus on specific questions which are significant to the program. The process of defining the program yields objectives which are necessary in formulating evaluation questions, but the objectives themselves do not constitute such questions. Program objectives are statements of intended educational outcomes, and the evaluation process requires translating these objectives into hypothesized effects which can be empirically tested. The formulation of evaluation questions facilitates this translation process.

Stating objectives is primarily the domain of program implementers and planners because they know the intentions of the program best. The statement of hypotheses is the realm of evaluators, who know the scientific and technical issues best. Formulation of evaluation questions is the middle ground where all program associates participate equally in exploring the implications of objectives and establishing the bases for developing hypotheses. This intermediate step is a helpful process for assuring that the evaluation design fairly and completely represents program intentions. It also promotes interaction among all staff in laying the foundation for evaluation of the program.

Evaluation questions are derived from the group of program objectives through exploration of their content and implications. This exploration process should include program implementers, planners, sponsors and evaluators, who should aim at specifying observable consequences associated with the objectives and reasonable standards of success. Such specification permits the formulation of evaluation questions.

The development of evaluation questions may be illustrated by using the sample objective, "to increase the career maturity of participating students." Examination of the intent behind this objective might yield "career planning skills" and "confidence in making a career choice" as the appropriate variables represented by the objective. Further, it might be decided that the standard for judging success should be demonstrable progress of students during their participation in the program. In this case appropriate evaluation questions would be: "Do students increase their confidence in making a career choice?" For both questions the demonstrated progress of students during their participation in the program would be the standard for judging whether the objective has been met.

This procedure should be followed for all program objectives. Each objective will result in at least one evaluation question, and many objectives, upon exploration, will require more than one evaluation question to represent adequately their intent. The process of developing these evaluation questions ensures that the implications of program objectives have been examined, that the program intentions are reflected reasonably in the evaluation and that the groundwork for representative hypotheses is completed.

STATEMENT OF HYPOTHESES

The collection of evaluation questions displays the desired scope of the evaluation component. The next set of tasks is concerned with establishing the means whereby these questions can be answered reliably and validly. Although ways will be found to address most questions adequately, it may be anticipated that some questions will have to be eliminated on technical or cost grounds. The priorities established earlier will help in making these decisions.

Hypotheses are propositions or assumptions constructed to draw out and test the logical or empirical consequences of the announced objectives as they represent the program "theory." Formulating hypotheses involves the refinement of evaluation questions into testable propositions, which necessitates establishing a standard of success for each question. For example, an evaluation question might be: "Do students gain career planning skills through participation in the program?" The most basic hypothesis in this case would be: "Students will score higher at the end of the program year than at the beginning on the XYZ test of career planning skills." The standard is higher performance on a relevant measure over a year of exposure to the program. The meaning of "higher" may be defined further as some standard unit score gain or gains of any magnitude which are statistically reliable.

This form of hypothesis allows for determining changes which occur during the intervals between tests, but it does not permit conclusions about whether the program was responsible for the changes. Other factors such as maturation, peer group interactions, media exposure and other events may have had some effect during the same time period. The typical method of taking these non-program influences into account is to compare the growth of program students with the progress of similar non-program student groups on the same measures. If this option is elected, hypotheses then become comparative statements. The sample hypothesis used above would become: "After exposure to the program, students in the program will score higher than comparison students not in the program on the XYZ test of career planning skills." The term "higher" again should be defined in terms of statistical standards.

This process of refinement must be carried out for each evaluation question. It will be found that some questions will be more amenable than others to restatement as hypotheses. Comparative hypotheses, such as the career planning example, lead to the most definitive tests of results and should be used wherever possible. Many intended program effects such as improved self-concept, reading skills, vocational

attitudes and others may be cast legitimately in a comparative form.

Some evaluation questions will not fit into a comparative hypothesis paradigm because they relate only to program participants. Such questions are not appropriate for non-program comparison groups. Examples would be: "Are student interests met by the program?" "Does the business community support the program?" In such cases comparative hypotheses are not possible, and standards of success must be established entirely within the program reference. Sample hypotheses might be: "Expressed student interests are matched by program activities at least 80 per cent of the time." "Participating businesses and agencies recommend involvement to others in at least 80 per cent of the cases."

Testing both comparative hypotheses and within-program hypotheses requires acceptable subject groups, instrumentation and statistical procedures as discussed below.

SUBJECT GROUPS

After hypotheses have been formulated, it is necessary to select subject groups that can provide the data needed to test them. Evaluation is possible without comparison groups, but the usefulness of such results generally does not justify the expense of generating them. For the purpose of this manual, it will be assumed that comparative hypotheses are to be included. Two kinds of comparison group designs are discussed: true experimental and quasi-experimental.

The true experimental design requires that subjects be randomly assigned to the experimental and comparison groups. The experimental subjects participate in the program, while the comparison group members are engaged in other activities which are distinct from the experimental program. In most educational evaluations the comparison groups are enrolled in a traditional curriculum or another competing program.

Randomly assigning subjects to the experimental and comparison groups eliminates the problem of selection bias, which typically confounds other designs. Since each subject has an equal chance of being assigned to either group, the likelihood of obtaining groups imbalanced on any characteristic is minimized. This method presents the best conditions for conclusively testing hypotheses because observed group differences in measured outcomes more likely will be due to program differences rather than possible differences in the groups themselves.

Random assignment usually is possible where the number of program applicants exceeds the number that can be admitted. In these cases random assignment is actually the fairest way of determining who should be enrolled in the experimental program. Each applicant has an equal chance.

It should be understood that randomization precludes the possibility of selecting subjects on any special criteria unless such subjects are to be excluded from the program evaluation. A random assignment plan restricts the influence of staff on the composition of subject groups so that energies are directed toward ensuring that the applicant pool contains the desired target population mix. As desirable as random assignment is for evaluation purposes, it may be objectionable to those who seek to have certain individuals or groups in the experimental program and could become an issue at the administrative level. There is no sure solution to conflicts of this type; competing interests must be weighed.

The quasi-experimental design utilizes comparison groups which are not random in their composition but which can be justified as providing legitimate

comparative data. Such an approach may be necessary either as a substitute for or supplement to a true experimental design. Substitution may be required where the applicant pool is not large enough to form both experimental and comparison groups or where administrative considerations preclude randomization. Supplementation may be recommended where hypotheses call for comparisons with identifiable groups which cannot be constituted randomly from the applicant pool. Examples of such groups would be typical high-school students, work-study students and school dropouts.

Whether quasi-experimental groups provide the only comparisons or supplementary comparative data, the groups must be selected with great care to meet the needs of hypothesis testing. Criteria used in selecting the experimental groups must be documented so that any resultant special characteristics can be identified. Comparison groups not differing markedly from the experimental groups, but still permitting the desired comparisons, should be sought. Demographic, cognitive and affective characteristics of all groups should be determined to the degree of completeness possible. The potential effects of initial experimental-comparison group differences on the outcomes to be measured should be estimated to provide a background for interpreting the final results.

12

It must be recognized that the quasi-experimental design allows less confidence in conclusions than the true experimental design since the quasi-experimental groups are more likely to be different at the outset and these differences may be suspected of affecting the evaluation results. Statistical procedures can compensate to a degree, but the design is inherently weaker. Serious consideration should be given to the advantages and disadvantages of the various subject group designs before a selection is made, and administrative, as well as evaluative, consequences should be examined carefully. Once groups have been constituted, changes will not be possible within an experimental year, and they are often difficult between years.

Students are the principal subjects in most educational programs, and the establishment of student groups automatically creates parent groups. Other subject groups in the evaluation may consist of community resources, instructors, potential adopters and others. The specific array of groups necessary is defined by the range of hypotheses to be tested. Success in establishing appropriate subject groups determines the ability to test hypotheses.

INSTRUMENTS

Hypotheses determine what is to be evaluated; subject groups determine the samples with which hypotheses will be tested. The next step is to select instruments which reasonably can be expected to measure the hypothesized program effects. For each hypothesis at least one measure must be selected to represent the intended outcome. Such measures may range from performance on a standardized test to opinions about aspects of the program. Indirect measures such as attendance, assignments completed and frequency of resource use also may be appropriate as criteria for evaluating effects.

A series of instruments, along with scoring and interpretation packages, is available for use with RBS Career Education. These instruments are relevant to the measurement of career skills, life skills, basic skills and participant opinions. Depending on the scope of hypotheses, selection from among these instruments may suffice or additional measures may be needed. The instrumentation materials are described in detail in the *RBS Instrument Service Guide*.

Just as the selection of instrument content must be keyed to project objectives and hypotheses, the schedule of administration must be timed to permit the desired analyses. Some hypotheses may require data gathered at one time only, as with standards of participant opinion, which generally call for survey measurement at some point after participants have had sufficient experience with the program. Hypotheses dealing with growth require measurements from at least two points in time in order to assess change. This approach utilizes a pretest-posttest or repeated measures schedule. It is important to allow enough time between the test administrations so that the desired growth reasonably can be expected to occur.

Hypotheses dealing with comparison groups require a simultaneous test administration after all groups have participated in their respective programs for the specified period of time. This is a posttest-only schedule. If the groups are quasi-experimental, then all groups also must be pretested before the program begins so that initial differences can be taken into consideration. If the groups are true experimental, it still is desirable to pretest in order to enhance precision and minimize the weakening effects of dropouts during the program.

Thus, instrument content must match program objectives. And, in designing the schedule of instrument administration, it is important to provide for the timely collection of data required to test the stated hypotheses.

DATA SYSTEMS

The creation of a data system capable of accommodating all collected information is an important support task in the evaluation process. The absence of a systematic approach to data storage and maintenance greatly increases the occurrence of lost, irretrievable or unusable information. As soon as the evaluation design has been finalized, construction of data systems can be undertaken. The hypotheses, subject groups and instruments all serve to define the parameters of the system which will meet the needs.

The first task is to establish an identification system for members of subject groups. It usually is preferable to employ a numerical system to minimize the recognizability of individuals, except by designated persons who have the translation lists. Subject numbers can be constructed to include group identification, time of program entry or any other variable which might be helpful in file categorization. Whatever the numbering procedure selected, it is important to allow room for group members who may be added in the future and to assure that each subject will have a unique number.

The construction of a numbering framework establishes one dimension of the data system; it identifies the range of individuals across subject groups. The other major dimension is the specification of information to be collected within each subject group. Such data consist of the results obtained from all of the instruments administered to each group and may be in the form of individual item scores, subscale scores, total scores or any combination.

A basic information file might be diagrammed as follows in Figure 1. The first column lists the range of members in each group. The other columns list information and scores obtained for each individual. Most systems will be more complicated than this example because they will include more variables and multiple administrations of instruments, but the diagram may serve as a model which can be expanded.

The codes and formats for storing the data should be selected according to the information needs defined by the evaluation plan. The data system should be designed to facilitate the anticipated analyses by keeping the form and location of all data clear and retrievable for evaluative use.

After the data system has been outlined, the choice of implementing it as a manual or automated system can be made. This decision depends upon both the size of the data files and the complexity of planned analyses. Usually some degree of machine processing capability is desirable, which requires individual file formats

SUBJECT GROUPS	INFORMATION				
EXPERIMENTAL STUDENTS	SEX	AGE	TEST X	TEST Y	TEST Z
101 102 103 ↓ 199	– – – – –	– – – – –	– – – – –	– – – – –	– – – – –
CONTROL STUDENTS	SEX	AGE	TEST X	TEST Y	TEST Z
201 202 203 ↓ 299	– – – – –	– – – – –	– – – – –	– – – – –	– – – – –
COMMUNITY RESOURCES	SIZE	TYPE	OPINION SURVEY		
301 302 303 ↓ 399	– – – – –	– – – – –	—— —— —— —— ——		
PARENTS	OCC.	AGE	OPINION SURVEY		
401 402	– –	– –	—— ——		

FIGURE 1

designed for use with a computer card processing system or other automated procedures.

Designing and implementing the data system is a task requiring technical expertise and experience with the problems which typically are encountered. It should be done with great care and informed advice. Like the other elements in the evaluation process, the quality of the data system directly affects the clarity and usefulness of the evaluation results.

ANALYSES

The definition of hypotheses, subject groups and instruments is needed in the design of an analysis plan. Analyses should be selected to describe the results clearly, to test hypotheses statistically with the most rigor possible and to facilitate unambiguous interpretation of the evaluation outcomes. Hypotheses determine what effects are to be tested. Subject groups constitute the experimental samples among whom effects are hypothesized. Instruments provide measures of the criteria selected to represent the hypothesized effects, and analyses are the statistical techniques which support or deny the existence of effects within the hypothesis framework.

Planning specific analyses depends greatly on the decisions made in previous design stages, but some general guidelines can be suggested. More specific information on analyses is presented in the *RBS Analysis Service Guide.*

The first level of analysis should be descriptive. Appropriate distributional statistics should be displayed for each subject group on all available measures. These data serve to depict group characteristics and suggest between-group differences which may need to be considered.

The next level of analysis is intended to uncover any differences between initially selected groups and the groups available for final analysis. Since the groups were chosen to represent specific target populations, it is necessary to know how they changed in composition over the course of the year. Initial groups will be decreased in size both by attrition from the program and testing absence. It is important to estimate the effects of such reductions in the samples by statistically comparing the subjects remaining for final analysis with those who have been eliminated. These comparisons should include any subject characteristics for which pretest information is available. The results will allow an estimate of the representativeness of posttest data in terms of the initially drawn samples and may suggest subsidiary analyses in the hypothesis testing. Absence of such estimates constitutes a weakness in interpreting results whenever group attrition is substantial.

When an estimation of the representativeness of available data has been provided through procedures such as those just outlined, the final level of analysis may be designed: the testing of hypotheses. Where a criterion or standard of success has been established for a subject group, the group performance mean, or other representative statistic, may be compared directly with the designated standard. If development within groups has been hypothesized, then statistical tests comparing

the pretest and posttest performance levels may be conducted. For hypothesized between-group differences, analyses comparing the performance of the various groups should be carried out. Selection of specific statistical techniques depends upon the nature of the data and the questions posed.

This general flow of analytic procedures provides descriptive information, assessment of data representativeness and testing of the stated hypotheses. The specific elements of the analysis plan should be designed well before the analyses are conducted. This timing is important because the analysis design serves as input for implementing the necessary data systems, and also because unanticipated or unannounced analyses may be viewed as searching for desirable results.

REPORTING PROCEDURES

All of the steps in this evaluation process contribute to the production of evaluation findings. These findings are communicated in reports which should be geared to the audiences that will receive them. Three major audiences can be identified: 1) participants in the program, 2) sponsors of the program or potential sponsors of similar programs and 3) external education and research groups. For each group the pertinent questions and when they need to be answered must be specified so that a schedule of reporting can be designed.

Participants in the program require the most detailed and frequent evaluation reporting. For example, staff will be able to use individual student results in guiding students through the program. Members of any subject group will be interested in overall results for their group. Students will want to know how they scored on achievement tests. Program leaders will want to be alerted to apparent problems. Each of these possible reporting categories requires a timely internal feedback system. Reporting in this sense is an ongoing communication activity, often without much formal interpretation. It serves an important function in supporting the operation and development of the program, but it also necessitates a field test of the data collection, storing and manipulation procedures. Testing these procedures at an early point can be helpful in avoiding problems later.

Sponsors and potential sponsors usually require a different level of information and reporting. They are interested in summary data on progress and outcomes as well as interpretations of the meaning of results. Typically this information calls for a mid-year and year-end report in which the evaluation process is described, results outlined clearly and concisely and outcomes interpreted in terms of program success and recommended future direction. Such reports also will be of interest to the program participants.

Often it is valuable to prepare reports for external groups. Portraying the program at a general level would be useful at regional, state or national educational forums. Groups implementing similar programs may be given assistance through reports on problems encountered and solutions found. Research and evaluation audiences might be interested in reports on technical issues and research significance. Such reporting must be designed to meet the needs of the particular audience.

Reporting is the final stage of the evaluation process. In many senses the report is the culmination of that process since all of the preceding stages combine to generate it. Reporting is the evaluation product. As such, it should be planned carefully to utilize available data to their fullest.

RESOURCES REQUIRED FOR EVALUATION

This manual has presented an outline of the evaluation planning activities which are recommended for experimental or demonstrational programs. A final topic concerns the resources necessary for designing and conducting a worthwhile evaluation. Needed resources vary with the scope of the program objectives, numbers and sizes of subject groups and the complexity of analyses planned. For this reason projections must be fairly general, with substantial room for adjustment to meet local needs.

Although the preceding sections of this manual have dealt primarily with evaluation planning rather than implementation issues and problems, resource estimates for both planning and implementation are included here. Implementation estimates are provided because such costs are generally a planning concern.

In order to establish some basis for resource projections, a hypothetical career education program will be used. In this illustration it is assumed that approximately 100 students, equally divided between experimental and control groups, are to be included in the evaluation. These student samples would create parent groups totaling at least 100 members. Since this program is to utilize community-based career education experiences, approximately 50 resource sites with a total of 100 key site personnel would participate in the study.

The program objectives are assumed to focus on the development of career skills, life skills and basic academic skills. One major testing instrument is to be employed in each skill area along with a student background questionnaire. All participants also will be administered an opinion survey. The skills tests will be administered on a pretest-posttest schedule; the opinion surveys will be given only once during the year.

Systematic feedback of evaluation results to program staff would be available, as would automated instrument scoring and a computer-based data system. Progress of the experimental group in each skill area over the course of the year will be analyzed, and the superiority or inferiority of the experimental group relative to the control group will be assessed. Opinions, perceptions and suggestions of program participants are to be documented. Standard statistical procedures will be used for analysis purposes; all results will be presented in evaluation reports.

It is assumed that the services of a trained and experienced evaluator will be available locally to accomplish most of the tasks. External evaluation technical assistance and services are projected to facilitate major steps in the evaluation process.

Given this hypothetical example, a generalized allocation of resources may be projected for evaluation planning and evaluation implementation. Figure 2 presents projects for the planning process.

FIGURE 2

ESTIMATED EVALUATION PLANNING RESOURCES

Task Area	Staff Days	Technical Assistance Days
1. Program Definition	1 - 2	
2. Evaluation Questions	3 - 4	1 - 2
3. Statement of Hypotheses	1 - 2	
4. Subject Groups	4 - 5	1 - 2
5. Instruments	2 - 3	1 - 2
6. Data Systems	5 - 6	1 - 2
7. Analyses	4 - 5	2 - 3
8. Reporting Procedures	2 - 3	0 - 1
	22-30	6-12

Estimates of time requirements are included for each evaluation planning task. The "staff days" refer to the program evaluator, and "technical assistance days" denote consulting services from an external agency such as Research for Better Schools. Participation of other program staff has not been accounted for but will be necessary according to local needs. Support services and non-staff resources likewise have not been calculated because they are dependent on local conditions and usually can be readily extrapolated from the staff costs.

With these qualifications, it is estimated that the evaluation planning resources needed should approximate 22-30 staff days and 6-12 technical assistance days. These resource requirements are affected by the extensiveness of the program, but they increase at a less than proportional rate. The planning and design tasks for a 200-student program are not much different from the tasks for a 100-student program. In this sense, planning costs are much less variable than implementation costs.

Figure 3 presents projections for evaluation implementation that are based on the hypothetical program; changes in conditions would have a directly proportional effect on resources needed.

FIGURE 3

ESTIMATED EVALUATION IMPLEMENTATION RESOURCES

Task Areas	Staff Days	Technical Assistance Days	Other Costs
1. Subject Groups – implement, maintain	3 - 4	0 - 1	–
2. Instruments – purchase, administer, score	10 - 12	1 - 2	$1200 instruments and scoring
3. Data Systems – implement, maintain	9 - 10	2 - 3	$ 300 computer services
4. Analyses – perform, interpret	17 - 20	3 - 4	$ 600 computer services
5. Reporting Procedures – prepare feedback, interim, final and other reports	20 - 40	3 - 4	$ 600 production
	59 - 86	9 - 14	$2700

These accumulated implementation estimates suggest the need for an approximately one-third-time staff evaluator supported by 9-14 outside technical assistance days and $2700 in other resources. Technical assistance, scoring, computer and production services are available from Research for Better Schools. These resource estimates do not include support services, physical facilities, supplies and materials, postage and other non-staff costs.

It should be emphasized that these resource allocations are generalized estimates. More precise projections would require planning information specific to the individual project to be evaluated. Expansion of the program objectives, student groups or intended analyses beyond the hypothetical example used for these projections would necessitate proportionate increases in evaluation resources.

CONCLUDING NOTES

The creation of an evaluation plan requires a crucial and technically demanding set of tasks in the evaluation process. The scope and sophistication of the plan do much to determine the usefulness and conclusiveness of the evaluation findings. This manual has attempted to discuss evaluation planning in a concise but comprehensive way by structuring a series of major sequential steps. In this final section, the steps will be capsulized and several overall concerns will be noted.

The planning process begins with defining the program to be evaluated, formulating evaluation questions and refining the questions into testable hypotheses. These steps establish the evaluation needs and formally state expectations for the program. Next, the selection of subject groups and instruments enables the testing of hypotheses by specifying the effect variables and the samples among whom effects are intended. Finally, data systems must be designed to accommodate the evaluation information, and an analysis plan must be developed to show how the hypotheses will be tested.

Although these steps can be discussed as separate stages in the evaluation planning process, their interrelatedness should not be minimized. The decisions made at each stage strongly influence the requirements of succeeding stages. Likewise, difficulties in later stages may call for revisions at earlier points. Changes in the program or evaluation components which affect one stage necessitate a review of the entire process to ensure consistency. The activities within evaluation planning thus are interdependent and must be conducted with that perspective.

The evaluation planning process is sufficiently complex to benefit from advisory assistance; an external review is always appropriate. Omissions, errors of judgment and inconsistencies in evaluation planning generally are magnified and harder to correct during evaluation implementation. Weaknesses in the design become limitations in the usefulness and interpretability of the findings.

The objectives which are to be accomplished by an evaluation effort vary substantially from program to program. The intended role of evaluation may range from simply completing a funding requirement to providing extensive information in the operation and development of the program. The evaluation planner may help to shape that role and must be aware of the real evaluation objectives as they relate to the program as a whole. It is important to have this awareness in the planning process in order to maximize the usefulness of evaluation results.

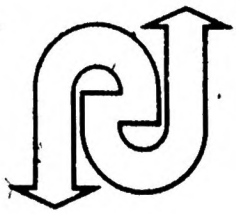
The evaluation planner also must be able to deal with non-evaluation factors

which affect evaluation design. Other program activities or decisions may alter or impede evaluation plans, and in these cases flexibility and creative problem-solving skills are required to adapt to the environment while maintaining the integrity of the evaluation effort.

The interpretation of results is a planning issue which typically receives insufficient attention. A complete evaluation plan should project how major alternative outcomes will be interpreted should they be obtained. This exercise in projection accomplishes two important aims. It uncovers possible findings which would be uninterpretable and may call for redesign. It also establishes the potential significance of the results. Thus, interpretation is a planning as well as an implementation concern.

After a satisfactory evaluation plan has been developed, its implementation can proceed. Implementation introduces a whole series of problems and issues which could not be addressed in this manual. Even the best plans have limited value if they are not rigorously implemented. Successful implementation turns the potential of the design into reality. If evaluation planning and implementation are accorded proper attention, the probability of obtaining conclusive, unambiguous and pertinent results will be greatly enhanced.

The publication of RBS CAREER EDUCATION: EVALUATION PLANNING MANUAL has been funded in part by the Vocational Education Act (P.L. 8910, Title IV) as amended in 1968, in cooperation with the National Institute of Education, under research contract #NE-C-004--0011. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.



Research for Better Schools, Inc.
1700 Market Street, Suite 1700
Philadelphia, Pennsylvania 19103
Robert G. Scanlon, Executive Director