DOCUMENT RESUME

ED 181 343 CG 014 055

AUTHOR Jung, Steven M.
TITLE Trying Out Activities and Monitoring Early Implementation Efforts. Module 10.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Jul 78
NOTE 78p.; For related documents see CG 014 049-057, ED 124 823 and ED 133 511. Reprint.
AVAILABLE FROM National Consortium Project, American Institutes for Research, P.O. Box 1113, Palo Alto, CA 94302 ($3.20)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Career Development; Counselors; *Evaluation Criteria; Evaluation Methods; Guidance Personnel; Inservice Programs; *Occupational Guidance; *Performance Factors; Professional Development; Program Design; *Program Development; *Program Effectiveness; Secondary Education

ABSTRACT

This staff development module is part of one of three groups of career guidance modules developed, field-tested, and revised by a six-state consortium coordinated by the American Institutes for Research. This module is the tenth in a series on developing a comprehensive career guidance program at the high school level, designed to aid guidance personnel responsible for developing student-focused programs. The goal of this module is to convince users of the value of the empirical approach to program development, and to impart those skills required to conduct and measure the effects of activity tryouts and early implementation efforts. The module format consists of an overview, goals, objectives, outline, time schedule, glossary, readings, skill development activities, and bibliography. A Coordinator's Guide is also included with detailed instructions for presenting the module in a workshop setting as well as the facilitator's roles and functions, and the criteria used in assessing the participants' achievement of module objectives. (Author/HLM)

MODULE 10

TRYING OUT ACTIVITIES AND MONITORING EARLY IMPLEMENTATION EFFORTS

Steven M. Jung

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

August 1975

Reprinted with minor revisions, July 1978

ISBN 0-89785-574-4

Illustrations in this module are by Jurgen Wolff. Editorial work is by Charles Dayton and Phyllis DuBois.

This module builds on cooperative staff development activities of:

Mesa Public Schools; Mesa Arizona,
Fremont Union High School District; Sunnyvale, California
Santa Clara County Superintendent of Schools Office; San Jose, California, and
American Institutes for Research; Palo Alto, California

Support for these efforts was received through the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963 and Title III of the Elementary and Secondary Education Act of 1965. Points of view or opinions stated do not necessarily represent U.S.O.E. position or policy.

TABLE OF CONTENTS

Page

INTRODUCTION

Module Goal and Outcomes ---------- 1

Module Outline ---------- 2

Model and Module 10 ---------- 3

Glossary ---------- 5

TEXT

Setting the Stage ---------- 6

When and when not to pilot test ---------- 7

Discussion on the concept of "subset optimization" ---- 10

Planning the General Strategy of Pilot Testing ---------- 12

Before the pilot test ---------- 12

Preparing to pilot test ---------- 13

Discussion on a general strategy of pilot testing ----- 16

Activity 1 - Analyzing a program's internal logic ----- 20

Estimating Costs ---------- 21

Activity 2 - Estimating Costs ---------- 23

Monitoring During Early Implementation Efforts ---------- 26

Measurement Techniques ---------- 27

The four basic techniques ---------- 30

Considerations of adequacy ---------- 32

Activity 3 - "Really understanding" measurement and measurement techniques ---------- 40

Results of Preliminary Tryouts and Early Implementation --- 44

Summary ---------- 45

Discussion Questions ---------- 46

POSTASSESSMENT ---------- 47

APPLICATION ---------- 48

APPENDIX

Optional Group Simulation Description ---------- 51

References ---------- 53

COORDINATOR'S GUIDE

INTRODUCTION

MODULE GOAL AND OUTCOMES

Module Goal

To convince you of the value of the empirical approach to program development and to impart certain skills required to conduct and measure the effects of activity tryouts and early implementation efforts.

Module Outcomes*

Following completion of the learning activities of this module, you should be able to perform the following:

1. State the primary purpose of carrying out preliminary activity tryouts and describe at least two situations in which such pilot testing is useful and two in which it is a waste of time.
2. Verify the internal logic of a planned career guidance activity, given a written description of such an activity covering its goals, student performance objectives, and process objectives.
3. Develop a relatively accurate estimate of the cost of implementing a career guidance activity.
4. Develop measures of attainment of the *process* objectives of a career guidance activity, given a written description of the activity and its process objectives.
5. Develop relatively objective, reliable, and valid measures of the *outcomes* of a career guidance activity.
6. Develop a plan for trying out activities and monitoring early implementation efforts at your school.

*The criteria for attainment of the outcomes are available in the Coordinator's Guide for Module 10.

MODULE OUTLINE

Approximate Time	Activity	Outcomes
15 minutes	Introduction The Coordinator will explain the basic purpose of this module. Several additional textbooks are referenced herein and may be used for extended learning. The basic outcomes of the module, however, can be attained without these texts.	
3 hours	Text Presentation of important information and opportunities to practice skills related to activity tryouts and early implementation monitoring.	1-7
½ hour	Postassessment Assessment of your acquired knowledge and skills.	1-7
1 hour	Application Planning for your own tryouts and monitoring efforts.	8

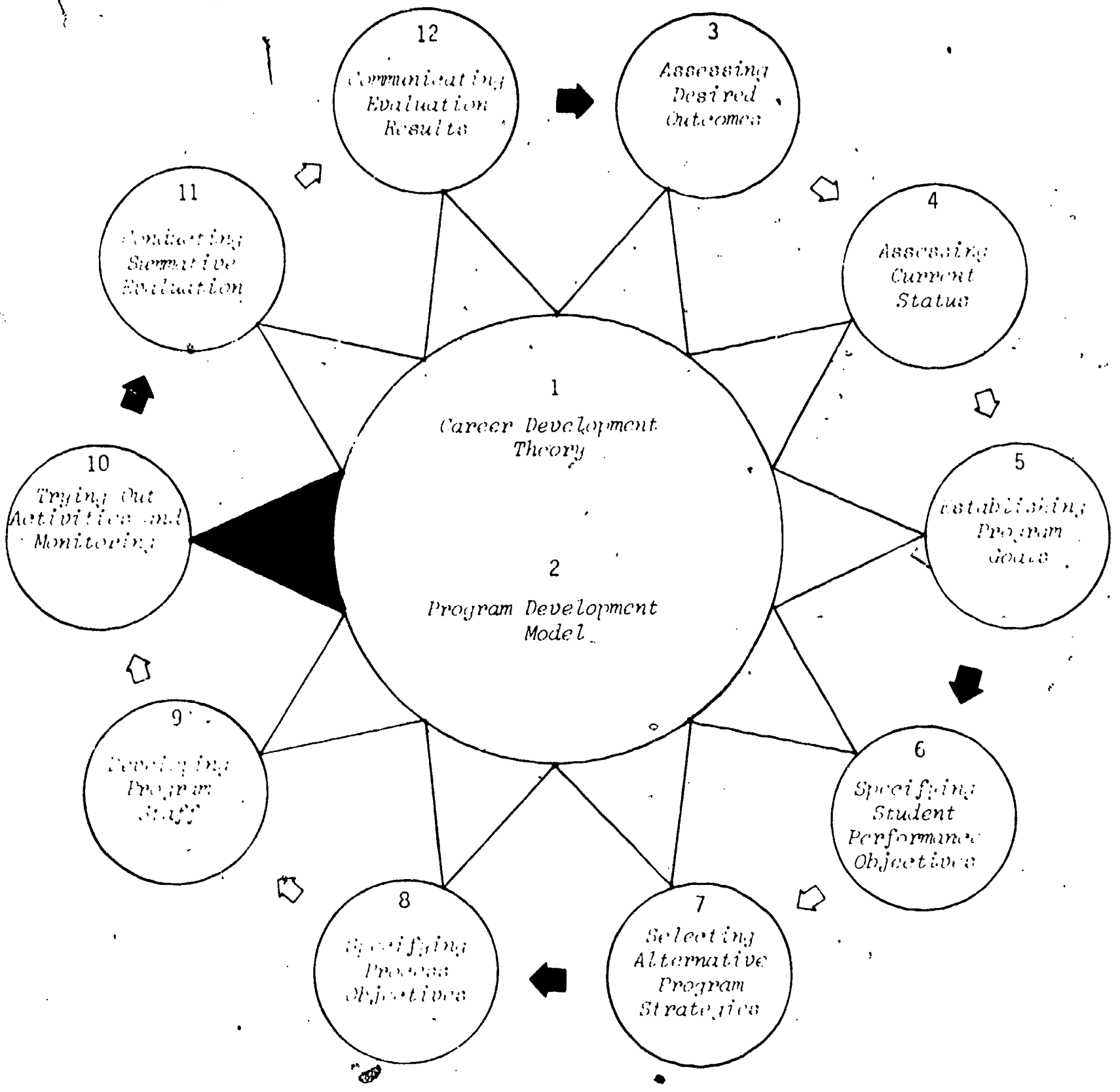
MODEL AND MODULE 10

The model on the next page shows the relationship among the modules. *Trying Out Activities and Monitoring Early Implementation Efforts* comes at the end of Phase 3 and is Module 10.

This module deals with the tasks important in the early stages of actual implementation. Thus, it builds on the process objectives dealt with in Module 8, while foreshadowing the summative evaluation to come later. It is the bridge between looking forward and looking back, between preparing and assessing results.

A Model for Developing

Comprehensive Career Guidance Programs



GLOSSARY

Efficiency - A property of a measurement which refers to the degree to which the technique is inexpensive, easy to administer, and easy to score.

Evaluation design - An arrangement of persons, an activity, and measures of effect so that inferences can be made about the probable effects of the activity.

Formative evaluation - A process of collecting and using information during program development in order to improve the functioning of the program.

Monitoring - A process of conducting data collection activities during early program implementation to permit immediate revisions in the program.

Non-reactivity - A property of a measurement technique which refers to the absence of undue influence upon the person to whom it is applied.

Objectivity - A property of a measurement technique which refers to the degree to which the technique produces the same score regardless of who applies it.

Pilot testing - A process of conducting limited tryouts of specific activities that have been tentatively adopted as a result of planning and design procedures.

Reliability - A property of a measurement technique which refers to the consistency of its scores and their relative freedom from chance variation over time.

Subset optimization - Dr. Richard Schutz's phrase for ensuring that subcomponents of a program are performing optimally.

Summative evaluation - A process of collecting information to facilitate judgments about the overall worth of a program; especially appropriate to later implementation stages.

Validity - A property of a measurement technique which refers to the degree to which the technique actually measures what it is supposed to measure, intuitively and by empirical demonstration, thus yielding a relatively "true" score.

SETTING THE STAGE

A. *Definitions*

1. *Formative evaluation*
2. *Pilot testing*
3. *Monitoring*

B. *When and when not to pilot test*
Four scales for rating activities

As you have no doubt learned from earlier modules, two of our basic themes are continuously trying out activities and making modifications based on obtained results. These themes reflect a basic commitment to the virtues of continual "formative" evaluation, a term coined by Professor Michael Scriven (1967) to refer to a process of collecting and using information to improve the functioning of educational programs. The concept of formative evaluation is the subject of an article by Garth Sorenson that you may wish to read. It is included in the Appendix of this module.

Formulative Evaluation Defined

In this module, we are interested in application of formative evaluation procedures to a very specific *stage* in the program planning and development process: the stage which *follows* the determination of goals, objectives, and intervention strategies and *precedes* any summative evaluation of these strategies. This stage includes pilot testing and monitoring early implementation efforts.

Pilot Testing Defined

Pilot tests are limited tryouts of specific activities which have been tentatively adopted as a result of the previously discussed planning and design procedures. These tryouts should usually *precede* larger scale (i.e., schoolwide or systemwide) program implementation efforts. However, because of time or money limitations, it may be necessary to proceed to actual program implementation efforts without pilot tests of component activities.

During early implementation efforts, in which all components of the career guidance program are being tried out in concert for the first time, it is important to engage in data-collection activities designed to permit immediate revisions. Such activities have been given the term "monitoring" in this module. Summative evaluation, designed to facilitate judgments about the *overall* worth of the program, should not begin until the program components have had an opportunity for revision based on monitoring results. If summative evaluation is attempted prematurely, because of program accountability demands for administrators, for example, it is possible that correctable flaws in design or implementation will cause premature termination of the program.

Monitoring Defined

When and When Not to Pilot Test

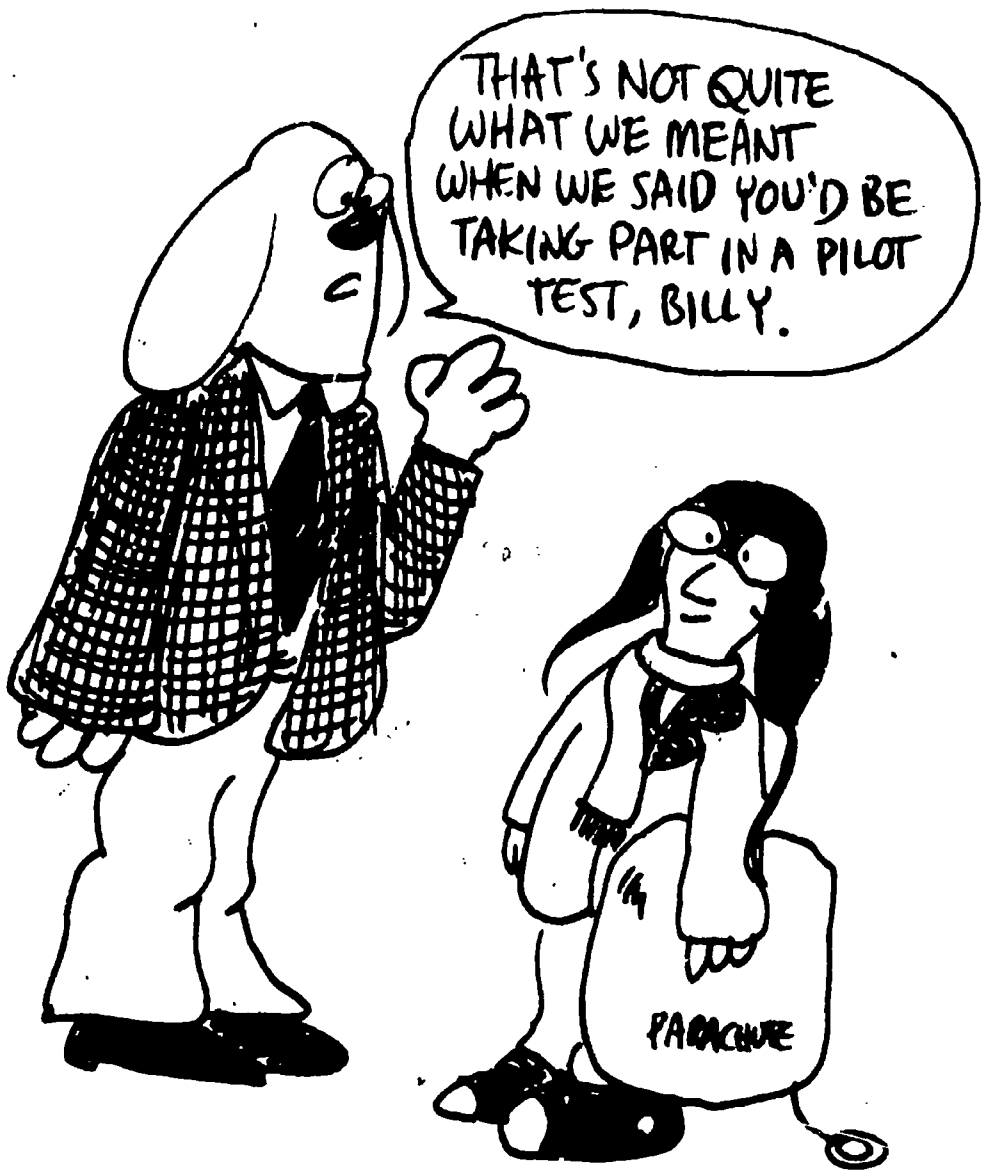
Pilot testing does require time and money which might well be devoted to other pursuits. Moreover, pilot tests rarely result in dramatic improvements for activities which are already well conceived and planned; smaller, incremental improvements can be expected.

Thus, pilot testing of planned activities is *not* always desirable. If an activity is *not* intended to be *replicable*, there is little use for preliminary tryout and formative evaluation. In other words, one-shot events, or activities which depend for their implementation on unique situations, spontaneity, etc., probably should not be pilot tested. A good example of this is the "Career Night" type of activity.

Don't pilot test if activity is not replicable

Similarly, activities which are already in their final form and for which there is *no intent of revision*, should not undergo pilot testing. These would normally be traditional activities whose substance is dictated by precedent rather than by intent to produce prespecified student outcomes, e.g., an initial orientation assembly.

or not revisable



Activities which should receive *highest* priority in conducting pilot tests are those which are *most crucial* to the success of the overall career guidance program and about which there is the *most uncertainty* in terms of desired outcome attainment.

Do pilot test activities which are most crucial and will produce uncertain effects

In determining whether or not to conduct empirical pilot tests, it might be useful to *rate* the activities you have selected for implementation on four simple scales such as the following:

Four scales for rating activities

REPLICABILITY

1	2	3
There is no intent to repeat this activity.		If this activity proves successful, it will be used repeatedly in the future.

REVISABILITY

1	2	3
The structure of this activity is inviolate.		There is a strong desire to modify and improve this activity.

IMPORTANCE

1	2	3
The success of this activity is not crucial to the success of the career guidance program.		The success of this activity is absolutely essential to the success of the career guidance program.

UNCERTAINTY

1	2	3
There is no question about the potential effects of this activity.		There is considerable doubt about the potential effects of this activity.

Sum the ratings and rank order the sums. Activities which score highest should receive whatever time and financial resources are available for pilot testing.

DISCUSSION ON THE CONCEPT OF "SUBSET OPTIMIZATION"

Dr. Richard Schutz coined the phrase "subset optimization" in a 1970 address to the American Educational Research Association.

> The final point involves the distinction between comparative and cumulative experimentation. The experimental tradition in the behavioral sciences is comparative. One compares effects of different phenomena introduced concurrently or simultaneously. An equally venerable experimental tradition involves comparisons over time which cumulate in more optimal performance. This tradition has often been rejected in education because of its industrial connotations. Cumulative optimization methodology can, however, be applied to educational endeavors without considering people as machines, just as comparative experiment methodology may be applied without considering people as fertilizer.
>
> The point revolves around the concept of *subset optimization*. It is both reasonable and necessary to use subset criteria. For example, although specified changes in pupil behavior represent the ultimate criterion which instructional development is attempting to optimize, it is very unwise to use this as the sole feedback basis. For example, our staff have, at times, been very disappointed when specific procedures such as teacher training, audiovisual segments, etc., have not improved pupil performance. One could throw out the prototype and look in a different area. But on closer examination and analysis, we have each time determined that the intended function was itself not being performed by the instructional components. That is, the teachers learned nothing or the wrong things from the training, the audiovisual segments were being "misused", etc. With subcomponents performing optimally, one has a much better likelihood of accomplishing larger functions optimally. Optimization is unlikely achievable with unreliable subcomponents. This sounds obvious, but it is almost universally overlooked in education.[1]

---

[1] Richard Schutz, "Programmatic Instructional Development", Paper delivered at the 55th annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March 5, 1970.

Do you agree with the last sentence above? Can you cite examples from your own experience where (1) the improper functioning of some component of a program has detracted from the effects of the program as a whole; and/or (2) the improvement of some component of a program has improved overall program functioning? If you can cite examples, please list them on the blackboard and discuss the actual circumstances with others in the group.

PLANNING THE GENERAL STRATEGY OF PILOT TESTING

A. *Reviewing internal logic*
B. *Preparing pilot test*
   1. *Selecting a sample*
   2. *Designing evaluation*

Before the Pilot Test

Review Internal Logic

Prior to pilot testing, it is almost always useful to perform a brief *review* of the *internal logic* of planned activities. Assuming that program goals, student performance objectives, and process objectives have been developed in accordance with techniques suggested in preceding modules, this review can be relatively simple and straightforward. It should include the following questions, roughly in this sequence:

1. Do the planned activities *relate* to the intended student performance objectives? Is there a reasonable chance that the activity, if carried out as indicated, will produce the desired outcomes?
2. Do the proposed student performance objectives *relate* to the program's stated goal or goals?
3. Is there reason to believe that most members of the target audience can *already perform* the student performance objectives, making the planned activities unnecessary?
4. Are the activities described specifically enough to be observed? Can their attainment be *documented?*
5. Is there an indicated *time schedule* or sequence for multiple activities?
6. Are the planned activities and sequence *practical* with regard to the constraints likely to exist in this situation?

7. Are the planned activities and sequence *appropriate* to the level of the target audience?
8. Are there likely to be *side effects* which developers have probably not anticipated?

From the above, it may be seen that initial pilot testing precedes actual empirical tryouts. Negative answers to the above questions should be pursued with designers until a satisfactory positive response can be identified.

Preparing to Pilot Test

Normally, early tryouts of activities should be low cost, involving few students and abbreviated time schedules. Although these limitations are naturally relative to the scale of the activity in question, tryouts requiring more than ten classes and/or five days are extremely rare. The *general strategy* of pilot testing is to implement the planned activity under well-controlled circumstances, using members of the target population, collecting (1) objective[2] information on degree of attainment of process and student performance objectives, and (2) any other objective or subjective information which is likely to be useful in indicating changes that *can* potentially be made in the activity. Care must be taken to avoid the tendency to overcollect data at this stage, since the range of available modifications is likely to be limited, and data-processing and analysis procedures have to be rapid so that immediate feedback can be obtained.

A General Strategy of Pilot Testing

"Less is usually more"

[2] By *objective* in this context, we mean a property of outcome measures, to be explained later, which refers to the degree to which a test or scale produces the same score rega dless of who applies it. This is not to be confused with *objective* in the context of a statement of desired outcomes.

Select a sample for the tryout. In carrying out a pilot test of the selected activities, you are interested in *inferring* from the results of the tryout the probable results of actual large-scale implementation with your target population. Therefore, it is essential to select persons for the tryout who are at least broadly representative of the target population. In practice, this often means identifying persons who may be predicted, because of past performance, to do exceptionally well, poorly, and average on the activity. Strict *random sampling* procedures, in which every person in the target population has an equal chance of being selected for the tryout, are rarely used because of the low probability that small random samples will contain representatives of the extremes of the characteristics (ability, experience, motivation, etc.) which are presumed to cause above- or below-average performance. When reporting on the results of a tryout, it is important to *describe* the characteristics (age, sex, prior experience and achievement levels, grade, etc.) which were used in identifying the tryout sample.

Strive for Broad Representation of Target Population

Evaluation design. The word *design* in this context

Evaluation Design Defined

refers to an arrangement of *persons*, an *activity*, and *measures of effects* so that inferences can be made about the probable effects of the activity on a larger group of similar persons. Design problems can quickly become very complex, especially in educational experiments which are conducted to demonstrate the relative effectiveness of competitive experiments. Comparative experimental evaluation designs, however, have no place at this particular stage of the program planning and development process. Two so-called "pre-experimental" evaluation designs (Campbell and Stanley, 1966) are normally sufficient for the inferential needs of this stage. These are the "one-group post-test only" design and the "one-group pre-test-post-test" design. These may be diagrammed as follows, where X stands for implementation of the activity and O stands for outcome measurement.

One-group post-test only design	X O
One-group pre-test-post-test design	O_1 X O_2

The major tradeoff in selecting the latter design over the former is the value of having pre-implementation information versus the danger of *producing* effects solely through administration of the pre-test. The latter dangers can be reduced by using *nonreactive* measures, which are discussed in the next sub-section. Thus, the one-group pre-test-post-test design is recommended for most tryout applications.

The "Best" Design

A DISCUSSION ON A GENERAL STRATEGY OF PILOT TESTING

Please read the following fables.

(1) In May, a team of developers at Research Associates, Inc. was given a contract to prepare an innovative career education curriculum for seventh graders, on the understanding that the developers would be able to deliver the following March. The team members got to work at once, but since summer was over by the time they finished the materials and developed the accompanying book and A-V support materials, the team had everything produced and began a full field test in August without bothering to try the components out first. When results were finally in, the field test unfortunately revealed a number of serious problems. The revisions required were so extensive that the materials could not be completed by the deadline, and it was therefore the following autumn before the program could actually be implemented in the schools.

It is not certain, but it is at least possible, that a tryout would have revealed some of the problems before so much work had been done.

Moral: Shortcuts can produce long delays.

(2) In response to a directive from the school board to emphasize career education, a group of teachers at Urban High School got together to develop a career handbook. They carefully specified their objectives, chose instructional techniques, and tested their book with typical students when they were done. The students liked it, and testing showed that they were able to master the material.

However, when those same students got out of high school and began looking for jobs, they discovered that the employment picture was vastly different from that described by the handbook. When they revisited the high school and told the teachers who had developed it, the developers were extremely shocked and puzzled. What happened?

Admittedly employment is an area in which changes occur almost daily; however, the developers failed to consult experts in the field while they were designing the handbook, and they did not have it reviewed by people conversant with the real situation after they had prepared it.

Moral: Unless you are the One Great Expert on the material you are presenting, seek an expert opinion on the appropriateness of your activity's goals and objectives before pilot testing. It is a truism that if the goals of an activity aren't worth achieving, then it is uninteresting how well they are achieved.

(3) A Career Education textbook writer recently explained that he did indeed try out his book before publication. He had several teachers use it in their classrooms for a semester and then report their reactions. When asked what aspects of the book he was interested in evaluating, he replied, "None in particular. I was interested in reactions in general."

While this approach is better than nothing, that is all it is. Ideally, a tryout should be predicated on very clear notions of what aspects of the activity are being examined. Only in this way is it possible to insure that important elements are indeed evaluated.

Moral: If all you are interested in is "reactions in general," don't be surprised if your pilot testing yields nothing specific.

(4) Another team at Research Associates, Inc. had developed a career education leader training program with a short film illustrating several types of leader/group interactions. The team had carefully prepared a pre-/post-test to measure learning, and the film was shown to a small number of persons from the intended target audience. When the tests were administered and scored, the team found that most of the subjects, while performing at about the "chance" level on the pre-test, got about two-thirds of the items right on the post-test. They were still discussing the implications of this when one of the team happened to hear a participant remark, "You know, that film was a nice idea, but we all thought the narration was a bore."

The staff made a solemn vow at their next meeting that from then on they would always include an instrument to collect information on learner reactions in their assessement.

Moral: Anticipate "unanticipated" side effects and try to measure them during tryouts.

(5) Stan Binet was monitoring the tryout of a self-instructional booklet intended to help teachers improve their ability to construct tests. Since it was summer vacation, he had been able to find a group of experienced teachers to use as subjects, and for convenience he had asked them to come to his office and work in a room set aside for that purpose. When the tryout began, Stan found that the teachers spent a lot of their time drinking coffee and talking to each other, and their comments on the materials seemed to be a consensus rather than a collection of individual reactions. After he inadvertently mentioned that he had helped write the booklet, the severity of their criticism decreased perceptibly.

Moral: Try to control the circumstances of the pilot test, especially instruction time. Use common sense.

(6) The Humpty Dumpty Toy Company was working on a children's game designed to improve reasoning ability. The game was tried out with five groups of children of different levels of reasoning ability, and trained observers noted the interactions and reactions of each child as he or she played. The observers accumulated five folders full of new data, in addition to pre-/post-tests on reasoning ability. When the tryout was over, the data were delivered to the staff member assigned to revise the game. She spent the month analyzing and introspecting about the observation data and produced a 79-page pilot test report. The report arrived in the Humpty Dumpty offices the day after a major competitor unveiled a similar game in nationwide Saturday morning TV advertisements.

Moral: Pilot tests are not efficient if they consume too much time or resources.

These fables are presented to illustrate some of the important aspects of a *general strategy* of pilot testing. Try to derive such a general strategy from them and write it down in the space below; then compare answers with members of your group.

Discuss these questions with the members of your group:

What types of activities seem most appropriate for pilot testing? Can one-to-one counseling be "pilot tested?" "Group counseling?" Why or why not?

24

ACTIVITY 1 - ANALYZING A PROGRAM'S INTERNAL LOGIC

Please read the description below which contains a brief written account of a planned career guidance career program, covering one goal, two student performance objectives, and two process objectives. Then construct a written analysis of the internal logic of this program, listing factors discussed in the text and presenting your judgment of the program's probable effectiveness with regard to each factor.

The guidance staff of Fliver Junior High School is considering the implementation of a new career guidance program for ninth graders.

The primary goal of the program is to increase students' awareness of post-high school vocational training programs which are available in Fliverville and to increase their knowledge of the minimum entry requirements for each training program.

Two student performance objectives of the program are as follows:

1. Following the program, each full-term participant will be able to describe the post-high school vocational training programs available in Fliverville.
2. Following the program, each full-term participant will be able to describe the minimum entry requirements for all of the post-high school vocational training programs in Fliverville.

To carry out a program designed to help each participating student meet these objectives, the guidance staff intend to implement the following activities:

1. A special study carrel will be opened in the guidance office containing copies of Lovejoy's *College Guide* and the *Occupational Outlook Handbook*. The availability of the carrel and its contents will be announced to all ninth graders during the first assembly of the school year.
2. Prior to the Christmas vacation, all ninth graders will be given a mimeographed handbook on vocational training programs in Fliverville. The handbook will contain a two-page section on each of seven available programs listing their goals, current enrollment, minumum entry requirements, length, and any available information on the success currently being encountered by graduates of the program. A written certification will be required from the parent or guardian of each ninth grader to the effect that the handbook has been brought home and noted by the parent or guardian; students who have not returned a certification by January 31 will be called to the guidance office for a special meeting to go over the handbook.

ESTIMATING COSTS

A. *Five basic cost categories*

B. *Calculation of rates*

Importance of Cost Estimates

One of the most useful purposes a pilot test can serve is to provide more accurate estimates of the likely full-scale implementation costs of an activity. As is pointed out in more detail in the summative evaluation module (Module 11), program outcome data are almost always considered in a cost-effectiveness *context* in making decisions about overall program worth. It is not unrealistic to expect preliminary tryouts to point out activities whose cost-to-effectiveness ratio is not likely to be judged favorably by decision makers. Steps can then be taken to either increase (or provide more convincing measures of) effectiveness or reduce costs if large-scale implementation is still intended.

Cost estimation proceeds in several steps, each of which can be made more simple if school or district accounting procedures allow for the central storage and retrieval of activity costs. The suggested steps are as follows.

Five Basic Cost Categories

First, establish some general expense categories for which costs can reasonably be recorded, such as the following:

1. Personnel costs, including salaries and fringe benefits for administrators, counselors, teachers, specialists, consultants, etc. who are involved in directly implementing the activities
2. Instructional equipment and materials used in implementing the activity
3. Other program expenses (e.g., special postage, data-processing costs, etc.)
4. Building costs (a proportion of total building use and maintenance costs)
5. Fixed administrative charges (a proportion of general administrative costs)

Second, record all costs expended by the activity during the span of its tryout. Third, sum category costs and try to calculate a daily or weekly unit *rate* for each category or subcategory (e.g., teacher days, counselor days, etc.). These rates can then be used for roughly projecting costs of full-scale implementation.

Rate Calculation

ACTIVITY 2 - ESTIMATING COSTS

Your school last year decided to use one classroom as a career center to house career related materials, hold classes, show films, have discussion groups, and so on. You have been asked to determine the economic feasibility of the effort; i.e., specifically, to determine the expense for the first year (10 months) of operation. You have the following information to go on and can use your judgment to make estimates where necessary. The worksheet on the following page should prove helpful in making the estimates.

Personnel costs for the effort included:

- One paraprofessional at $5.00/hour, six hours a day, five days a week, for a 40-week school year, to serve as a librarian/ resource person.
- One day per month of a school principal's time, at $20,000/year, to overview and administer the center.
- One day per month of a consultant's time, at $25,000/year, to plan for the materials to be used and help evaluate the center's effectiveness.
- An hour per week of 10 teachers' time, at an average salary of $12,000/year, taken from other instruction, to devote to career development for their classes.
- Five hours per week of a counselor's time, at an average salary of $12,000/year, to provide instruction and lead discussions.
- Volunteered time from community residents, school graduates, and students to help with particular presentations, talks, and locally relevant programs.
- One month of a secretary's time, at $9,000/year.

Equipment and materials costs included:

- One projector, at $600.
- One filmstrip projector, at $200.

Activity Cost-Estimate Worksheet

Cost Category	No. of Units	Rate/Unit	Estimated Cost
1. Personnel			
2. Equipment/Supplies			
3. Other Program Expenses			
4. Building Costs			
5. Fixed Costs			

- One screen, at $50.
- Two portable cassette tape decks, at $100 each.
- A supply of career related games and assorted materials, at $200.
- Books and printed materials, at $1,000.
- General supplies (paper, pencils, etc.), at $200 for the year.

The cost of the above seven items may be considered to be spread across ten years, so that only 1/10 of the total needs to be considered part of the first year's expenses

Other Program Expenses:

- Cost of mailing a letter to all district parents to let them know about the center. There are approximately 2,000 such parents.
- Cost of using a two-page hand-scored feedback form to collect users' reactions to the center and suggestions for improvement. There were approximately 1,000 users the first year.
- Cost of writing a report at the end of the year on the center's effectiveness. Paper and printing costs only need to be figured here. The report was approximately 20 pages long, and 100 copies were printed.

For all of the above, figure about $.03 per printed or xeroxed page.

Building Costs:

- A general rule of $1,000 per classroom per year is the means used to estimate building costs.

Fixed Administrative Charges:

- This is figured as a 5% additional charge added to the total of all of the above categories.

MONITORING DURING EARLY IMPLEMENTATION EFFORTS

An adequate program monitoring system should continuously provide program directors and their supervisors with data on (1) the degree to which planned activities are being implemented as intended, (2) the degree to which desired outcomes are occurring as intended, and (3) any unplanned outcomes (i.e., positive and negative side effects).

Feedback

An essential element in planning a monitoring system is the "feedback" mechanism, by which process, outcome, and side-effect data are transmitted immediately to program staff who have the authority to make modifications in the program. As in the case with pilot tests, this means immediate processing of results in relatively simple and uncomplicted formats.

In order to facilitate this feedback, it is necessary to plan in advance for the administration of process and outcome measures, selected or developed according to specifications which are described in the next section of this module.

Similarity to Pilot Testing

Because the concepts of early program monitoring so closely match those of pilot testing activities, differing mainly in the degree to which activities are separated in time, no additional discussion will be presented in the text at this time.

MEASUREMENT TECHNIQUES

A. *Four basic measurement techniques*

1. *Paper and pencil instruments*
2. *Performance tests*
3. *Behavioral observations*
4. *Interviews*

B. *Considerations of adequacy*

1. *Objectivity*
2. *Reliability*
3. *Validity*
4. *Efficiency*
5. *Non-reactivity*

In conducting activity tryouts and monitoring early program implementation, it is essential to collect data which show the degree to which *both* process and student performance objectives have been achieved, as well as data on side effects. Measurement techniques of choice will differ according to the type and specificity of the outcomes to which they are applied. For example, process objectives are usually stated with such directness and precision that measurement of their attainment may involve the construction of a *simple checklist* of "Yes, it was done," "No, it wasn't" items. An example of some process objectives of this sort and a checklist to measure their attainment are included as Table 1. Similarly, a student performance objective which calls for "an increase of x points on Test Y" involves the fairly straightforward administration of Test Y. Attainment of student performance objectives which call for changes in observable student behaviors may often be measured simply by rearranging the objective into one or more test items which presumably will then measure attainment of the objective. All measurement techniques, however, should be constructed to meet basic adequacy criteria. These criteria will be discussed shortly.

Measuring Attainment of Process Objectives: a Simple Checklist

TABLE 1

Community Involvement Component Process Objectives

1. By May 15 the Community Council, composed of representatives from local organizations, school faculty, and administrative members, will be established.

2. By July 1 the target groups to participate in the program will be selected.

3. By July 1 the project schedule will be approved.

4. After federal acceptance of the project for funding, members of the Community Council will be involved in interpretation of the project to the community.

5. After federal acceptance of the project for funding, the Community Council will supply information from interested citizens to the director regarding community reaction to the project.

6. By August 1 the Community Council will approve the district's selection of project staff.

7. During the course of the project, the Community Council will be called upon by the Project Director to assist in acquiring volunteers.

Community Involvement Component Process Objectives Checklist

1. By May 15 the Community Council, composed of representatives from local organizations, school faculty, and administrative members, was established.

yes X
no ____

2. By July 1 the target groups to participate in the program were selected.

yes X
no ____

3. By July 1 the project schedule was approved.

yes X
no ____

4. After federal acceptance of the project for funding, members of the Community Council were involved in interpretation of the project to the community.

yes X
no ____

5. After federal acceptance of the project for funding, the Community Council supplied information to the director from interested citizens regarding their reaction to the project.

yes X
no ____

6. By August 1 the Community Council decided upon the district's selection of project staff.

yes ____
no X

(This is to be accomplished by September 30)

7. During the course of the project to this time, the Community Council has been called upon to assist in acquiring volunteers when the need arises.

yes X
no ____

Certain widely used measurement techniques are the basic tools of educational evaluation. Although there are roughly as many measurement techniques as there are measurement specialists (perhaps more, due to a rich legacy left by some of the legendary specialists of the past), mastery of certain basic techniques will permit the assessment of most tryout outcomes. It is estimated that 95% of the data needs of activity tryouts and early monitoring can be satisfied by the application of one of the following four techniques.

The Four Basic Techniques

Paper and Pencil Instruments

Paper and pencil instruments. Paper and pencil instruments include standardized and non-standaridized tests, questionnaires, checklists, rating scales, etc. In general, they pose some fixed question and require a *written* response from all respondents. They have the advantages of ease of administration and scoring and can easily be adapted to measurement of both cognitive and affective (attitudinal) outcomes.

Performance Tests

Performance tests. Performance tests pose some fixed question or situation and require all respondents to *do* something. The consequent response is then observed and scored according to predetermined standards. Performance tests are especially adaptable to measurement of skill and competency outcomes which require more complex responses than those required on paper and pencil instrument..

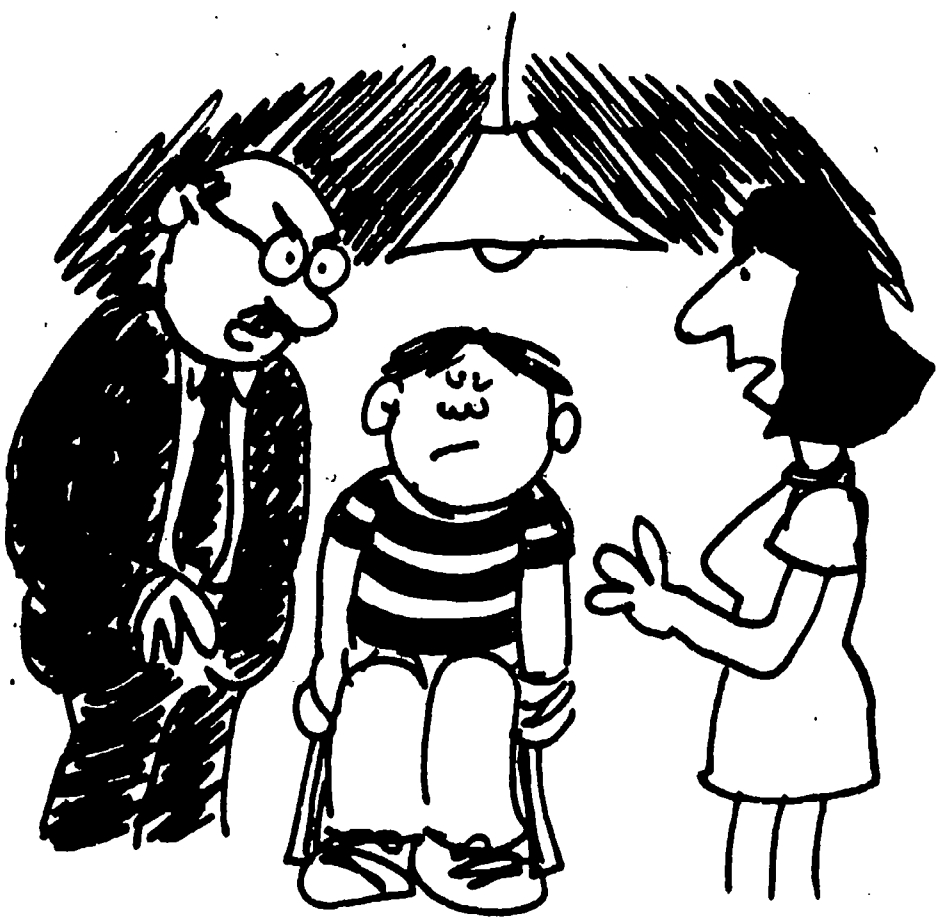
Behavioral Observations

Behavioral observations. Behavioral observations, unlike performance tests, do not involve questions or posed situations. Observations are usually performed by trained observers in a free-response situation, where the person or persons being observed are not instructed in any systematic sense. Scoring is usually performed by the observer, who categorizes behaviors into discrete predetermined units and tabulates them as a function of time. During activity

tryouts, it may be extremely useful to schedule periodic observations of the participants as they perform to get clues about portions of the activity that are frustrating, boring, too time consuming, etc.

Interviews

Interviews. Interviews require direct communication between the respondent and the data collector (interviewer). The interviewer's questions may permit a range between wide response options (unstructured or open-ended questions) or limited response options (structured questions). Interviews are generally most valuable if administered after implementation efforts to acquire information about possibly unanticipated outcomes.



Interview technique

Considerations of adequacy. Paper and pencil tests are the most widely used measurement techniques, especially standardized[3] commercially-available ones. However, small-scale activity tryouts are more likely to require unstandardized, user-constructed paper and pencil tests, performance tests, observations, and interviews. Because only standardized tests usually contain built-in attention to problems of measurement adequacy, this makes it even more crucial for users at this stage to be aware of and to conscientiously apply certain well-accepted considerations to avoid misinterpretation of measured results. The application of such *considerations of adequacy* to measurement techniques is nothing more than an attempt to keep you from fooling yourself and others. You should apply these considerations whenever you select or construct outcome measures.

Objectivity

Objectivity. Will the technique yield the same score regardless of who is applying it?

It should be recognized that all measurement techniques have some *deficit* in *objectivity*. In other words, there is a continuum running from relatively objective methods to relatively subjective methods. This same principle also applies to reliability, validity, efficiency, and reactivity, which will be discussed subsequently. Objective methods are those which yield similar scores no matter who is doing the scoring. Subjective methods are those which can yield widely disparate scores when applied by different persons.

In order to improve objectivity, it is necessary to

[3]Standardized here refers to the fact that the tests have been administered to a large (usually representative) reference (or norm) group whose score distribution permits comparative interpretation of new scores; e.g. determining how Johnnie's test score compares to the national *norm*.

establish scoring *rules* which enable unequivocal assignment of scores to each obtained response. For example, an "objective" paper and pencil test may require a very discrete written response, such as "True/False" or "a, b, c, or d" which, with the aid of a test key, can then be scored "right" or "wrong" by virtually everybody who might be scoring that response. A less objective essay test item requires a less discrete written response, and consequently there may be considerable disagreement among scorers as to the numerical value of that response. Objectivity can be improved for this type of item by establishing a type of key which states the general scoring rules, then lists examples of typical responses and the proper scores for each.

Develop Scoring Rules of Keys

Another factor in improving objectivity is to remove identifying information from each response, so that a scorer will not know who wrote it. This helps control such threats to objectivity as the "halo" effect. In the "halo" effect, the responses of a person known to the scorer tend to be scaled according to the scorer's general opinion of that person rather than on the merit of the response. Information on identity can be removed by using code numbers rather than respondent name on all scorable documents.

Remove Identifying Information

Halo Effect

All things being considered, paper and pencil tests are usually more objective than performance tests, observations, and interviews. During activity tryouts, however, you are generally often more interested in obtaining intuitive or idiosyncratic judgments which are not amenable to highly objective scoring. Thus, even relatively subjective techniques may have an important role at this stage if proper care is taken in interpreting their outcomes.

Reliability

Reliability. Does the technique produce data which are free from random error and thus yield a relatively "constant" score?

Reliable measurement techniques are those which yield *constant* scores which are relatively free from *chance* variation over time. For example, if you measure an object twice with a wooden ruler, the results are likely to be very consistent, or reliable. On the other hand, if you used a rubber band to perform such multiple measurements, you may observe considerable inconsistency; the rubber band is a relatively unreliable instrument for measuring length. Lack of objectivity can lead to low reliability. Trick questions or inaccurate recording devices (e.g. unwound clocks, inaccurate test keys) can also lead to low reliability.

Application Considerations

A more important consideration in improving reliability is the way measurement techniques are *applied*. An application which poses the same question or situation under identical conditions to all persons being measured is generally more reliable. If possible, then, instructions and conditions should be the same for all examinees. Questions, or test items, should be phrased so that they discourage wild guessing. All examinees should be familiar with the type of response being requested. Always give a few practice items or a warmup period so that performance will not be unduly influenced by unfamiliarity with the

type of item being used. Finally, combining several items usually produces a more reliable score than a single item. For example, if you are attempting to measure mastery of a student performance objective, a more reliable estimate of mastery can be obtained by requiring satisfactory performance on four out of five items rather than on one item.

Validity

Validity. Does the technique actually measure what it is supposed to be measuring, intuitively and by empirical demonstration, and thus yield a relatively "true" score?

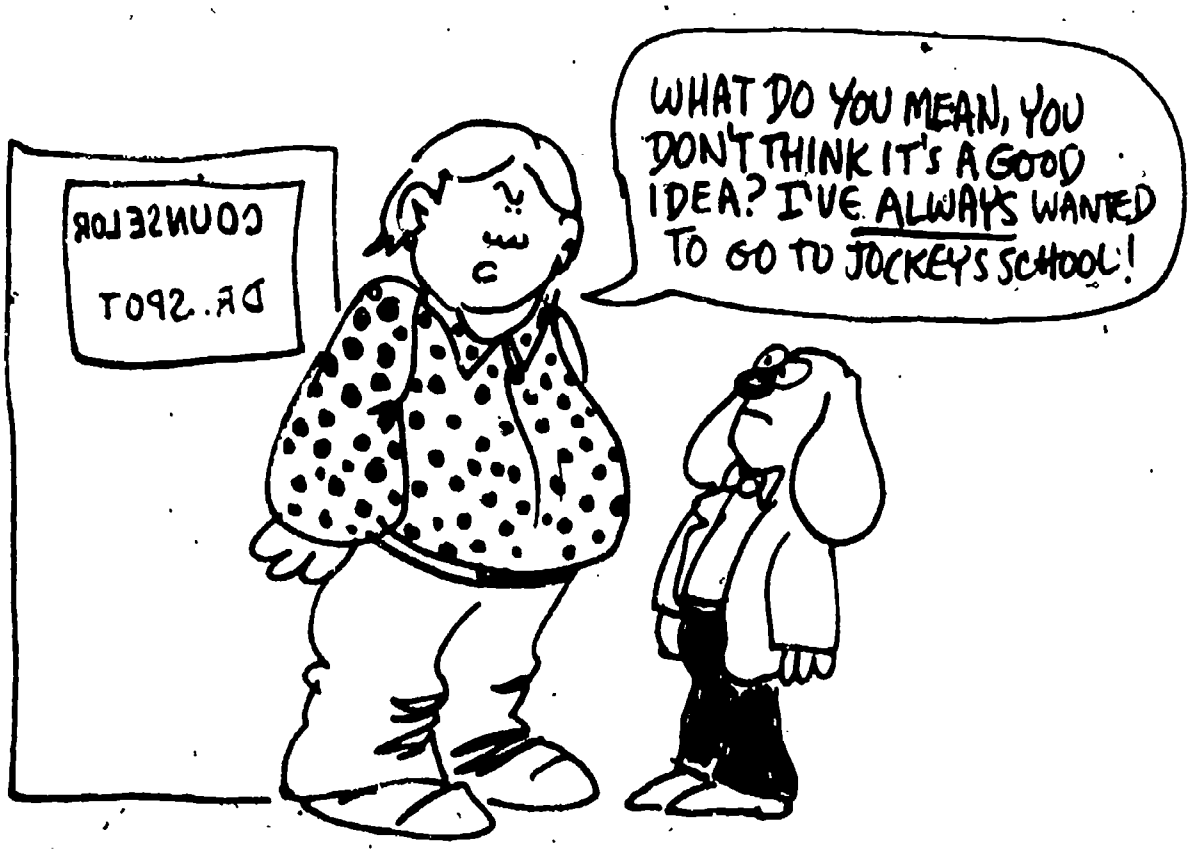
For present purposes, we are interested mainly in that aspect of validity which refers to the congruence between an underlying characteristic or objective and an obtained score. For example, if a score indicates a student has mastered an objective, has he in fact mastered it? In this connection, we may be interested in knowing two basic facts. First, is the obtained score a measure of the *desired* characteristic and not some other extraneous or peripherally-related characteristic? If we are interested in measuring, say, "knowledge of occupational opportunities," does our technique measure this and not reading ability, test-taking ability, etc? Second, is the obtained score adequately representative of the *entire* characteristic and not some very limited aspect of the characteristic? In the example posed above, does our technique measure knowledge of a broad range of occupational opportunities or just a few limited ones? It is desirable to try to insure that measurement techniques are both *closely related to* and *broadly representative of* the characteristic being measured.

Construct Explicit Rationales

The first step in doing this is to construct an *explicit rationale* for each measurement technique being utilized, to show in writing how the technique relates to the objective or characteristic being measured. For the above example, suppose one student performance objective for a career guidance program is for "participating students

to be able to describe vocational and educational opportunities which are available to them." There are many ways to measure attainment of such an objective; in fact, paper and pencil tests, performance tests, behavioral observations and/or interviews might all be adapted to the task. Assume we decide to adopt a paper and pencil item such as "List at least two educational opportunities open to you." The explicit rationale for this item might look something like the following:

> Objective is primarily *cognitive*, of basic *knowledge* type. Demonstration of attainment requires a *constructed* response (as opposed to a *selected* response or *multiple-choice*), with no cues or prompts. Response must show some awareness of the *practical* range of opportunities open to each individual respondent, e.g., a youth with a C- grade point average and no course work in math should not list a college pre-med course as an educational opportunity. This requires scoring by persons knowledgeable about each respondent. Additional scoring criteria must be designed to eliminate overlap (e.g., require "discrete" statements of opportunity) and require adequate specificity (e.g. "attend Foothill College in the general studies program," not "go to college").



The written rationale tends to (1) force attention to important indicators of the objective (away from more trivial indicators), (2) highlight practical administration and scoring details which might otherwise be overlooked, and (3) ensure that the required response is within the capability of the examinees.

Measure All Known Aspects of Objective or Characteristic

The second step in helping to ensure validity is to provide sufficient measures of *each* important objective or characteristic. In the above example, the designated paper and pencil item, while probably an acceptable one, is insufficient in itself to adequately measure attainment of the stated objective. At the very least, another item on *vocational* opportunities would be required. Scores on the multiple items would then be summed to measure attainment of the objective.

In general, measurement techniques which are relatively objective and reliable are more valid.

Efficiency

Efficiency. Is the technique relatively cheap and easy to administer (*at least* within the capabilities of the person who is performing the measurement) and score?

Relatively efficient techniques are those which yield reliable and valid scores at a low cost in terms of money and examiner and examinee time. In general, this means that instruments which can be administered to *groups* rather than individuals, *once* rather than on multiple occasions, and under *normal* rather than contrived circumstances, are *more efficient*. It also means that exercises which can be *scored and processed quickly* (e.g., with a test key, a behavior frequency tabulator, etc.) and easily (e.g., by a clerk, a test-scoring machine, etc.) are more efficient than those which require more time (e.g., rating an essay, analyzing the content of a tape recording) and expertise (e.g., by a committee of experts). Efficiency is especially important in measuring the effects of activity tryouts

because the results are usually needed for immediate application in modifying the activity before actual implementation.

Non-reactivity. Does the technique unduly influence the subsequent behavior of the person to whom it is being applied?

Non-reactivity

The classic example of a highly reactive technique is uprooting a seedling daily to measure its growth. Reactivity is to be avoided *unless* the reactive effect of the measure is designed to be part of the activity being scrutinized, e.g., when a pre-test is designed to sensitize students to the material which they are supposed to learn.

Relatively non-reactive measures include physical *traces* of past events, routinely collected *records* or statistics, and unobtrusive *observations*. An example of using physical traces as a measure is the case of a well-known research firm which made daily records of the number of cigarette butts found in ashtrays to estimate the effects of a lung cancer prevention campaign. An example of using routine records as a measure is the use of school attendance figures as a measure of the effectiveness of an anti-truancy program. Unstructured observations of such things as time to complete an activity and frequency of certain behaviors (like long disgusted sighs accompanied by throwing pencil down on desk) can often tell more than any other measure about the effectiveness of an activity. Non-reactivity is promoted when the observer is placed so as not to intrude into the situation s/he is observing.

The number of possibly useful non-reactive or unobtrusive measures is limited only by your imagination. Since the validity of such measures often tends to be low, however, multiple measures may become necessary, and special care must be taken in the construction of sound explicit rationales.

ACTIVITY 3 - "REALLY UNDERSTANDING" MEASUREMENT AND MEASUREMENT TECHNIQUES

Please read the following passage from Fred Kerlinger's 1964 book on the foundations of behavioral research

> "In its broadest sense, measurement is the assignment of numerals to objects or events according to rules." [1] This definition of measurement succinctly and accurately expresses the basic nature of measurement. To understand the definition, however, requires the definition and explanation of each important term —a task to which much of this chapter will be devoted.
>
> Suppose that we ask a male judge to stand seven feet away from an attractive young woman. The judge is asked to look at the young woman and then to estimate the degree to which she possesses five attributes: niceness, strength of character, personality, musical ability, and intelligence. The estimate is to be given numerically. In the number system a scale of numbers from 1 through 5 is used, 1 indicating a very small amount of the characteristic in question and 5 indicating a great deal of the characteristic. In other words, the judge, just by looking at the young woman, is to assess how "nice" she is, how "strong" her character is, and so on, using the numbers 1, 2, 3, 4, and 5 to indicate the amount of each characteristic she possesses.
>
> After the judge is finished, another male judge is asked to repeat the process with the same young woman. The numbers of the second judge are checked against those of the first judge. Then both judges similarly judge a number of other young women.
>
> This example may seem to be a little ridiculous. Most of us, however, go through very much the same procedure all our lives. We often judge how "nice," how "strong," how "intelligent" people are simply by looking at them and talking to them. It only seems silly when it is given as a serious example of measurement. Silly or serious, it *is* an example of measurement, since it satisfies the definition. The judges assigned numerals to objects according to rules. The objects, the numerals, and the rules for the assignment of the numerals to the objects were all specified. The numerals were 1, 2, 3, 4, and 5; the objects were the young women; the rules for the assignment of the numerals to the objects were contained in the instructions to the judges. Then the end-product of their work, the numerals, might be used to compute measures of relation, analyses of variance, and the like.
>
> The definition of measurement includes no statement about the quality of the measurement procedure. It simply says that, somehow, numerals are assigned to objects or to events. The "somehow," naturally, is important—but not to the definition. Measurement is a game we play with objects and numerals. Games have rules. It is of course important for other reasons that the rules be "good" rules, but whether the rules are "good" or "bad," the procedure is still measurement.

Why this emphasis on the definition of measurement and on its "rule" quality? There are three reasons. First, measurement, especially psychological and educational measurement, is badly misunderstood. It is not hard to understand certain measurements used in the natural sciences—length, weight, and volume, for example. Even measures more removed from common sense can be understood without wrenching elementary intuitive notions too much. But to understand and accept the fact that the measurement of such characteristics of individuals and groups as intelligence, aggressiveness, cohesiveness, and anxiety involves *basically and essentially* the same thinking and general procedure is much harder to do. Indeed, many say that it cannot be done. Knowing and understanding that measurement is the assignment of numerals to objects or events by rule, then, helps to erase erroneous and misleading conceptions of psychological and educational measurement.

Second, the definition tells us that, if rules can be set up on some rational or empirical basis, measurement of anything is *theoretically* possible. This greatly widens the scientist's measurement horizons. He will not, in short, reject the possibility of measuring some property because the property is, say, a complex and elusive one. He understands that measurement is a game that he may or may not be able to play with this or that property at this time. But he never rejects the possibility of playing the game, though he may realistically understand its difficulties.

Third, the definition alerts us to the essential neutral core of measurement and measurement procedures and to the necessity for setting up "good" rules, rules whose virtue can be empirically tested. No measurement procedure is any better than its rules. The rules given in the example above were poor. The procedure was a measurement procedure; the definition was satisfied. But it was a poor procedure for reasons that should become apparent later.[4]

This passage was quoted in its entirety because it represents an extremely nice (if slighty chauvinist--remember it was done in 1964) job of presenting the definition of measurement and some of the problems associated with it. Anything is theoretically measurable, but not necessarily with complete accuracy. Quality of measurement is *relative*, and measurement techniques yield relatively better or poorer scores according to the way they are applied and the use to which the scores are put.

In the text, we talked about four basic *types* of measurement techniques. Each type has many variations and each variation has certain rules of construction which, if followed, will make the scores *better* measures.

---

[4] Fred N. Kerlinger, *FOUNDATIONS OF BEHAVIORAL RESEARCH* (New York: Holt, Rinehart, and Winston, Inc., 1964), pp.411-412.

In order to practice your skills in constructing evaluation measures, please read the following situation.

You are the guidance counselor at an elementary school. You have developed an activity designed to help students who come to the school nurse's office with minor injuries caused by in-school accidents. The goal of the activity is to help students to recognize things they can do to prevent such accidents. This is just one subcomponent of your overall guidance program goal, which is to help students to overtly identify sources of pain and frustration and formulate and practice realistic steps to eliminate or reduce them. The desired outcome of the new activity is that each student who reports to the school nurse's office with a minor injury will be able to describe, to the satisfaction of the nurse: (1) the immediate cause of the injury, (2) what s/he did prior to the injury, (3) how this act led directly to the injury, (4) what s/he might have done differently to avoid the injury, and (5) what s/he can do in the future to avoid similar injuries. The activity will involve giving all nurses a set of structured questions to ask each student who comes in with a minor injury. The questions will elicit responses consistent with each aspect of the objective stated above and will be repeated in slightly different forms until each student has voiced a suitable response.

17

Please write one process objective of the activity. Write an item to measure its attainment?

Please write one student performance objective of the activity. Write an item to measure its attainment using each of the following techniques, showing particular concern for objectivity, reliability, and validity.

1. A paper and pencil instrument

2. A performance test

3. An observation of student behavior

4. An interview

Are there likely to be some unanticipated outcomes? What might be one? Indicate how you would measure its attainment.

RESULTS OF PRELIMINARY ACTIVITY TRYOUTS AND EARLY IMPLEMENTATION

If desired outcomes are achieved with a minimum of undesired side effects, improvements in efficiency and effectiveness can be expected during later implementation efforts.

However, if desired outcomes are not achieved or are accompanied by undesired side effects, further analysis is indicated. Closer analysis of the *process* and student product *outcome* data is required to determine if (1) outcome deficits were probably caused by improper *implementation* of an activity, (2) outcome deficits were probably the result of an underlying fault in the basic design or *rationale* of an activity, or (3) outcome deficits were probably the result of a failure in the interactions between two or more activities.

Possible Defects

Improper implementation may often be spotted and corrected rather easily, especially when the process objective checklist shows crucial "No, it wasn't done" responses. Decisions to proceed with implementation in the face of negative results may even be made immediately if there is considerable confidence in the activity's logic and in the corrective action that has been taken.

If everything was apparently done as planned, however, the problem is more serious. It may be necessary to select alternate strategies or activities to achieve the desired outcomes. Further tryouts may be necessary prior to large scale implementation.

"It's worth it."

It is almost always preferable to discover implementation problems and/or theoretical deficits in an activity's logic *prior* to large scale implementation. All too frequently, well-conceived and costly guidance programs initially fail to achieve their goals because of the failure of one important component. By then student, parent, and community expectations may have been seriously deflated and even effective remedial action may arrive too late to save the program.

SUMMARY

These readings have briefly discussed the valuable practice of conducting preliminary tryouts and monitoring early implementation efforts. A quantitative rating scheme was suggested to help you determine which activities are most in need of pilot testing prior to implementation. The ratings consider aspects of activity replicability, revisability, importance, and uncertainty of effect. Prior to carrying out empirical pilot tests, the internal logic of the activity should be verified. Actual pilot testing involves implementing the planned activity, under well-controlled circumstances, using representative members of the target population, collecting information on degree of process and student performance objective attainment, side effects, time for completion, and costs. Monitoring involves collecting similar data during early implementation efforts for combined activities. Outcome measurement is facilitated if certain standard techniques are utilized with consideration for well-accepted criteria for adequacy. Inferences about the effects of implementation are facilitated by the use of a simple evaluation design. Deficits in desired effects, if discovered at this stage, may be corrected in time to optimize the career guidance program of which the activity is a part.

DISCUSSION QUESTIONS

The following questions are designed to let you check on the knowledge you have gained and discuss with the Coordinator and those in your group any remaining questions or problems. Discussion may range to other questions and issues related to the module also, if you desire.

1. What is formative evaluation?
2. What are two types of career guidance activities for which tryouts or pilot tests are not useful?
3. What are two types of career guidance activities for which tryouts are *most* useful?
4. What steps should be taken in reviewing the internal logic of a planned activity?
5. What is the general strategy of pilot testing?
6. Why are sample groups for activity tryouts rarely chosen by random selection?
7. What is "evaluation design"?
8. What are four considerations of measurement adequacy?
9. What is a way of improving a measure on each of the considerations of adequacy you discussed in question 8?
10. What are the major expense categories to consider?

POSTASSESSMENT

1. In a sentence describe the primary purpose of carrying out activity tryouts.

2. Please list two situations in which pilot testing of activities is useful and two in which it is a waste of time.

   Useful: 1.

   2.

   Waste of Time: 1.

   2.

3. One of the process objectives for a career guidance program is the following.

   > By June 30, all the books and materials required for the special study carrel will be either ordered or in hand.

   Please construct a checklist item for measuring attainment of this process objective.

4. The box below contains a brief description of the trial implementation of a career guidance activity. Please develop a paper and pencil test item, a performance test item, a behavioral observation, and an interview schedule to measure the intended outcomes of this activity. These can be brief and to the point.

   > In order to estimate the probable effectiveness of the proposed handbook about vocational training programs, Ms. Longpause, the head of the Fliver High School Guidance Department, decided upon a small scale tryout of a prototype handbook. She personally constructed a two-page description of the goals, current enrollment, minimum entry requirements, length, and 1971 graduate placement data on the local Abacus School of Electronic Computer Programming. She asked the first ten students who came into her office for appointments the next Monday morning to individually read the description. She then asked them to take the description home and discuss it with their parents.

5. Check your responses to question 4. How could the paper and pencil item you wrote be made relatively *objective*?

6. Would the four items you listed in question 4 above constitute a relatively *valid* measure of the attainment of the objectives of t.e vocational training handbook? Why or why not? How could they be improved?

7. Later that week, Ms. Longpause decided to check on unanticipated outcomes by calling the parents of each of the ten students to see if any had discussed the Abacus program at home. How could she secure more *reliable* information from these calls?

APPLICATION

You are now ready to map out a plan for applying the skills you developed in this module to your own setting. If you are not working in a particular setting and would like a hypothetical setting with which to think through this Application, turn to the "Optional Group Simulation Description" located in the Appendix. Thinking of the sections of the module and the questions you have discussed, consider the tasks that must be accomplished to try out activities and monitor early implementation efforts for your program, the person who should be primarily responsible for each task, and the date by which the task should be done. Since these tasks must be performed after the process objectives have been established, the dates you choose must coordinate with those set for earlier program planning tasks. If possible, this exercise should be done under the direction of someone from your district who would be good at taking charge of this effort and who would like to do it. Use the chart on the following pages for listing your tasks, the people responsible, and the completion dates. The major headings are already noted on the chart. Use additional paper as needed.

APPLICATION

TASKS	INDIVIDUAL(S) RESPONSIBLE	COMPLETION DATE
I. Determine when to pilot test.		
II. Devise a general strategy of pilot testing.		

TASKS	INDIVIDUAL(S) RESPONSIBLE	COMPLETION DATE
III. Estimate costs.		
IV. Determine the best measurement technique.		

APPENDIX

OPTIONAL GROUP SIMULATION DESCRIPTION

The description below is intended as a frame of reference for those of you who are not working in a particular setting. You may wish to describe one of the other schools that might be in the "Lillington School District," to change parts of this description, or to use it for ideas in creating your own setting. Feel free to modify it to meet your particular needs.

The Lillington School District is a suburban district made up of seven elementary schools, four junior high schools, and two high schools, with a total school population of 10,000 students. The district is rapidly changing from one that is largely rural to one that is urbanized. It has several active industries, including a furniture factory and an automobile assembly plant. The minority population is growing, and the unemployment rate is higher than that in the rest of the state.

Lillington's Chester Arthur High School has an enrollment of about 1,600 students in grades 9-12. Approximately 75% of the students are from lower and lower-middle class families. The racial/ethnic composition is approximately 82% white, 11% black, 5% Spanish-surname, and 2% Oriental and Indian. The school has experienced some group conflicts, but none serious to date.

No follow-up studies have been conducted in recent years, but the school counselors estimate that about 40% of their graduates continue their education, mainly at the nearby community college. The dropout rate there and in Arthur High is unknown. The counselors guess that at least half of the girls marry within a year after graduating from Arthur High; most of them do not seek jobs until after they have had children. Most of the remainder of the female graduates and a large majority of the males seek jobs locally, principally in the furniture and auto assembly plants. No information is available as to their success or failure in obtaining jobs.

Chester Arthur High School is headed by a fairly pragmatic principal and a progressive, forward-looking vice-principal. The teaching staff is a mixture of older, conservative individuals who want to avoid most "new fangled ideas" and young, liberal instructors who are eager for change. Each group has its well-liked leaders.

The guidance staff consists of the director of guidance, who is a frustrated Freudian psychoanalyst, and two counselors: a middle-aged ex-space engineer who loves paper work and a young vegetarian who likes to "rap with kids." Two part-time aides assist with filing and other clerical tasks. The department has never formulated its goals and objectives, and its principal guideline is "How have we handled that in the past?" The department seems to be drifting, each member "doing his own thing" without cooperation or coordination.

The school's efforts in career education are haphazard. Some teachers, it is thought, include in their classes some career information about the subject area. Counselors supply specific information to students who request it. The school library includes a file of occupational information, but it is rarely used and poorly maintained.

Chester Arthur High, as well as other Lillington District schools, has experienced a number of problems recently. There are persistent rumors of widespread drug sales and usage on campus. Racial tensions are increasing. Students complain a great deal about the curriculum and the school in general. School spirit is at an all-time low.

Parents and community members are becoming increasingly dissatisfied with the school system. They feel that taxes for education are far too high and that students are not adequately prepared for the world they face upon leaving school. A number of parents believe that faculty members are too liberal and that they should devote more time and effort to "the basics." Other parents insist that the school is old fashioned and that much of the subject matter is irrelevant and unimportant. Parents from both of these groups say that their children do not know what they want to do and are uninformed about career opportunities and job requirements.

Arthur High graduates, complain local employers, make poor employees. Many employers seem to agree with a statement made by a foreman at the furniture plant: "All these kids care about is the paycheck, but they don't know how to earn it."

Due to financial problems, there has been talk recently of having to reduce the school staff. One of the leading candidates for reduction is the guidance staff. While the staff members seem to be generally liked by both students and teachers, the school administrators are having trouble justifying the expenditure of money for so tenuous a purpose. They are seeking documentation and other evidence to justify retaining the guidance program and staff.

REFERENCES

There are several references which are "must" reading for those who would like to go deeper into the area of formative evaluation in general and activity tryout in particular. These are as follows.

Baker, Eva L. "Formative Evaluation of Instruction." In *Evaluation in Education: Current Applications*, edited by W. James Popham, pp. 531-585. Berkeley: McCutchan Publishing Corp., 1974.

This is an excellent chapter, most comprehensive yet communicative. Chapter subheadings include What is Formative Evaluation For?, Formative Evaluation of Instructional Prototypes, External Data Gathering in Prototype Tryouts, Prototype Data Sources: Summary, Issues in Prototype Testing, Operational Testing, Optional Exercises, and References.

Sorenson, Garth. "Evaluation for the Improvement of Instructional Programs: Some Practical Steps." *Evaluation Comment* 2, 4, (January 1971), pp. 13-17. A publication of the Center for the Study of Evaluation, UCLA.

As the title implies, this article presents practical steps in the formative evaluation of educational programs. It includes eight "principles," examples, references, and a "Formative Evaluation Checklist" for use by developers of programs which have definable goals.

Briggs, Leslie J. *Handbook of Procedures for the Design of Instruction.* Pittsburgh: American Institutes for Research, 1970, pp. 173-177.

The referenced pages constitute Chapter 8 of this document, entitled "Formative Design, Formative Evaluation, and Summative Evaluation." Briggs presents an outstanding case study example of the use of formative design and tryout procedures to bring about dramatic increases in the performance of learners on a multimedia first aid course. He also presents eight valuable suggestions for conducting formative evaluations and six self-test items.

Van Dalen, D.B. and Meyer, William J. *Understanding Educational Research*. Second Edition. New York: McGraw-Hill, 1966, pp. 301-325.

These 25 pages are the best we have found for presenting simple straightforward instruction on the construction of paper and pencil instruments, including questionnaires, tests, inventories, rating scales, checklists, sociometrics, etc.; performance tests; behavioral observations, including check lists and schedules, time sampling, etc.; and interviews, including individual and group, structered and unstructured.

Much more detailed coverage of the same material may be found in chapters 26-32 in

Kerlinger, Fred N. *Foundations of Behavioral Research*. New York: Holt, Rinehart, and Winston, Inc., 1964, pp. 467-602.

Additional References

Flanagan, John C. *Measuring Human Performance*. Palo Alto: American Institutes for Research, 1974.

This text is especially valuable if you are interested in attitude or value measurement. It suggests that too much dependence in this area is placed on self-report ratings and scales and not enough on the observation of behavior--a valid indicator (or indirect measure) of "attitude" or "value." See especially pp. 222-227 on uses of human performance measures in guidance.

Greenberg, B.G. "Evaluation of Social Programs." *Review of the International Statistical Institute* 36, 3, (1968), pp. 260-277.

Isaac, Stephen, and Michael, William B. *Handbook in Research and Evaluation*. San Diego: Robert R. Knapp, 1971.

A very confusing compendium of various and sundry techniques and principles, unless you know what you're looking for. Pages 82-91 contain concise explanations of validity and reliability, along with techniques for generating *quantitative* estimates of these concepts; those could be very useful to you if anybody ever says "What was the *reliability* of your measure?!" with a sly smirk. Pages 62-63 contain a nice discussion of measurement reactivity. Pages 92-93 contain useful information on designing and carrying out a mailed questionnaire survey.

Popham, W. James. *An Evaluation Guidebook: A set of Practical Guidelines for the Educational Evaluator*. Los Angeles: The Instructional Objectives Exchange, 1972.

Scriven, Michael. "The Methodology of Evaluation." In *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally, 1967.

This document is the first in a series of American Educational Research Association Monographs on curriculum evaluation.

Webb Eugene J., *et al. Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally, 1966.

The bible of non-reactive measurement techniques. Also a very witty, knowledgeable, and well-written book.

Weiss, Carol H. "Utilization of Evaluation: Toward Comparative Study." In *Readings in Evaluation Research*, edited by Francis G. Caro. New York: Russell Sage Foundation, 1971, pp. 136-142.

COORDINATOR'S GUIDE

MODULE 10

TRYING OUT ACTIVITIES AND MONITORING EARLY IMPLEMENTATION EFFORTS

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

August 1975

Reprinted with minor revisions, July 1978

TABLE OF CONTENTS

	Page
Coordinator's Role and Functions	1
Introductory Activity	3
Outline of Introductory Remarks	5
Activity Feedback	6
Assessment Criteria	9
Introduction to Application	12
Sample Evaluation Instruments	13

Your role as coordinator is crucial. It may be thought of in four categories.

Set the Tone

Set the right mood. Don't make things deadly and boring. Inject humor into the activities and discussions, let people joke around some and have fun. On the other hand, make it clear that there is a very serious purpose behind it all. People should be relaxed, but alert, interested, and motivated.

Set the Pace

Maintain the right pace. If things bog down, inject some humor, ask some provocative questions, get a lively discussion going. Some sections can be summarized orally to speed things, and this can be planned ahead. If things are going too fast and people are getting lost, slow it down, let them ask questions, spend time orally covering the points. Keep the flow smooth at junctures in the module--winding up one activity with a satisfying resolution and easing participants into the next. Take breaks as you sense they are needed. Be flexible in structuring activities, adapting to individuals and situations as needed. Regard times listed in the "Module Outline" as flexible.

Facilitate

Encourage discussion and interaction from the participants. Bring out the shy people; don't let the aggressive ones dominate. Seek out questions and uneasinesses, get them into the open, talk them over, especially at the beginning. Watch facial expressions and body language. Be a trouble shooter. Spot problems and work them out. In short, act as a guide through the module, but try not to get in the way.

Evaluate

Make sure participants are headed in the right direction; nudge them that way when they're not. Judge whether they perform adequately in the postassessment items, the activities which are part of the assessment (see the "Assessment Criteria"), and the Application. Keep a record of how each participant does. In general, maintain the quality level of the workshop.

Specific Functions of Coordinator

Prior to workshop:

Study the module thoroughly ahead of time. Be familiar with all participant materials and this Coordinator's Guide.

At the workship:

1. Introduce yourself to participants, and them to each other. Briefly explain your background and the role you will play in the module.
2. Establish time limits (lunch, when day ends) and schedule for the day, and do your best to stick to them.
3. Conduct Introductory Activity if you plan to use it (see "Introductory Activity" herein).
4. Introduce the basic purposes and structure of the module (see "Outline of Introductory Remarks"). Answer any questions.
5. Start participants on the text. Lead discussion and practice activities as you go through this module. Provide feedback on the practice activities (see"Activities Feedback" and "Assessment Criteria"). Collect evaluation data on the participants.
6. Conduct the Postassessment. Collect the results. Evaluate each of the participants on her/his performance (see "Assessment Criteria"). Keep a written account of this.
7. Start participants on the Application (see "Introduction to Application"). Evaluate the plan produced (see "Assessment Criteria").
8. Conduct a Wrap-up session. Your tasks here are to
   a. Summarize what has gone on and been accomplished.
   b. Resolve any unanswered questions.
   c. Point our sources for additional study. Go through the Reference section briefly; add any sources you know of.
   d. Mention any technical assistance available--experts related to module topics to whom participants might be able to turn.
9. Throughout, observe how things go; collect suggestions for ways to improve the module. Keep a written account of these.
10. Submit the results of 5, 6, 7, and 9 to the overall workshop director.

OUTLINE OF INTRODUCTORY REMARKS

Learners generally do better if they are presented an overview of what it is they are to learn. This is your main job at the beginning of the module. Having reviewed the materials, briefly summarize them for the participants, preparing them for what is to come. Go over the Module Goal and Outcomes to be sure that the knowledge and skills to be gained are clear. Go over the Module Outline so that participants will understand how their time is to be structured. Review the Model and how Module 10 fits in. Encourage and answer any questions you can.

No outside materials are required for this module. However, Sorenson's article would be useful to have available for participants. Also, if at all possible, have copies of the Van Dalen text available. (These are described in the References.) If by any chance you have access to the Baker and Briggs texts also, so much the better.

ACTIVITY AND DISCUSSION FEEDBACK

One of your most important functions as Coordinator is to provide feedback to participants as they work through the readings, discussions, and activities. Try to make sure they are understanding the central points and are able to do the things requested in the activities. To help you in this role, some feedback suggestions are provided below.

Discussion on "Subset Optimization

Try to help participants see that pilot tests normally involve small "diagnostic" tryouts of individual program components.

Discussion on a "General Strategy" of Pilot Testing

Try to help participants derive a strategy such as the following:

> The general strategy of pilot testing is to implement the planned activity under well-controlled circumstances, using members ofthe target population, collecting (a) objective information on degree of attainment of process and student performance objectives, and (b) any other objective or subjective information which is likely to be useful in indicating changes that can potentially be made in the activity. Care should be taken to avoid over-collection of data so that immediate diagnostic feedback can be obtained.

Activity 1 - Analyzing a Program's Internal Logic

This task requires a good deal of common-sense analysis, such as verifying that process C will in all likelihood lead to attainment of performance objective B, which can reasonably be expected to contribute to attainment of goal A. Specifically, the written analysis should contain evidence of consideration of the following points:

a. Are the activities described specific enough to be observed? Can their attainment be *documented*?

b. Is there reason to believe that most members of the target audience can *already perform* the student performance objectives, making the planned activities unnecessary?

c. Is there any indicated *time schedule* or sequence for multiple activities?

d. Are the planned activities and sequence practical with regard to the constraints likely to exist in this situation? Are they appropriate to the level of the target audience?

e. Are there likely to be *side effects* which the program developers have probably not anticipated?

f. Do the planned activities *relate* to the intended student performance objectives? Is there a reasonable chance that the activities, if carried out as indicated, will produce the desired outcomes? Do the proposed student performance objectives relate to the program's stated goal or goals?

Activity 2 - Estimating Costs

The total expenses for the first year would come out something like this. Figures may vary within a reasonable range.

COST CATEGORY	NO. OF UNITS	RATE/UNIT	ESTIMATED COST
Personnel			
Paraprofessional	40 weeks	$150/week	$6,000
Principal	10 days	~ $100/day	$1,000
Consultant	10 days	~ $125/day	$1,250
Teachers	400 hours	~ $ 7/hour	$2,800
Counselors	200 hours	~ $ 7/hour	$1,400
Secretary	1 month	$750/month	$ 750
		TOTAL	$13,200
Equipment/Supplies			
Film Projector	1	$600	$ 600
Filmstrip Projector	1	$200	$ 200
Screen	1	$ 50	$ 50
Tape Decks	2	$100	$ 200
Games and Materials	several	$200	$ 200
Books			$1,000
General Supplies			$ 200
		TOTAL	$2,450
		FIRST YEAR	$ 245
Other Program Expenses			
Mailing Letters	2,000	$.10	$ 200
Reproducing:			
Letters	2,000 pp.	$.03	$ 60
Feedback form	2,000 pp.	$.03	$ 60
Report	2,000 pp.	$.03	$ 60
		TOTAL	$ 380

COST CATEGORY	NO. OF UNITS	RATE/UNIT		ESTIMATED COST
Building Costs				
	1 classroom	$1,000		$ 1,000
			TOTAL	$ 1,000
Fixed Administrative Charges				
	$14,825 budget (total of above four categories)	5%		$ 750
			TOTAL	$ 750
			SUBTOTAL =	$15,575

With this information, the decision becomes much easier. It's a matter of whether the benefit derived was worth $15,575. Since there were approximately 1,000 users the first year, another way of stating the question is whether it was worth $15 per student to have the career center.

Activity 3

The feedback for this activity is in the form of some *possible* answers to the questions asked. These are not the only answers and are intended to be only suggestive of the responses that participants might generate. Please discuss all responses in your group and bring out differing perceptions.

Process objectives would state something to the effect that the school nurse is to ask the following questions of all students who appear requesting treatment for minor school-related injuries.

What happened to cause this? What were you doing before that? Why do you think what you did caused your accident? Could you have avoided the accident? How?

Measurement of process objective attainment would involve observa- of the nurse and marking a "Yes, it was done in this case," "No, it wasn't done in this case" checklist.

Student performance objectives would state something to the effect that students would be able to respond *vocally* to the above questions in an appropriate manner. Criteria of response adequacy might require certain standards of probable accuracy (to avoid lying), common sense (to avoid unlikely causes), and realism (to avoid unsound avoidance strategies).

Although the objective of the activity clearly calls for a "performance" test as the ultimate criterion, many other measurement techniques could be used to estimate side effects. The range of possible alternatives is almost unlimited, so it is very difficult to suggest an answer a priori to this question. For each of the four techniques, however, some evidence should be provided to show consideration of the following techniques for improving objectivity, reliability, and validity.

Objectivity. Develop scoring keys or rules. Remove identifying information from responses.

Reliability. Increase objectivity. Equalize testing conditions for all examinees. Discourage random responding, e.g., wild guessing. Use familiar and uncomplicated response forms. If possible, combine measures to produce a composite score.

Validity. Improve reliability. Construct explicit rationales for each technique. Measure all aspects of an objective or characteristic.

Some of the possible unanticipated outcomes include the following: (1) students are injured less frequently at school, (2) students are injured with the same frequency but come to the nurse's office with less frequency to avoid a "lecture," (3) administering first aid to students becomes so time-consuming there is always a queue outside the nurse's office, (4) teachers send students to the nurse's office for more minor injuries than before because students now seem to come back so "pensive," (5) the nurse becomes a non-directive counselor and aces you out of your job.

Measurement of such outcomes involves tabulating records, interviewing students and teachers, observing behavior, and checking your pay envelope every month for a pink slip.

End of Module Discussion

Again, allow for variations and individual emphases. The point here is to review some of the major topics of the module and be sure everyone has understood and internalized the information. Don't feel restricted to this set of questions if others can be added that seem appropriate.

1. Formative evaluation is a process of collecting and using information to improve the functioning of educational programs.

2. Pilot tests are not useful for activities which are (a) not replicable and (b) not revisable.

3. Pilot tests are most useful for activities which are (1) most crucial to the overall success of the guidance programs and (2) about which there is most uncertainty in terms of desired outcome attainment.

4. Eight steps in reviewing internal logic of an activity.
   a. Do the activities relate to the desired outcome?
   b. Do the activity's outcomes relate to the program's goals?
   c. Can the target population already perform the activity's objectives?
   d. Can outcomes be documented?
   e. Is a time sequence indicated?
   f. Are the planned activities and sequence practical?
   g. Are the planned activities and sequence appropriate to the level of the target audience?
   h. Are there likely to be unanticipated side effects?

5. The general strategy of pilot testing is to implement the planned activity under well-controlled circumstances, using members of the target population, collecting (a) objective information on degree of attainment of process and student performance objectives, and (b) any other objective or subjective information which is likely to be useful in indicating changes that can potentially be made in the activity. Care should be taken to avoid over-collection of data so that immediate diagnostic feedback can be obtained.

6. Small samples from a population are rarely representative of the extremes of that population.

7. An evaluation design is an arrangement of persons, activities, and measures of effects such that inferences can be made about the probable effects of the activities on a larger group of similar persons.

8. Considerations of measurement adequacy.
   a. Objectivity. The degree to which a technique will produce the same score regardless of who is applying it.
   b. Reliability. The degree to which a technique will produce data which are free from random error and thus yield a relatively "constant" score.
   c. Validity. The degree to which a technique measures what it is supposed to be measuring, thus yielding a "true" score.
   d. Efficiency. The cost, in terms of money, time, and effort, in administering and scoring a technique.
   Or Reactivity. The degree to which technique influences the subsequent behavior of those to whom it is being applied.

9. Improvements in each category above.

   a. Objectivity. Develop scoring keys or rules. Remove identifying information from responses.

   b. Reliability. Increase objectivity. Equalize testing conditions for all examinees. Discourage random responding, e.g., wild guessing. Use familiar and uncomplicated response forms. If possible, combine measures to produce a composite score.

   c. Validity. Improve reliability. Construct explicit rationales for each technique. Measure all aspects of an objective or characteristic.

   d. Efficiency. Use group measures. Administer sparingly. Avoid contrived circumstances. Use scoring aids.
   Or Reactivity. Use physical traces, archival records, and unobtrusive observations.

10. a. Personnel costs

    b. Equipment and materials

    c. Other direct expenses

    d. Building costs

    e. Fixed administrative costs

ASSESSMENT CRITERIA

Outcome 1. Measurement: Postassessment items 1 and 2.
Criteria:

Item 1. The sentence should essentially state the following: "To bring about small improvements in important activities which are meant to be replicable and about which there is some uncertainty."

Item 2. Useful instances:

1. When the activity is replicable.
2. When the activity is revisable.
3. When the activity is important.
4. When there is considerable doubt about the effects of the activity.

Any two of these four constitute an acceptable response.

Waste of time: The reverse of any two of the above four.

Outcome 2. Measurement: Activity 1.
Criteria: As specified in the activity feedback, the written analysis must contain evidence of consideration of the following points:

a. Are the activities described specific enough to be observed? Can their attainment be *documented*?

b. Is there reason to believe that most members of the target audience can *already perform* the student performance objectives, making the planned activities unnecessary?

c. Is there any indicated *time schedule* or sequence for multiple activities?

d. Are the planned activities and sequence practical with regard to the constraints likely to exist in this situation? Are they appropriate to the level of the target audience?

e. Are there likely to be *side effects* which the program developers have probably not anticipated?

f. Do the planned activities *relate* to the intended student performance objectives? Is there a reasonable chance that the activities, if carried out as indicated, will produce the desired outcomes? Do the proposed student performance objectives relate to the program's stated goal or goals?

Outcome 3. Measurement: Activity 2.
Criteria: The activity feedback section provides an extensive example of approximately what the cost estimates should look like. Make sure the participant has completed the worksheet and estimates and that the estimate approximates that provided in the feedback.

Outcome 4. Measurement: Activity 3.
Postassessment item 3.
Criteria: The skill that is important to judge here (in both the activity and postassessment) is the ability to write an item which measures the attainment of a process objective. In both cases, the item should be a restatement of the objective, with "Yes, it was done," "No, it wasn't" options.

Outcome 5. Measurement: Activity 3
Postassessment items 4 - 7
Criteria: The items developed in Activity 3 and on item 4 of the postassessment should be some variation on the following themes:

a. Paper and pencil test item. "List all the minimum entry requirements for Abacus School of Electronic Computer Programming."

b. Performance test item. "What are the minimum entry requirements for Abacus Tech?"

c. Observation. How long does it take each student to read the handbook? Do students take the handbook with them or wad it up and toss it in the trash can?

d. Interview. "What did you think of it?"

Postassessment items 5 - 7 should respond as follows:

Item 5 - Any indication that (a) scoring rules (a test key) should be constructed, and (b) the scorers should not be made aware of the identity of the respondents.

Item 6 - Any negative response associated with the notion that they test knowledge and attitudes outcomes related to only *one* of *seven* vocational training programs in Fliverville. The validity could be improved by writing items which measure outcomes related to the other six vocational training programs.

Item 7 - Ask everybody the same question, call at approximately the same time of day (or evening), chat with everybody for a few minutes before asking questions, ask several questions trying to get at the same point, etc.

Outcome 6. Measurement: Application plan produced
Criteria: The time and task analysis produced should line out the tasks under each of the major headings listed, assign responsibility for each task to someone, and set a completion date for each task. In addition, the plan should be reasonably

1) *Logical* - do the tasks flow in logical sequence?
2) *Thorough* - is it detailed enough to be helpful?
3) *Feasible* - is it not too detailed to be burdensome; are the times allowed for the tasks reasonable?
4) *Fair* - are the responsibilities assigned equitably and fairly?

On the next five pages are two sample instruments, the Module Performance Record and the Evaluation Questionnaire for Staff Development Workshops. You may wish to use these instruments to gather information for evaluating any workshop in which you administer this module, and for making decisions about future workshops. The Module Performance Record (MPR) is a form for tallying participants' achievement of objectives. The Evaluation Questionnaire seeks participants' opinions on four dimensions: (1) perceived value of the workshop; (2) effects of participating in the workshop; (3) role and performance of the coordinator; and (4) recommended improvements in the workshop. As it now stands, the questionnaire should take participants 10-20 minutes to complete. You, as module coordinator, should complete the MPR form based upon the results of the postassessment or other evidence supplied by participants. If you duplicate the Evaluation Questionnaire for participants to complete, we suggest you print it as a four page booklet.

NATIONAL CONSORTIUM ON COMPETENCY-BASED STAFF DEVELOPMENT

MODULE PERFORMANCE RECORD

MODULE TITLE: _____

WORKSHOP DATES: _____

WORKSHOP COORDINATOR(S): _____

_____

Participants' Names (Alphabetically)	OBJECTIVES (Place a check (✔) mark for each objective achieved.)						
	1	2	3	4	5	6	7
1.							
2.							
3.							
4.							
5.							
6.							
7.							
8.							
9.							
10.							
11.							
12.							
13.							
14.							
15.							

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare.

NATIONAL CONSORTIUM ON COMPETENCY-BASED STAFF DEVELOPMENT

EVALUATION QUESTIONNAIRE FOR STAFF DEVELOPMENT WORKSHOPS

Your responses to the brief questions in this booklet will help us evaluate the workshop you just completed and make decisions regarding future workshops. Please take 10-20 minutes to answer honestly and thoughtfully. You need not sign your name, but we do need your help. Please answer each question. Thank you.

Name (Optional) _____ Date _____

Module Title _____

A. General Issues Related to the Workshop

Respond by checking the column (A, B, C, D, or E) of the statement which best expresses your feeling or opinion on each item in the following list. If none of the possible choices precisely represents your view, pick the one that comes closest.

STATEMENTS	Column A: I strongly agree	Column B: I agree	Column C: I disagree	Column D: I strongly disagree	Column E: I don't know. I have no basis for answering.
1. This workshop made a valuable contribution to my professional development.					
2. I acquired no new knowledge during this workshop.					
3. I would not recommend this workshop to anyone else.					
4. I am glad I attended this workshop.					
5. In this workshop, I experienced at least one positive change in my knowledge, attitudes, or skills.					
6. As a result of this workshop, I expect that I will help improve the career guidance program in my work setting.					
7. I experienced at least one negative effect from this workshop.					
8. This module must be improved.					
9. The coordinator was an asset to this workshop.					
10. The coordinator was unprepared.					
11. The coordinator was poorly organized.					
12. The coordinator was clear and to the point.					
13. The material and activities in this workshop were not helpful.					
14. The material and activities were applicable to my needs.					
15. The material and activities were routine and boring.					
16. The workshop's objectives addressed the training needs that I need they could.					

Developed by the National Consortium on Competency-Based Staff Development in cooperation with the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

B. Please list and briefly describe up to three major positive changes that you have experienced in your knowledge, attitudes, or skills because of this workshop. Continue on the back of this booklet if necessary. If you did not experience any positive changes, please check the appropriate space.

_____ There were no positive changes.

C. Please list and briefly describe any negative effects you have experienced because of this workshop. Continue on the back of this booklet if necessary. If you did not experience any negative effects, please check the appropriate space.

_____ There were no negative effects.

D. Please list and briefly describe any improvements you anticipate in your career guidance program as a result of this workshop. Continue on the back of this booklet if necessary. If you don't expect any improvements as a result of this workshop, please check the appropriate space.

_____ I don't expect any improvements in my career guidance program as a result of this workshop.

E. Please list and briefly describe any other comments on this workshop, criticisms of it, or suggestions you have for improving it. We are especially interested in your ideas on topics or activities that should receive more or less emphasis. Continue on the back of this page if necessary.

78