DOCUMENT RESUME

ED 043 136 24 EA 003 105

AUTHOR	Crawford, Jack, Ed.
TITLE	CORD National Research Training Manual (2nd ed.)
INSTITUTION	Oregon State System of Higher Education, Monmouth. Teaching Research Div.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C. National Center for Educational Research and Development.
BUREAU NO	BR-9-I-002
PUB DATE	69
GRANT	OEG-9-9-47002-0006
NOTE	598p; A project of the RED TRAIN Program
AVAILABLE FROM	Teaching Research Division, Oregon College of Education, Monmouth, Oregon 97361 ($12.50 or $15.00 with Workbook: EA 003 106)
EDRS PRICE	EDRS Price MF-$2.25 HC Not Available from EDRS.
DESCRIPTORS	Administration, *Behavioral Objectives, Data Analysis, Educational Change, Educational Philosophy, Educational Specifications, Evaluation, *Instructional Design, *Instructional Improvement, Instructional Innovation, Instructional Programs, Measurement, Methodology, Research, *Research Design, Research Proposals, Statistics, Systems Approach
IDENTIFIERS	*Consortium Research Development Program

ABSTRACT

The purpose of this manual is to present some principles useful in the systematic improvement of instruction. Nine chapters present a methodical approach to the design, implementation, and measurement of the effectiveness of an instructional system. The first four chapters concern themselves with the development of an instructional system, beginning with the statement of behavioral objectives and continuing through to the trials and modifications of the instructional design. The next three chapters outline methodological principles and techniques to measure the effectiveness of such an instructional improvement program. The last two chapters round out the research perspective of the document by dealing with the administration of a research organization and successful research proposal writing. A related document is EA 003 106. (RA)

CORD

National Research Training Manual (2nd ed.)

edited by Jack Crawford

A project of the RED TRAIN program

a research, evaluation, and instructional development training program of the Teaching Research Division of the Oregon State System of Higher Education

EA 003 105

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

James H. Beird

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Copyright, 1969, Teaching Research
Division of the Oregon State System of Higher Education

ED0 43136

Acknowledgement

The development of this manual was supported by a grant from the Regional Research Program, Bureau of Research, U.S. Office of Education. Appreciation for continued support of this project is extended to Dr. Paul Messier, Director of Regional Research Programs, U.S.O.E., and a special thanks goes to Dr. Walter Hirsch, Director of Educational Research in Region 9.

ERIC Full Text Provided by ERIC

NATIONAL RESEARCH TRAINING MANUAL

Second Edition

Table of Contents

i I II III IV V VI VII VIII IX

ERIC

Introduction

The purpose of this manual, and its accompanying workbook, is to present some methodological concepts and principles useful, and probably requisite to the consistent improvement of instruction. The manual has been composed by ten different authors. Surprisingly, they seem to be in harmonious descant. The manual combines an instructional systems approach with an introduction to those research skills required to design and measure potentially effective change. The first four chapters are directly concerned with the systematic development of improved instruction. Chapters V through IX deal with approaches generic to both research and instructional development.

The outstanding characteristic shared by all ten authors is a fervent missionary zeal. They are mutually rapt in their conviction that all instruction could be substantially improved now, and that the approach offered in this manual presents the logical, practicable, and measurable way to do it.

This conjoined crusading spirit has always affected me with some awe and admiration. I joined this group somewhat later in life than the others, and entered habituated to the university atmosphere in which each professor feels that most of his colleagues are either talking intellectual nonsense or wasting their time on trivia.

The unrelenting intensity of the manual authors might become grating. It doesn't, because conviction is fused with perspective, reasonableness and above all openness to evidence. An unusual blend.

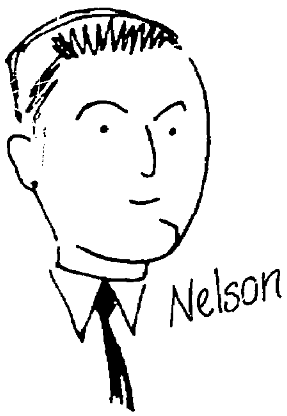
It will be no surprise to the reader that the ten authors work for the same organization. This organization, the Teaching Research Division of the

ERIC Full Text Provided by ERIC

Oregon State System of Higher Education was created by the state system several years ago to somehow improve instruction through the art and application of research. The organization has grown from a staff of two to a staff of a hundred people who are doing the kinds of research, evaluation, and instructional development activities necessary to accomplish this, and whose basic concepts are represented by the following chapters.

Each chapter is accompanied by workbook exercises, which are bound separately. The exercises are designed to start developing your skill in applying the concepts of the chapter. To facilitate your use of the workbook while reading this manual, a device has been used to let you know at what points you might leave the manual and dig into the workbook. It consists of a shovel accompanied by the number(s) of the workbook exercise(s) you might attempt. For example, if you saw the symbol shown here, you would leave your reading and turn to Exercise 7.1 in the workbook. Several of the individual chapters are now a part of an encompassing instruction system. Readers who are interested in developing those competencies that cannot be attained by simply reading a chapter on the topic, are advised to examine the complete instructional packages in those areas. The workbook represents a sampling of only limited dimensions of the instructional activities available in the complete packages.

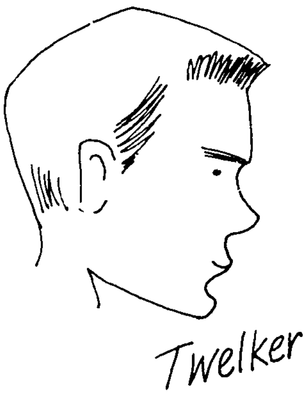
7.1



The first chapter, Frank Nelson's and Bud Paulson's chapter on Behavioral Objectives, deals with what is usually the first and foremost stumbling block in improving instruction. Frank attempts to place behavioral objectives in the perspective of other goals and purposes,

to identify the distinguishing attributes of behavioral objectives, their function, components, and sources. Frank, himself, is hard-working, hard-nosed (at least he is to me, when he serves as the evaluator on my projects), and often carries a major part of the world's weight on his shoulders. The chapter reflects his thoroughness, breadth of outlook, and good sense.

The problem of arriving at the specifications of the components of an instructional system, is handled by Paul Twelker. Paul takes the reader



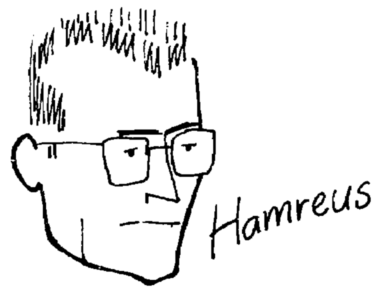
through a detailed account of how to specify the components of the instructional system, from terminal objectives through space and hardware requirements. I found it exhaustive, exhausting, and worth the effort. Paul has developed this chapter into an individualized self-instructional unit. The unit provides for higher levels of competencies than can be attained by reading a chapter. As you may infer from reading his chapter, Paul is quite bright and keenly analytic. He is also distressingly young. His ardor for detailed specification and systematic analysis of all topics is tempered by a barely restrained, ebullient sense of humor. Humor as often directed at his own systems, as to the precious idols of his colleagues.

Dale Hamreus presents both a model and an overall view of the Instructional Systems Approach. The model presented is one of the clearest introductions to his systems approach.

The frequency with which it is copied and noted, with and without his permission is a compliment to Dale. In one sense, the Behavioral Objectives

chapter and the Instructional Systems Specifications chapter represent requisites within the development of an instructional system. In this perspective, Dale's chapter encompasses both.

Dale writes and talks about instructional system development with incessant conviction. He is a true believer in the approach. Yet, he is a practical, real-life developer who realizes the limitations of present knowledge and the tentative nature of all solutions.



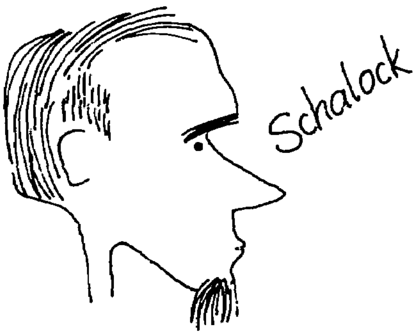
I've always felt there must be something in the systems approach when I met Dale's family. He has the nicest and most uniform set of red-headed daughters I've ever seen in one home (I think there were nine at the last count).

The chapter on Evaluation was written by Bud Paulson, who has developed some refined and systematic notions on this sticky topic, which should place him in the forefront of the field. The principles Bud presents should be as useful to those who need to develop evaluation plans for an existing instructional system, as for those who are developing new systems.

Bud tends to play the role of remonstrating conscience and intellectual worry-wort in the group. Always thoughtful, carefully reflective, Bud is usually attuned to the risks and blunders the rest of us often overlook.

Del Shalock presents some general concepts about measurement which go well beyond the usual

psychometric ideas, e.g., reliability and validity. His extensions above current standards include not only fundamental attributes of measurement, but techniques as well. To fully appreciate this chapter, I suggest the reader take a quick review of the conventional treatment of reliability and validity.



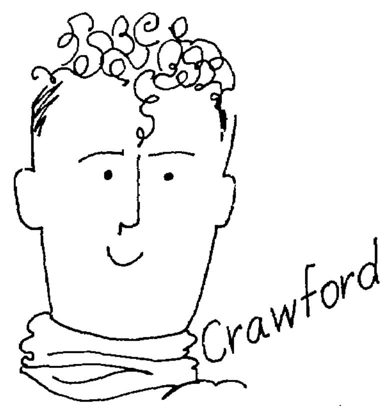
I always appreciate occasions when Del is able to talk with me. He is usually out in the cosmos developing schemes to unify all inter-galactic educational knowledge. Since my own ambition tends to be limited to something like sticking my toe in a puddle and observing it, the fact that we do communicate well and easily is pleasantly astonishing.

Jim Beaird presents a unique and useful decision tree for the selection of data analysis tools. With minimal, or no statistical competency, the reader of this chapter can learn to make intelligent decisions in choosing an appropriate data analysis procedure. This chapter is probably the most systematic and programed of any in the manual. For once, I find myself agree-ing with an author that his writing should be, and is, crystal clear.

Jim is a cool and careful head. In fact, he is now head of Teaching Research. His ability to pare a problem to its essentials, examine all alternatives, and come to prompt decisions is reflected in the chapter. We lost a good researcher when Jim made the agonizing decision to become an administrator.

The experimental design chapter was written by Cathy Kielsmeier and myself. The chapter focuses on the information yield and relative cost of certain generic designs. It is non-mathematical. We have used the same approach to design with high school juniors, successfully. The chapter is part of a package. To learn to build adequate designs we suggest the entire package. However, the chapter will present some useful concepts--and after reading it, you should be able to <u>talk</u> about them.

Cathy has contributed sixty percent of the ideas and ninety percent of the work to our joint effort. The right kind of colleague. In addition, she is the only really good looking author in the entire group. The other nine are uniformly ugly.



Jack Edling, the founder of the Teaching Research Division, addresses himself to the problems of Research Management. The chapter is brief, to the point, provocative, and <u>tough</u>--a mirror of the author. Jack knows more about building and managing a research and development organization than anyone I've met. He built the division from nothing. Yet, it was difficult to get him to write extensively about the subject. He felt that a summary statement could contain the useful messages he had to offer. A contrast with the other authors, most of whom were eventually gagged.

That this manual exists at all is due mainly to Clark Smith. Clark built the workbook and shaped more of the content of the chapters than each author cares to admit. Section by section and page by page he cajoled and bullied initial production and then, unobtrusively suggested clarifications (which were then, unobtrusively and without acknowledgement, adapted). That he accomplished this and simultaneously increased the friendship and respect of the authors, describes Clark.

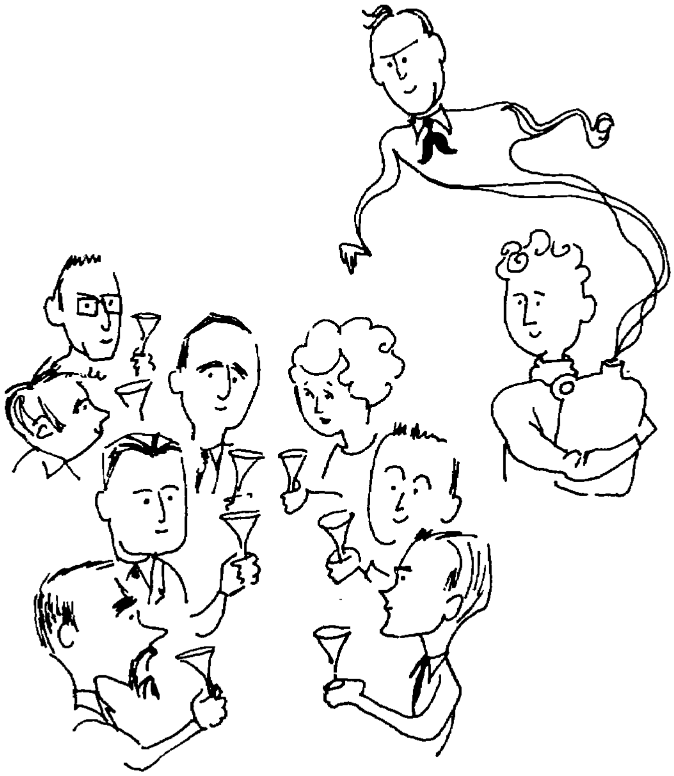
The proposal chapter represents an attempt to present the reviewer or field reader perspective to the proposal developer. It, too, is part of a fairly complete instructional system now. In the chapter you can examine the crucial components, typical weaknesses and weakness prevention techniques, and identify some criteria of a proposal.

That's the thumbnail version of the chapters and authors. We present this manual to you as a tentative draft, the best we have put together as of today. Most chapters contain a set of objectives for the reader, and an annotated list of references for further reading. A perusal of the objectives may give you a better idea of how deep you wish to plunge into each chapter. Although we believe these chapters hang together in an integrated way, each can be read separately. To assist the less experienced reader through this

volume, the editor has prefaced each chapter with some guiding comments. These should be taken with discretion.

In as much as the ten contributors appear to continue in fervent enthusiasm, and seem to be in moderately good health, I suggest that if you have any difficulty in wading through any of this--inform the author. He will respond. Of course, remember that you may have difficulty extinguishing his response.

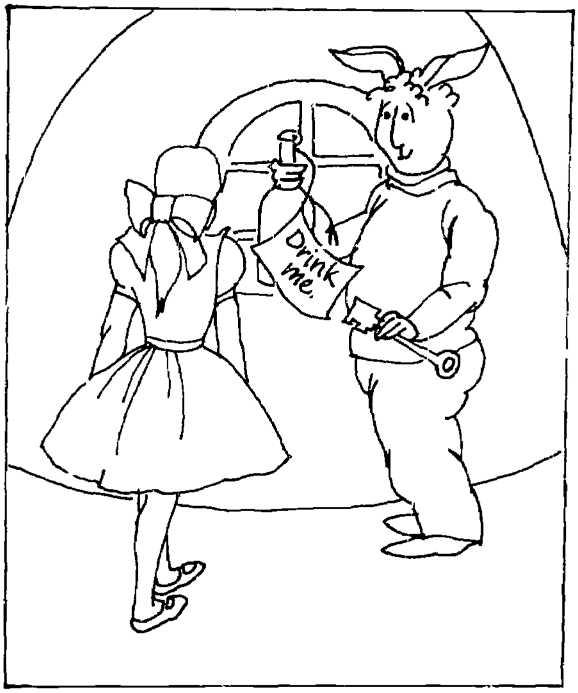
Good reading. Whenever you happen to be in the vicinity of Monmouth, Oregon, drop by.



VIEL GLUECK, ERFOLG und SPASS

Jack Crawford

ERIC



"For, you see, so many out-of-the-way things had happened lately that Alice had begun to think that very few things indeed were really impossible."

Section I: Behavioral Objectives

Editorial Foreword

This initial chapter by Frank Nelson and Bud Paulson presents one component in the development of instruction. This single component, specifying behavioral objectives, is the cornerstone of the entire approach presented in this manual, yet it has so eluded educators that the state of the art is aptly summed, "Behavioral objectives are more written about than written."

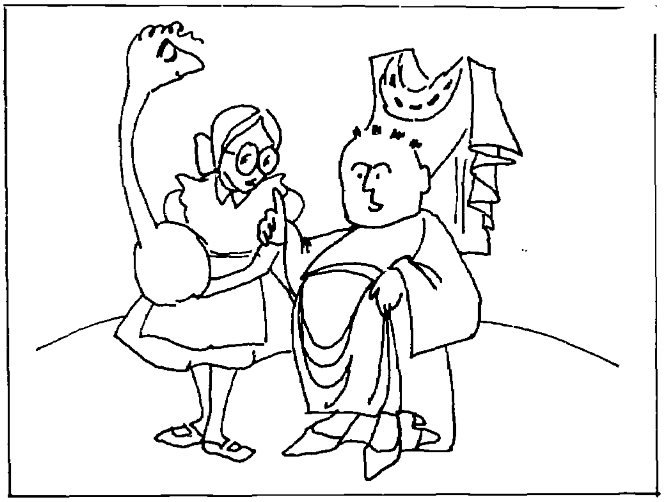
The chapter presents:

1. A rationale for importance of specifying objectives in terms of overt learner behavior.
2. The place of such behavioral objectives in the context of the broader goals and values of education.
3. Probable sources for discovering appropriate behavioral objectives.
4. Essential attributes of behavioral objectives, presented in a four-step check list.
5. Advice on the writing of such objectives.

The chapter repeatedly emphasizes the need to frame instructional objectives in terms of learner performance. Several useful cautions are given by the authors; e.g., the danger that relatively trivial objectives will be selected primarily because they are overt, specific and easily measured; and, the typical neglect of affective objectives.

Probably editor and reader both would have rejoiced if the authors had woven more exemplars into the text. At least the reader will find non-exemplar

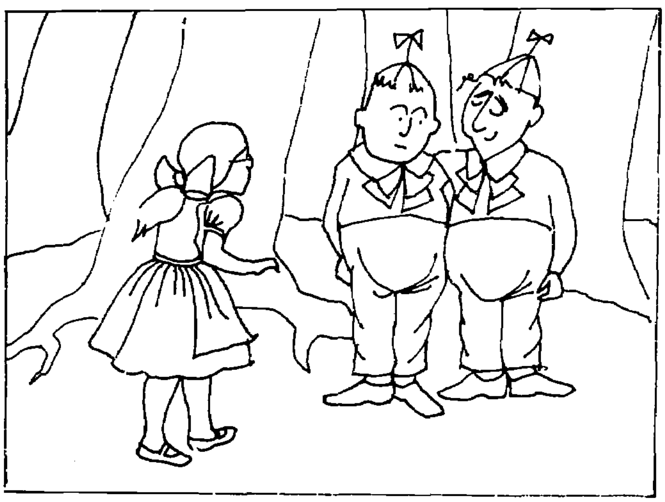
instances permeating instruction throughout his real environment.



"Tut, tut, child. Everything's got an objective if you can only find it."

A point, not in conflict with the authors' views, but scarcely emphasized, is that a behavior may be specified but used merely as an indicator that an objective is reached. In itself, that behavior is not the objective of instruction, but one of a possible set of signs that indicate that an underlying condition or process is attained.

The reader should be cautioned that new disciples of the "identify your behavioral objectives" approach become a bit fanatic. You may find yourself defending the legitimate existence of your own goals.



"You know very well you're not real--and you won't make yourself a bit realer by crying."

Alice: "I am real!"

Casper F. Paulson and Frank G. Nelson

Behavioral Objectives

Preface

Much of the current emphasis within educational research and innovation focuses on the idea of purposiveness of human behavior. This is evident in new instructional programs, new textbooks, and some forms of new media that are currently being published. This concept permeates our entire work life but has been largely ignored by educationists. Whenever we begin a task we frequently have a somewhat precise image of what we want to do and how we want to do it. We may arrive at the desired point even without knowing precisely where and how to proceed, but certainly we will not attain the desired degree of success or completion with any degree of efficiency. Unless there is a clearly defined end point, we may not even be able to identify when we have arrived. The homeowner, do-it-yourself type, has probably experienced this phenomenon when he attempted to build a cabinet or other object without a plan; it is difficult for him to know when he has achieved precisely what he intended to achieve. The project seems to be subject to continuous modifications and chances are he is left without the feeling of satisfaction of completion. On the other hand, if a rather precise plan had been developed and a systematic effort launched to complete the object, the finish would probably come in far less time, leaving the individual with a feeling of closure.

The point is, that we may arrive at the desired point or status by not knowing precisely where and how to proceed, but this procedure will not enable us to follow the most effective or most efficient route. To satisfactorily begin and complete a task we need to know what the task is and

what it looks like when it is finished. Only when success is clearly defined and identified can we ever be certain that we actually have achieved what we intended to achieve.

Objectives, in this sense, serve to orient both the instructor and the learner. They provide firm criteria by which a learner's progress may be assessed or his level of competence determined at a specific point in time. Without this information it is difficult, if not impossible, to outline a course of instruction or determine when and if a learner has attained the desired level of competence. Objectives, expressed in behavioral terms, are useful in this respect in that they clearly define the intended or desired behavior of the learner.

The ability to write and to use behavioral objectives therefore, becomes a basic skill necessary for anyone who is attempting to design or conduct research on the instructional process. The intent of this publication is to improve both your teaching and instructional research skills. Of prime concern, therefore, is your ability to write and use statements of objectives. Consequently, when you have completed this chapter of the manual, it is intended that you will be able to:

(1) Define verbally each component of the philosophy that guides your educational endeavors.

(2) Identify several sources for objectives.

(3) Define the relationship that exists between various levels or forms of objectives including when and where each type is most useful.

(4) Write statements of objectives that are useful in the design and evaluation of instruction oriented toward both cognitive and affective outcomes.

Introduction

An educational program, as with all purposeful activity, is directed by the expectation of certain outcomes. Education's major responsibility, and consequently the net effect of expected outcomes, is to effect changes in individuals that in some way add to the knowledge they possess, that enable them to perform skills which otherwise they would not be able to perform, and to develop certain understandings, insights and appreciations.

Changes which are appropriate, and desired, may be considered either educational goals or educational objectives, depending upon the level of specificity involved. For example, statements that education should transmit culture, reconstruct society, or provide for the fullest development of the individual merely serve to identify instructional parameters. They outline the broad area of concern.

Identification of goals such as these provides an orientation to the main emphasis in an educational program. Goals on this level normally serve to describe the philosophy of a particular institution. They represent the first step in translating the needs and values of society, and of individuals, into an educational program. While they are inadequate guides for making specific instructional design decisions, they serve an essential function in a total development/design effort. For example, if the instructional management situation becomes the means to expedite subject matter and social competence, the four basic processes probably involved will be : (1) analyzing the characteristics of subject matter competence; (2) diagnosing preinstructional (entry) levels of potential learners; (3) carrying out the instructional process; and (4) measuring learner outcomes. These processes when empirically applied with sufficient expertise assure opportunities for optimal learner achievement.

ERIC

Due to the potential power of this instructional process to effect behavioral change, it is essential that the basic purposes of the larger educational system be carefully defined with this definition process basically a series of decisions regarding what is valued and what should be taught. Strong conceptualization of goals permits an educational program to be planned and systematic efforts made toward continued improvement. As noted earlier, objectives, at this level are not for facilitating the making of specific decisions, but rather for providing a basis for specifying objectives which can. Such derived objectives serve as criteria for selecting materials, outlining instructional content, developing instructional procedures, and preparing tests and examinations.

This process of defining educational goals and objectives is a social process continually influenced and conditioned by political and economic realities. Decisions about what to teach are manifestations of what is valued by those responsible for the school. To the extent that such value judgments by school people are different from judgments by those of the societal segment served by the educational institution, there may be cause for serious conflict. It behooves school personnel therefore, to make these judgments within the framework of a comprehensive philosophy which is sensitive to the society it serves, to the children toward whom efforts are directed, and to the disciplines involved.

The philosophy expressed by value judgments sets forth for the individual and the system in which he functions, the nature of a good life and a good society as perceived by those being served. Careful definition of this philosophy establishes the parameters of the value system within which all aims and objectives of the educational program are to have relevance

ERIC

and typical values against which they will be evaluated. By way of example, typical values for an effective and satisfying social and personal life within a democratic society might include an emphasis upon the importance of the individual regardless of his social and economic status, freedom to interact within and with the society, differentiating rather than stereotyping individual personality, and finally a belief in the ability of the individual to initiate action as opposed to reliance upon an authoritative power to deal with important problems.

The philosophical statement, in addition to establishing value parameters, provides a point of origin in the task of building reliable objectives for the educational program. It "suggests the kinds of desired behavior patterns, that is the types of values and ideals, the habits and practices which will be aimed at" (Tyler, 1966, p. 22). Acceptance and adherence to these value dimensions by the educational system facilitates retaining objective statements.

The basic purposes of any educational program stem from the formulation of a philosophy. Practice translates them into reality, With a sound philosophic base from which to build, an effective collection of objectives may be created. The importance of this base cannot be over emphasized. Taba (1962, p. 211) comments:

> An organized statement of objectives should be more than a mere grouping of individual objectives. It should also convey the fundamental rationale on which the very conception of objectives is based. This rationale should indicate what is important in education and where the subsidiary values lie. Such a statement should be useful in establishing priorities in the grand design of the curriculum as well as in the smaller decisions such as those about sampling content for particular units or whether to spend time on analyzing historic documents. It should yield some criteria for the scope of the educational effort and set

ERIC Full Text Provided by ERIC

some limits for the specificity or depth desired. In this sense an organized statement of objectives expresses the philosophy of education of a particular school system or of a particular school.

General aims such as those which are identified in the educational philosophy of an institution may be satisfied when the individuals served by that institution acquire certain knowledge, skills, techniques and attitudes. These latter represent a more specific definition of goals and as such are usually referred to as educational objectives.

What is a Goal

Let's turn first to the definition of an educational goal. Most often a goal is defined as an aim or a purpose, an object toward which an organism strives with conscious or unconscious purpose that guides activities toward a specified end. The important concepts from this definition are aim and purpose in the guidance of activities. Earlier, the importance of establishing a viable philosophy for the institution was discussed. This philosophy enables personnel within the institution to establish goals or orientations which have consistent direction. Goals seldom achieve a more precise definition than that which has been offered here. In most cases the desired end will be stated in extremely general terms, so general that attempts to communicate specific intent are wanting of explanation, and attempts to operationalize them without elaboration is next to impossible. It is imperative therefore that some distinction be made between the goals of an educational enterprise and the specific objectives which guide the components of that enterprise to fulfillment of the goals.

Objectives

Objectives, as compared to goals, are relatively explicit formulations of the ways in which students are expected to change within the education

process. Good provides us with a rather comprehensive definition which, when analyzed, clearly indicates what an objective is, the kind of objectives we should consider, and the functions of objectives and instruction. He states "An objective is an aim, an end in view, or purpose of a course of action. It may be a belief. It is that which is anticipated as desirable in the early phases of activity and serves to select, regulate, and direct later aspects of the act so that the total process is designed and integrated" (Good, 1959, p. 371).

Three important dimensions are implicit in Good's definition of an objective. First, an objective must be visible. If one is to establish guidelines for a course and identify anticipated outcomes then one must be able to recognize when and if an objective has been attained. Secondly, if an objective is to serve a design and integration function, then there must be some control over the conditions surrounding its attainment. As Good notes, it serves to "select, regulate and direct the various dimensions which surround its attainment." Finally, there must be some commitment to its attainment before it actually becomes an objective. The number of potential objectives for a given course is infinite. Statements of intent only become objectives when we set out to achieve them, wherein the difference between commitment and aspiration is clearly shown. It is in this respect that a careful definition and delineation of one's philosophy is critical, for it is full acceptance of the philosophy that facilitates commitment to attainment. It is the difference between commitment and aspiration that carries the mountaineer that last impossible one hundred yards to the peak and that makes the effective teacher pursue a point until his students demonstrate that they have learned.

The notion of objective visibility needs some additional emphasis. Obviously, just being able to recognize something of value can be of considerable importance. As educators, however, we seek not simply to recognize valued dimensions in the total education endeavor. Rather our task is one of naming those things which we consider valuable and then labeling them in such a manner that they are easily recognizable. An educator, when he knows what valued behavior and success look like, is in a better position to pursue them.

A true objective then must possess these characteristics:

(1) It is defined clearly enough that we can recognize it.

(2) We can carry out whatever activities are necessary to make it occur.

(3) We seriously intend to achieve it even at considerable cost.

By now you are undoubtedly aware that an objective must represent some point or event that is identifiable. You must be able to tell when you have arrived, and if others are expected to pursue the same objectives they must likewise be able to determine when they have arrived. This adds another important consideration in the development of objectives. Not only must objectives make the behavior of intent visible, they must also be explicit enough to communicate this information to someone else. If we are to ensure systematic implementation of objectives, then we must be able to inform colleagues and students of our intentions.

Additionally, as with the number of potential objectives, the number of possible or desirable outcomes of instruction is infinite. In all likelihood, however, we will not have direct control over all types of outcomes and consequently the number of potential objectives is much smaller than the number of possible or desirable instructional outcomes.

ERIC

What are Some Different Kinds of Objectives?

Numerous objective classification systems exist, far too many to possibly list them all here. For us the most important distinction that must be made in objective forms is the difference between behavioral and non-behavioral forms.

Earlier, a differentiation was made between goals and objectives. Goals were described as derivatives of the philosophy held by the educational system and serve to provide the orientation or direction for that system. Objectives were described as operational definitions of the goals. It follows, therefore that objectives can be specified in several ways. We can specify those learning activities that we intend to take place, or the instructional activities that we intend to perform; but these are actually a means to an end--student learning. The total educational enterprise is concerned with changing people in some way, with developing characteristics that were not there previously. Frequently, however, the changes we seek to accomplish are difficult to observe, and unless we clearly identify the changes we seek to effect, there is no basis for determining whether or not they have occurred. Consequently, unless our objectives meet the criteria outlined above, and are cast in a behavioral form, instructional decisions are likely to be arbitrary.

While our intent is to make more visible those elements of the instructional process which are of most interest, it is mandatory that those which we feel should be of most interest be specified prior to the beginnings of instructional design. For example one of the myths of science is that the confident scientist decides beforehand what he wants to do, establishes a hypothesis, selects or builds a design to test his hypothesis and then

systematically implements this design. While the scientist may work this way, he is not operating under such a requirement. Instead, even though final reports may cite hypotheses and designs implemented, the scientist is normally free to play with ideas and speculate on tentative hypotheses. He may even discard a hypothesis when some more interesting phenomena appear in his work and he is sensitive enough to realize that it has greater potential than his current hypothesis. In many instances, a scientist may come to his hypothesis after, rather than before he has conducted a great deal of his research. Similarly, in the types of development frequently employed in education, freedom to explore before building elaborate lists of objectives does not necessarily indicate the lack of precision in one's approach. Instead it may be the most effective way of operating.

The above is not intended to imply that you should not have some direction or some notion of what you intend to do within a given project. It merely recognizes the fact that instructional design needs to be started in an atmosphere of some freedom rather than one that is severely constrained by too many, too precise, objectives.

Levels of Objectives

An effective operating procedure is to establish objectives on several levels which may represent a hierarchy of objectives. Another way of looking at the levels may be as an inverted cone, beginning with broad philosophical goals and moving to progressively more specific, behavioral, types of objectives.

Three levels appear to be most appropriate and useful in the context of systems development. First, and as discussed earlier, we need to be concerned with the establishment of a philosophy and statements of general goals.

ERIC

At the next level, objectives serve primarily to describe school related outcomes and concomitantly assure consistency of the program with the defined value system of the larger institution. In this way, goals can be translated into _operational objectives_. It is not necessary for operational objectives to be stated in behavioral terms; however, the more clearly these statements describe what the student must be able to do, or what must be the outcomes of instruction, the more readily can we derive our next level of objectives.

The two levels of objectives discussed above are perhaps most typical of the types of objectives specified by a school system. Most often, they are described only as actions of the teacher or the larger institution. Seldom, if ever, are objectives at this level concerned with the types of behavior which a learner must demonstrate in order to satisfy that objective.

Since education is charged with the responsibility of changing people in some way, with developing characteristics that were not there previously, instructional decisions should be based on evidence of such change. Landmarks such as topics covered, pages read, or time elapsed have a way of becoming firm objectives. These types of objectives, however, relate to something other than the learner himself. Even a cursory examination will normally reveal that it would not be necessary for learners to be present in order for an instructor to implement the objective.

To avoid this pitfall it is important that we render learning achievements more conspicuous. The third level of objectives then, is concerned with specific instructional objectives, or _behavioral objectives_. As noted, we can specify objectives and the instructional activities that we intend to perform in several ways. But these are actually only means to an end, which is that the student will "learn" something.

ERIC

The real challenge in constructing behavioral objectives is not to focus interest on what is readily apparent but to make more apparent that which is of most interest or value. These efforts may require a more careful structuring of students' environment so that their actions are more readily observed, or more careful structuring of our observations so that we are more sensitive to significant aspects of the student's behavior.

The preceding paragraphs have been concerned with establishing the notion that objectives at several levels of abstraction are useful and important in the educational process. Krathwohl (1964) has identified four reasons why these various levels of objectives, or analysis, are useful and needed in instructional design. Specifically, they are:

(1) Each level of analysis permits the development of the next more specific level.

(2) Mastery objectives can be analyzed to greater specificity than transfer objectives.

(3) Curricula gain adoption by consensus that what is taught is of value. Consensus is more easily gained at the more abstract levels of analysis.

(4) There are usually several alternative ways of analyzing objectives at the most specific level. Objectives at the more abstract level provide a referent for evaluating these alternatives.

The levels of analysis that should be considered in formulating instructional objectives are largely dependent upon the specificity required in the intended context. In developing programs of instruction the first level serves to identify types of courses and areas to be covered, the general goals desired toward which several years of education might be aimed,

i.e., an entire unit such as an elementary, junior, or senior high school.

At the second, somewhat more specific level, operational objectives assist in analyzing broad goals into refined components which are useful as the building blocks of a given curriculum. Most often this level of objectives will be used to specify the outcomes of a sequence of courses, for example, intended learner outcomes at the conclusion of a course in high school or elementary school.

Third, and finally, a level is needed which identifies specific learner behavior. Objectives at this level may serve to identify who the learners are, the behavior they are to be able to demonstrate, the conditions under which that behavior shall be manifest, and how the behavior is to be evaluated.

Operationally, this third level of objectives enables the teacher, or the instructional designer, to make decisions on what to cover, what to emphasize, what content to select, and which learning experiences to stress. This level contains the heart of educational objectives in their most precise sense. Clarification of the functions of objectives on this level is essential to arriving at a serviceable guide to system development.

Two general rationales that may be adopted in the establishment of objectives at the behavioral level are predominant. One of these requires objectives to be stated in terms of observable and measureable immediate behavior. The second outlines a taxonomic structure for objectives in terms of three large domains, specifically, the cognitive domain, the affective domain and the psychomotor domain.

Proponents of these two approaches do not deliberately see them in juxtaposition. However, it does appear that the two rationales can be mutually supportive. For example, in the taxonomic classification system

ERIC

a typical objective might be "the student will understand how to use a library." Under the second objective rationale such an objective might be stated: "When given the full name of the author, the sixth grade student can locate and record in his own book the call numbers of all holdings in the library, without assistance from librarian or other students." In practice, one might logically use both approaches, specifying the desired taxonomical classificaiton as a part of the objective. Unfortunately, most persons when they adopt a behavioral objective approach, focus their attention only on those behaviors from those lowest levels of the cognitive domain. The only satisfactory reason for such objectives is that these types of behaviors are more readily observable, and consequently measurable. If, however, the person concerned with design of an instructional unit wishes to emphasize higher mental processes, then it becomes his responsibility to define these in some observable or overt form. Corey (1967), for example, has noted that if a given behavior is amenable to instruction, then surely it is amenable to analysis at some level of specificity. Our concern, therefore, is to avoid focusing only on those kinds of behaviors which are most readily observable. Our efforts must also be directed to those kinds of objectives which will satisfy the broad or philosophical goals of the larger system and which focus on those types of behavior which we value very highly. Operationally, this requires definitions for desired attitudinal states, value systems, etc.

An additional constraint is that behavioral objectives should be constructed so as to indicate that the learner will be able to accomplish a particular task after some instructional event. In some instances, this behavior may be directly applicable to a life situation. In others, however, he may proceed neither to a specific task nor to some other instructional

ERIC

component but rather the situation may have been to prepare him for some future function. Objectives then, in a behavioral sense, are concerned not only with the immediate goals but also with preparing the students to satisfy the broader, philosophical goals of the educational institution.

For ease in instructional design one might consider the establishment of two levels within the behavioral form of objectives. These, once again, follow the rationale of broad to specific and may be labeled as terminal objectives and enabling objectives.

Terminal objectives are those which identify student action or performance on a meaningful unit. For example, a terminal objective for a math course might be: "At the conclusion of Math 201, the student will be able to satisfactorily derive the components for a truth table making fewer than three errors."

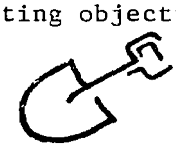
Enabling objectives on the other hand are the component actions, knowledges, skills, etc., which a student must learn if he is to attain the terminal objective. These objectives serve to bridge the gap between where the student is at the beginning of instruction and where he should be upon the completion of that segment of instruction. Enabling objectives define the transitional behavior of the learner within a given instructional unit. As will be noted later, enabling objectives may be established at any taxonomic level. They may be concerned with both cognitive needs and affective needs such as wants, desires, and interests.

The relationship between learner entry levels to an instructional system, terminal objectives and enabling objectives is perhaps best explained in the following diagram:

$\boxed{\text{Entry Level}} \xrightarrow{\text{Enabling Objectives}} \boxed{\text{Terminal Objective} \mid \text{Entry Level}} \xrightarrow{\text{Enabling Objectives}} \boxed{\text{Terminal Objective} \mid \text{Entry Level}} \xrightarrow{\text{etc.}}$

Since terminal objectives are meaningful units, they become the means for organizing instruction. As such, they represent a performance level that is to be attained through instruction. They establish both meaningful and measurable goals for the instructional program and subsequently all other aspects of program development.

Since enabling objectives represent the difference between where a student is now and where one wants him to be, and since they serve as the bridge between the learner entry level and each subsequent terminal objective, knowledge about the terminal objectives and existing capabilities of the learner population is necessary before enabling objectives can be established. It is important to understand that enabling objectives are primarily a device or framework to facilitate instruction and instructional design. If too much attention is devoted to identifying and specifying objectives at the enabling level one will soon become disenchanted due to the lack of meaningfulness in the level of objective specification. Unfortunately, many educators in their initial attempts at specifying behavioral objectives never get beyond the level of enabling objectives. As a result these persons frequently become disillusioned with the whole notion of behavioral objectives, never realizing that they have only begun to explore the full potential of stating objectives in this form.



1.1

<u>What are the Components of an Objective</u>?

Basically, stating behavioral objectives in appropriate form can be described as simply writing a sentence where the subject, verb and modifiers each have certain desired characteristics. To ensure that the information presented here "sticks" for some time in some way, we shall employ a mnemonic device. We shall be concerned with the ABCD's of writing behavioral objectives.

Audience. The "Subject" of the sentence used to describe a behavioral objective should describe who is to be doing the learning. The description should be in terms most relevant to the instructional task at hand. Sometimes this description simply will be an identification of the grade level and the subject of the class, for example, seventh grade geography students.

Some care must be exercised in establishing the audience for a behavioral objective. As you may recall one of the criteria in defining an objective was that there be some commitment to its attainment. In the foregoing example, that of a seventh grade geography student, we need to examine carefully, and determine precisely, whether or not we really propose to achieve a specified behavioral objective with all seventh grade geography students. Experience seems to indicate that there are few, if any, objectives that will be achieved by all students. For one reason or another some students inevitably fall by the wayside. If the objective is to be attained by all students then we shall impose some rather severe design constraints upon the instructional program. For example, the instructional unit will have to provide for students with low reading ability, low intelligence, or poor background in the subject matter. If, however, the audience is defined as students with average or better reading ability or intelligence in certain required subject matter skills, then the "audience" must be so defined in the subject of the sentence.

It s important that sufficient relevant information be provided in describing the audience, but at the same time it is nearly as important to avoid being compulsive about it. Unnecessary details should be avoided. Certainly learner populations can be described in great detail if you take the task to heart, but it takes time to write such descriptions and time to read them. The writer should only provide such descriptions as will be useful.

ERIC Full Text Provided by ERIC

Behavior. At this point it may be quite evident that the verb in a sentence that describes a behavioral objective is perhaps the most crucial element. A short cut used by those who edit and evaluate behavioral objectives is to skim through them looking at the verbs. The use of an inappropriate verb is the most frequent and disabling error in writing behavioral objectives. The notion that a verb may be inappropriate is consistent with comments made earlier about the value or meaningfulness of any given behavioral objective. No one would question another person's ability to continue the analysis. There would be some question, however, as to the validity of the micro-behaviors described in such objectives.

The verb should describe an observable action that the students will demonstrate as a result of their learning experience. Some verbs denote readily observable actions, others do not. It is difficult to visualize just what a student looks like when he is understanding, appreciating, or even listening. But it is fairly easy to observe whether or not he is speaking, writing, or constructing. Remember, if a construct is amenable to instruction, then it must be definable at some level such that an appropriate verb might be employed to describe behavior which demonstrates that construct.

Several authors have put together lists which they describe as words which are open to few interpretations. This list was comprised by Desmond Wedberg (   ) who has asserted that all behavioral outcomes can be classified in one of these ten categories. Whether or not this is the case, the list is at least suggestive of useful verbs.

1. Identify	4. Demonstrate
2. Name	5. Describe
3. Order	6. Construct

ERIC

7. State a rule

8. Apply a rule

9. Interpret

10. Distinguish

Robert Mager (1962, p. 11) uses the following two lists of words in juxtaposition to contrast the two types of words which are most appropriate to describe behavior:

Words Open to Many Interpretations	Words Open to Fewer Interpretations
to know	to write
to understand	to recite
to really understand	to identify
to appreciate	to differentiate
to fully appreciate	to solve
to grasp the significance of	to construct
to enjoy	to list
to believe	to compare
to have faith in	to contrast

The reader is referred also to the Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, (Bloom, 1956) and Handbook II: Affective Domain (Krathwohl et al., 1964). These two publications can be extremely useful as one begins to specify appropriate learning behaviors. Handbook I is devoted to a classification of objectives within the cognitive domain. This domain relates specifically to knowledge and intellectual skills relevant to the use of knowledge and follows an organizational scheme of simple to complex. The lowest and least complex in this hierarchy is the acquisition of knowledge. Following this are comprehension (translation, interpretation, extrapolation); application (use of abstractions particularly in concrete situations); analysis (analysis of elements

of relationships in organizational principles); synthesis (organization of elements and parts to form a new whole); and evaluation (making judgments about the value of material in methods for a given purpose). These levels and the terms associated with each level can be useful in defining behavior of interest, but more importantly in identifying the level of knowledge in the particular content area of concern. For example, a word that is frequently disapproved by writers of behavioral objectives is the verb "understand." Handbook I defines understanding in a way that may be useful in writing behavioral objectives. Specifically, comprehension is defined as understanding, which in turn is defined as the ability to translate, to interpret, and to extrapolate a given communication. The following objective, for example, is entirely appropriate if the definitions provided by the taxonomies are employed: "Understanding of the interrelations of chemical principles and theories" (Bloom, 1956, p. 77).

Handbook II, the affective domain, is directed to such concepts as attitudes, values and interests. It, too, is organized as an hierarchy, but uses internalization as the organizational construct. Using the definitions provided in Handbook II, one is enabled to write objectives that speak to such concepts as satisfaction and appreciation. For example, the following objective is from the "receive" level of the taxonomy: "Appreciation of family members as persons with needs and interests of their own" (Krathwohl, 1964, p. 108). The taxonomy goes on to describe how such an objective might be evaluated.

<u>Conditions</u>. Earlier it was stated that the purpose of behavioral objectives was not so much to focus our interests on what is observable as to render more observable that which is of interest. The degree to which we succeed in this will be determined largely by the care and ingenuity employed in

ERIC

describing the conditions of performance. It is here that we describe the setting for the evaluation, the materials and aids the student will be given, and the nature of the problem with which he will be confronted. It is important to note that this section of the objective does not deal with conditions in which the learning takes place, but conditions under which evaluation is conducted.

If the behavior called for in the objective is to identify each state of the United States, then an appropriate phrase describing the conditions might be: "Given a black outline map of the forty-eight contiguous states, on which only state boundaries are indicated..." But sometimes considerably more ingenuity will be required in devising evaluation situations. If the instructional objective is as subtle as increasing interest in reading poetry, it may be necessary to describe a free-reading situation, where the teacher systematically and unobtrusively notes how many of the students in the class choose to read poetry. Other conditions may be appropriate for the same objective. A situation could be described in which poetry books, along with others, were made available to the students and a log was kept of the books checked out.

While one of the most frequent sources of failure of an objective is inadequate description of the behavior of concern, the difference between an ordinary and an excellent objective is frequently the amount of ingenuity exercised in describing the conditions of performance.

Degree. It seldom happens that the behaviors described in an objective are completely mastered by all students. There are occasions when perfect mastery is required, and in those situations, instruction should be continued until it is attained. Most often, however, we are willing to tolerate the certain variability of performance and may be satisfied with

a reasonably high, but less than perfect degree of attainment. This portion of this statement of objectives establishes the decision point where instruction is deemed to have fulfilled its purpose and a new objective can be pursued.

If the degree of attainment is set too high it may require an extravagant use of instructional resources. It may even render the development of the instructional system of which it is a component economically unfeasible. This is a matter that can be settled only by the careful weighing of a number of values. The experience of the instructor as to what constitutes a realistic expectation of students and the extent to which mastery of this objective is prerequisite to subsequent attainment of other objectives must be considered when the developer of an instructional system prepares this portion of his objective. He should recall that it is he, not his students who is committed to achieve it. Failure cannot be resolved by assigning grades. Once an objective is established, then the developer is committed to insuring that students will attain the objective. Consequently the conditions and the degree of mastery of the objectives may not vary, only those conditions which lead to the objective are free to vary. If the objective is unattainable employing the strategies initially determined, then the entire system must be reworked, unless there is some basis for localizing the difficulty and reworking only that part.


1.2

What are the Sources for Objectives?

Numerous authors deal rather extensively with the form in which instructional objectives should be stated, but few speak to the derivation of objectives. While the form in which an objective is stated is important for communicating the intents of instruction, the validity of the objective is even

ERIC Full Text Provided by ERIC

more important. Behavioral objective statements may readily be written without first determining the nature of the specific requirements of the objective but such a process does not assure the preparation of either valid or relevant objectives. If objectives possess doubtful validity there is little justification for preparing them, and there is little likelihood that they will have an appropriate impact upon instruction.

> Ammerman and Melching (1966, p.4) have outlined the rationale underlying the construction of performance objectives as follows: ...(1) the derivation of job performance requirements must be accomplished prior to the preparation of statements of objectives; (2) the preparation of formal statements of objectives, incorporating the desired performance requirements is necessary for effective communication; (3) the use of these statements of objectives in the design and preparation of instruction, as well as in its management, must occur early to insure that instruction is consistent with the stated objectives.

It is apparent that behavioral objectives for an instructional program do not emerge by a process of spontaneous generation, nor are they entirely the product of creative imagination. Fortunately, there are a number of sources and procedures that the writer of behavioral objectives can utilize that will simplify his task and contribute considerably to the outcome.

It is usually advisable to gather information from several different sources rather than depending too heavily upon one. For example, it will frequently be observed that a professor will verbalize one set of objectives, teach to a second, and test a third. One can only guess which set of objectives is the most valid. Using only one of these sources would most likely result in an unacceptable set of objectives.

The most useful sources of information will depend upon whether the instructional program is intended to replace some existing instruction, or is intended to develop new understanding and skills now presently taught;

whether the objectives are primarily intellectual or manipulative; and whether the writer is looking at his own instructional problem or serving as a specialist for an instructional problem where the subject matter is outside of his areas of confidence.

In all cases, however, an individual investigating sources for objectives must be completely familiar with the value system of the institution for which the objectives are being generated. A concern for valid indicators of content, and intent, must continuously be the primary screen for one's objectives, and as such suggests what may serve as valid sources for objectives.

Task description and analysis. In many respects, this source of behavioral objectives represents a summary process which may be applied to many other sources of objectives. For example, when we observe a group of students in an environment similar to that established for instruction, analysis of learner behaviors, in context, may lead to formal objectives. It is the description of a task, and subsequent analysis of that task, wherein an instructor may find the majority of his objectives. Broadly defined, a task description specifies what criterion respones should be made to what task stimuli and under what range of conditions (Glaser, 1965). Glaser presents a succinct account of the necessity for precise task descriptions and analyses:

> In the learning laboratory, when the psychologist studies development and control of behavior, the task to be learned is carefully analyzed and described. Perhaps one of the indications of lack of interaction between experimental psychology and instructional practice is to be seen in the fact that the educational literature indicates concern with such terms as 'readiness,' 'understanding,' the 'whole child' and so forth. Certainly these are important words and need to be analyzed because the behaviors they refer to are amenable to experimental attack and manipulation only when they are behaviorally defined in stimulus and response terms, i.e., specific subject matter situations and observable student performance.

> This has been a necessity in the work of experimental psychologists in developing laws of learning; increasingly larger and larger units of behavior such as concept formation and problem solving have been studied and analyzed in such terms. In contrast there has been a general tendency among educators to submit student responses to analysis in stimulus-response terms. (Glaser, R. 1962, p. 8)

The basic question that one asks when embarking upon a process of task analysis is: What does a student need to know and be able to do in order to attain the final criterion level established for this instructional unit? In seeking answers to this basic question, task analysis emerges as a systematic method for determining the behavioral requirements of a task. In a psychological sense, task requirements can be stated by tracing through a cycle of task events, specifying the stimuli present and available, and the responses required on the available response network.

Several important points must be considered carefully throughout the task analysis procedure. For example, the vertical and horizontal dimensions of the instructional program, which were established in the philosophical and operational levels of objectives, become the basic units or components from which behavioral objectives are derived. In this derivation process, one's concern must be directed to those kinds of behaviors which a student must manifest after instruction. This is further broken down into behavioral components for some immediate task, or for some task to be accomplished at a future time. The prime concern, and this was emphasized earlier, is that the learner be prepared for something, e.g., for the utility of valued types of behavior.

The concept of utility may be further directed to a resolution of two basic questions which serve to restrict written objectives to elements which are considered important. The questions are:

ERIC

1. What is relevant to the intended performance situation? That is, what are the skills and knowledge that are likely to be useful in anticipated situations?
2. What is critical to instruction? Here one needs to identify the skills and knowledges that are most likely to be needed; those for which instruction in the program is most necessary.

Virtually any analysis procedure which provides valid and reliable answers to these questions will undoubtedly be useful. The value of the analysis, however, will depend entirely upon the validity of its products, on the procedures it employs for gathering and using information, and on the manner in which the procedures are applied. In turn, the manner in which the procedures are applied will be highly dependent on the sources of information used, the types of information gathered from those sources, and the ways in which the information is used to arrive at instructional decisions.

At this stage of deriving terminal behaviors, do not be concerned with the order of presentation, the strategies to be employed, or the methods of evaluation that might be employed. The major purpose in analyzing the learning task at this point is to identify the important behaviors the learner must eventually possess. For example, suppose you were developing an instructional unit on electricity. You might begin by asking what the learner needed to know and be able to do in order to construct an A.C. electrical circuit. Typical behaviors might include: (1) connecting at least 4 loads in a series circuit; (2) drawing the symbol for a resistance unit in a circuit; and (3) constructing a series circuit having two unequal resistors and computing the total resistance of the circuit.

When beginning to develop an instructional system it is desirable to ob-

ERIC

serve a number of individuals performing the desired task, both successfully and unsuccessfully. It is extremely difficult to analyze a complex performance into behavioral components if only successful performances are observed. However, if one can observe a variety of unsuccessful performances as well, the various reasons for their being unsatisfactory may suggest a variety of behavioral components required for successful performance. By thus breaking down one complex behavior into a number of simpler behaviors, the planning and development of an instructional system may be enhanced considerably. Procedures for sequencing these behaviors, identifying appropriate strategies, and methods for evaluating the objective will be detailed in later chapters.

In reading and working through the following sections of this chapter you should note that task analysis is the underlying process for identification of appropriate behavior. As you consider each of the following sources of objectives, you should also consider how the analysis process might be accomplished.

Verbalized Objectives. Perhaps the first step in preparing a set of behavioral objectives is to interview the person, or persons, who understands best what the outcomes of the instructional system should be, or who has the greatest stake in success. Even if one is writing objectives for an instructional system that he will build for himself, it is advisable to begin by expressing the objectives as articulately as possible.

A frequent mistake is to "behavioralize" the objectives too soon. This is particularly true when the writer is interviewing a subject matter specialist who is not accustomed to thinking in behavioral terms. Forcing the interviewee to speak in your terms, rather than trying to understand him on his terms, can easily lead to frustration and loss of cooperation.

An effective technique is one in which you gradually narrow the focus of the objectives until you reach specific behavioral descriptions. The process is one of first becoming familiar with the instructor's value orientations, his philosophical goals, and instructional aims. This view is consistent with that presented earlier where we established three levels of objectives, namely the philosophical or goals level, the operational level, and the behavioral level. The early part of such an interview can supply many valuable clues as to which behavioral objectives will be considered most relevant. If the instructor is forced to supply behavioral descriptions too soon, he may quickly narrow his attention to those outcomes most easily described, and never reveal those outcomes which are of most interest to him.

Once you have determined the "direction" toward which the objectives imply movement, it is time to generate a more explicit description. For example, one may ask: What does success look like? What does a student do that indicates to you that he has really learned what you wanted him to learn? If the student could do this and nothing more, would you be satisfied? What else could he do that would be just as good? How reliably should a student be able to do this? What proportion of the students should be able to do this?

The interviewer should have little difficulty generating meaningful questions of this sort. It is important that you do not confuse the interview with unfamiliar terminology.

It will be helpful to explain why the answers are important, for although answers frequently will come easily they may also come only with great difficulty.

It is usually neither feasible nor advisable to attempt to arrive at final statements of behavioral objectives in the interview, certainly not in the first interview. These can be written independently and later offered for approval. Lengthy discussions about specific wording problems may arouse defensiveness, lead to futile digressions, and obscure more important information. One should not expect too much from the interview or interviewee. Productive interviews are taxing at best, and much of the information required can and should be obtained from other sources.

Instructional materials and activities. Course syllabi and assigned reading materials can provide a considerable amount of raw materials for writing objectives. They specify the elements to be learned but usually reveal little about the way the student is intended to behave when he has learned them. They do, however, give the writer a comprehensive view of the range of objectives that may be required for the system.

It is often possible by observing the activity of an instructor to determine his appraisal of the relative value of different elements of subject matter and to determine what kind of learning behavior he expects from his students. He may for example reveal little concern for detailed factual data but emphasize analytical thinking. The most explicit, if not the most reliable, indicator of the learning outcomes that the instructor considers important is the procedure he uses for evaluation.

Test and evaluation. Since test performances do represent behavioral outcomes, it would seem that they are a natural place to look for behavioral objectives. There is a certain amount of risk involved in this because the nature of most tests is influenced heavily by factors other than instructional objectives. They often measure the most testable rather than the most

ERIC

significant outcomes. Few professors are trained in test construction and few enjoy it. Quite naturally, considerably more energy is devoted in preparation of instructions than to the preparation of tests. It is important to the instructor that a test be easy to construct, administer, and score and that it yield a distribution of scores suitable for the assignment of grades.

While such a test may certainly measure some of the desired learning outcomes it is quite unlikely that it will measure all of them. It is also helpful to examine student products that have been evaluated. Not only do these yield positive and negative instances of desired behaviors but they facilitate the establishment of minimum standards of acceptable performance. The conditions of evaluation that are observed may suggest conditions that should be included in the statements of behavioral objectives.

A different, and somewhat more inclusive, approach to identifying potential sources of objectives has been presented by Tyler (1966). (See Figure 1, page 32.) He suggests three sources of information which may be useful in making decisions about objectives. These are: (1) studies of the learner; (2) studies of contemporary life; (3) suggestions from subject matter specialists. More specifically he suggests that studies of the learner should focus on two basic types of needs. One type is identified as a difference between the present cognitive status of the learner and some acceptable criterion that can be identified. This type of need is cited as a "gap" need. The other sense to which the term "need" refers may be derived from the writings of Prescott, Murray and others. From this position a human being is viewed as a dynamic organism, an energy system, normally in equilibrium between internal forces, produced by the

energy from the oxidation of food, and external conditions. To keep the system in equilibrium it is necessary that certain "needs" be met.

Studies of contemporary life, accomplished at the national, state, and local levels are necessary because of a constantly increasing body of knowledge.

ERIC

Basic Curriculum Questions: What are purposes?
What learning experience should be provided?
How to organize these learning experiences?
How to evaluate these learning experiences?

Curriculum design begins with identification and formulation of objectives:

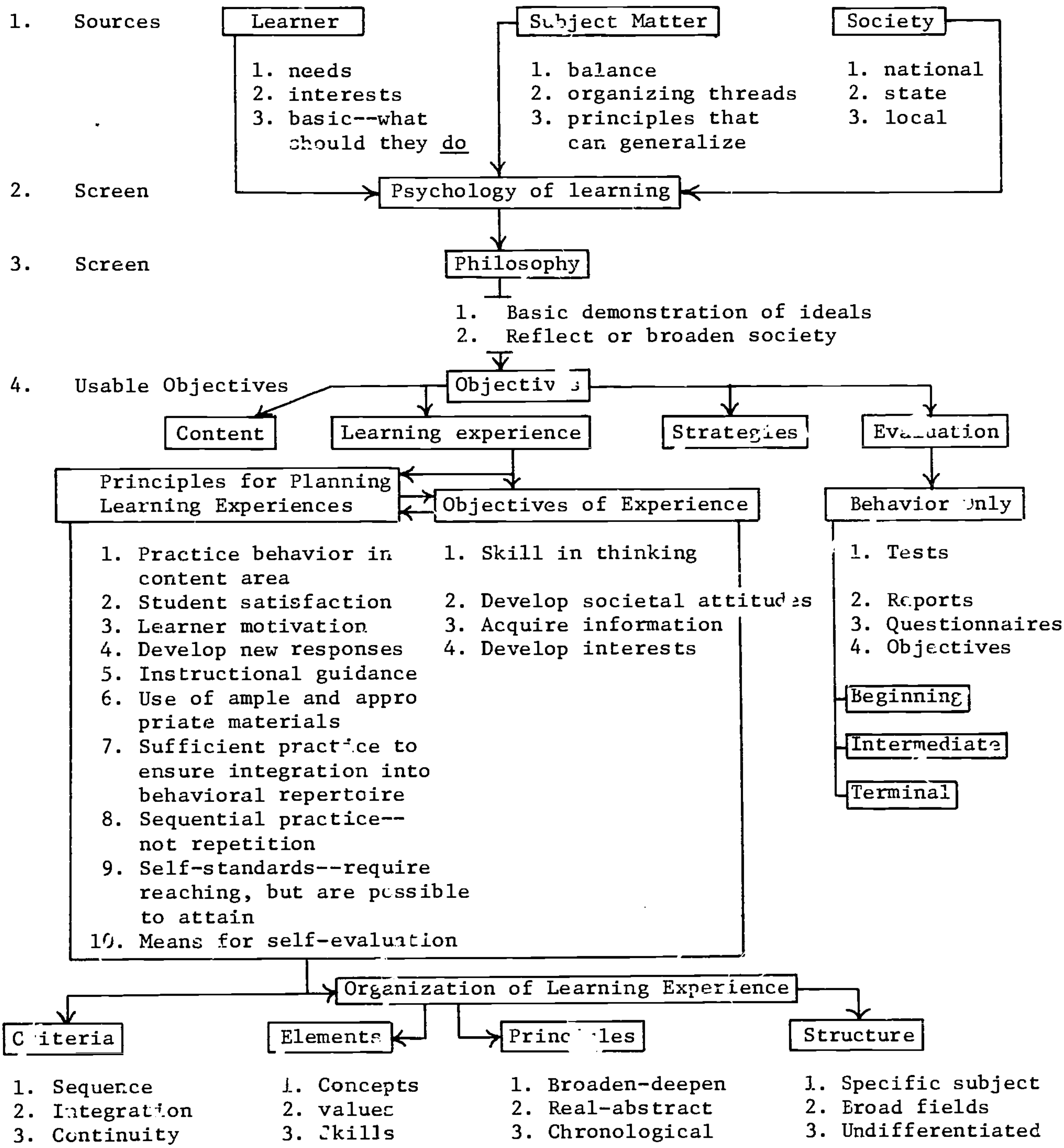


Figure 1. Curricular Development Model (Tyler, 1966)

It is no longer possible for the learner to study all material that has academic respectability. Instead, the schools are continually faced with the question of what knowledge has contemporary significance.

Analysis of contemporary life as a source for educational objectives derives importance from the fact that contemporary life is extremely complex and continually changing. It is imperative that we focus our instructional efforts upon those critical aspects which are of both lasting and immediate importance such that student time is not wasted in learning materials that had significance years ago but is no longer relevant.

Secondly, studies of contemporary life are related directly to the psychological concept of transfer. In these situations, the educational environment should be closely similar to the life situation to which it is directed. The student must be given some practice in seeking and identifying illustrations of his life outside school for the application of knowledge and skill gained in school.

The third potential source for objectives suggested by Tyler is the subject matter or discipline concerned. This is unquestionably the most frequently used source for objectives and, unfortunately, usually reflects only those concepts within the discipline which are deemed important by the academician. Objectives from this source are frequently pseudo-objectives. The importance of the discipline as a potential source of objectives should not be lightly discounted but rather an effective balance must be achieved between the learner, the subject matter and the society in which the education system functions.

It is important for one to recognize that these three questions only outline the broadest of concerns, or possible sources, for objectives. To

be useful, objectives derived under this rationale must be fully operationalized by implementation of the total rationale.


1.3

What are the Functions of Objectives?

Numerous authors present various dimensions to which statements of educational objectives must be addressed. Frequently, within the context of these rationales, curricular development or instructional design begins with the identification and formulation of objectives. Specifically, as in Tyler's model, the design process begins with answers to four basic questions:

(1) What educational purposes should the school seek to implement?

(2) What educational experiences should be provided that are likely to facilitate attainment of these purposes?

(3) How should these educational experiences be organized?

(4) How should these purposes be evaluated?

The most important function of educational objectives, as implied in these questions, becomes that of guiding decisions about the selection of content and learning experiences; and of providing criteria on what to teach, how to teach it, and how to determine when it has been learned.

Because the possible limits of knowledge and learning are unlimited, curriculum makers and teachers are constantly faced with the problem of selection, both of content and of learning activities, as to what is most necessary and most effective. Only a clearly defined and fully articulated collection of objectives can supply the criteria for making these decisions. "No matter what its nature, the statement of desired outcomes sets the scope and the limits for what is to be taught and learned" (Taba. 1966, p.197).

A careful examination of Tyler's Curricular Development Model, as shown earlier in Figure 1, may be extremely beneficial in identifying the primary

ERIC

functions of objectives and establishing the parameters and direction for an instructional program. The model illustrates how sources for objectives might be utilized, the screens one may employ, and the relationship between derivation and screening processes. Our concerns here, however, are with the function of objectives, and Tyler's Model specifies that the function of objectives is to delimit content, to assist in identifying appropriate learning experiences, to assist in identifying appropriate strategies, and to outline appropriate evaluation procedures.

In discussing the function of objectives it might be more useful to categorize the broad notions presented by Tyler into three dimensions of the instructional process. These "3-D's" of instruction may be called design, development and description. Within the design function, objectives serve to identify what content, what emphasis, and what strategies might be most appropriate. The development function refers to the management of a curriculum-building effort. Management in this case is very broadly defined and includes such things as facilitating decisions about content, sequence, pacing, utilization of evaluative information for refinement and clarification of the curriculum, etc. The descriptive function of objectives may perhaps be most clearly understood if we look at this dimension as guiding the evaluation of the total instructional effort. With respect to each of these dimensions, the statement and the use of objectives is of crucial importance.

Since the basic constructs underlying the use of behavioral objectives are not entirely consistent with traditional curricular development models, thinking in terms of "instructional systems" may be of assistance in overcoming traditional notions about objectives and their use in the instructional process.

The mechanistic connotation frequently associated with instructional systems is due to the degree of instructional precision attainable through their use, not to the foresaking of one's total value system. The planning of an instructional system does not signal the abandonment of philosophical goals, educational ideals, or aesthetic values. Instructional systems are every bit as sensitive to these important dimensions of the educational enterprise as are the traditional forms of curriculum building. In curriculum development or systems planning, however, they serve only as a very good place to start. Conversely, they represent a very poor place to stop in the planning effort. The success we have in systems planning will depend largely on the clarity and explicitness of expression of these value dimensions of the educational enterprise.

The more clearly they are stated, the more dearly they are held, the more urgent it becomes to express our values in a form that will allow us to plan effectively for their achievement. As discussed earlier, these types of statements provide us with a general direction, but not with adequate information to build an instructional system. It may be useful to know that we are heading east, but certainly we would want to know where we were going in that easterly direction.

Just what design functions should good objectives serve? In addition to giving directions to the instructional task, they should also delimit the task. They should describe something feasible of attainment with the resources at hand. A general goal such as appreciation of our cultural heritage not only requires transformation to more observable terms but also considerable delimitation if it is to be feasible of attainment. For example, few people would agree that describing the historical origins of the Bill of Rights would adequately develop an appreciation of our

ERIC Full Text Provided by ERIC

cultural heritage in students. It does, however, represent a legitimate objective that is highly relevant to the more general value statement.

Delimiting objectives allows the developer of an instructional system to effect closure. It is rather unlikely that one could ever complete the task of fully developing an appreciation for our cultural heritage. But the point of attainment of a behavioral objective is readily identifiable. Therefore, if the behavioral objective is to serve the purpose of providing closure and identifying completion of the instructional task, it must describe behaviors that will be observable within the spatial and temporal limitations of the instruction. Furthermore, its attainment should not depend upon external, uncontrollable events.

In addition to giving direction, focus and reasonable limits to the design of an instructional system, objectives should be stated in a manner that facilitates the planning of appropriate instructional procedures to achieve them. We must have at our disposal the means for attaining an objective, or else it is no more than a wish. Objectives of this type (wishful thinking) will normally prescribe an instructional activity that has a function highly congruent with the functions of a rain dance. It makes us feel better in that it relieves some anxieties, but it contributes very little to the total instructional process.

These concerns all emphasize again the importance of expressing objectives in behavioral terms. There is no scientific body of knowledge that speaks to the problem of developing hypothetical constructs, but educators and psychologists do know something about changing human behavior. While it may be fashionable to deplore the dearth of scientific principles of instruction available, the difficulty perhaps lies not so much in our

deficiency of knowledge as in our deficiency in stating instructional problems appropriately. Behavioral technology may not yet have all the answers, but this does not justify disregarding that which we do know.

Concurrent with the planning of the operation of an instructional system is the development of a test to determine whether the system achieves the stated objectives. This test differs from those used in conventional instruction in an important way; it is used to evaluate the instruction, not the student. Tyler (1966, pp.68-69) emphasized this point when he stated:

> Evaluation then becomes a process for finding out how far the learning experiences, as developed and organized, are actually producing the desired results, and the process of evaluation will involve identifying the strengths and weaknesses of the plans. This helps to check the validity of the basic hypothesis on which the instructional program has been organized and developed; it also checks the effectiveness of the particular instruments, that is, the teachers and other conditions that are being used to carry forward the instructional program. As a result of evaluation it is possible to note in what respects the curriculum is effective and in what respects it needs improvement.

This concept of evaluation outlines an approach that is essentially a process of determining to what extent the objectives specified are being realized by the instructional program. The objective form proposed in this chapter is designed to specify the types of change in learner behavior that we value or seek to attain. Consequently the evaluation function served by objectives is largely one of enabling us to determine the degree to which the desired changes in behavior are actually taking place. Thus, while the administration and grading of a final examination usually signals the fulfillment of an instructor's responsibility to his students and the end of his task, the administration of tests in this context usually signals not the end of the developmental task, but merely the inception of a new and crucial phase.

It is rare indeed that even the most carefully designed instructional system will function perfectly on the first trial. Even though well-established learning principles may have been meticulously applied in the design stage, their implementation will usually have unexpected or unwanted results. The first draft of an instructional system is a kind of complex hypothesis which is to be tested. The appropriate test of a system is not whether it was presented as planned, but whether it effected the desired changes in learner behavior.

The function of behavioral objectives in this case is to provide a sensitive indicator of the performance of the system, and to inform the developer of it strengths and weaknesses. Effective objectives enable the establishment of an efficient information feedback system that allows timely correction of the deficiencies of that system. It should be noted that this empirical process of developing instructional systems makes the naive student a unique and valuable resource person as a part of the developmental team. Frequently instructional planners know too much, and they have forgotten how difficult it is to learn. The student by his learning performance, can show the instructor where too much previous knowledge has been assumed, where the instructional pace is too rapid or too slow, and where the desired learning is or is not taking place.

Again, it is the system, not the student, that is being tested. It is fairly well known that good students can learn in spite of poor instruction. What is not quite so generally appreciated, however, is that effective instruction can overcome many learner deficiencies.

When the development of the instructional system has been completed and its effectiveness demonstrated, it becomes important to describe it in

terms of its most relative characteristics so that it may be used by others. If this is not done, the system constitutes an "instructional grab bag."

This description should be intellectually honest. It requires that the system be described in terms of instructional objectives actually attained and observed in a realistic trial. It is neither appropriate nor honest to describe "intended" outcomes as attributes of the system. While intentions may be attributes of the design, only the observed achievement of these intentions is appropriately regarded as an attribute of the system.

Intellectual honesty is further served when we make clear our educational goals with explicit statements or operational definitions of just what we consider to be adequate evidence of the achievement of those goals. It is not sufficient to say that our students have developed an understanding of democratic policies, even if our observations have convinced us that this is true. While most people would agree that such an objective is important there may be wide disagreement about what behavior constitutes achievement of the objective. For example, one instructor may consider that he has achieved the objective when his students have memorized the Constitution. Others, however, may disagree vehemently. The most honest and least deceptive approach is to describe the instructional system in terms of the observed attainment of behavioral objectives. If a potential user considers these objectives and the stated terminal behavior to be inappropriate, then he may disregard the system. At least he has not been misled by a deceptive label. It is up to each individual instructor to determine the relevance of the instructional objectives to the learner group with whom he is working, the context within which he is working, and his instructional value system. No one else is in a position to make this decision for him.

ERIC

No matter how effective an instructional system may be, its ultimate impact on education is determined by how generally, and how appropriately, it is used. Ideally, both instructional tools and instructional problems would be described in such a manner that they could be matched one with another.

Unfortunately, this is not generally the case. Tools of instruction, for example, whether books, filmstrips, or motion pictures, are usually described and evaluated in terms of their physical properties. Catalogs list the number of frames in a filmstrip, the length of a motion picture and whether it is black and white or color, and so forth. Yet the ultimate purpose of these tools is that they effect some change in the students that are exposed to them. It would seem that a far more appropriate criterion for selection and far more useful means of description would be the behavioral rather than the physical properties of these tools. The prudent car buyer will assess the performance characteristics of a car by taking it out on a trial drive. He will observe how it handles, the smoothness of the ride, and the quietness of the motor. The selection of instructional tools merits no less caution.

Implicit within the functions outlined is a central concern to which the reader's attention must be drawn. First, it goes without saying, the focus of any educational system is to insure that as learners leave its instructional control, they are capable of exhibiting certain desired skills and knowledges related to anticipated work, further school, or life situation. The larger institution assumes the responsibility for the orientation or direction of the instructional program through the philosophic base it has described. But, as Cronbach (in Glaser, 1965) has noted,

ERIC

"success of the behavior of the pupil is modified so that it meets certain specifications."

The rationale for behavioral or performance objectives presented thus far has emphasized that objectives become the indicators of the specific intents of the instructional program. They outline the conditions under which the learning is to occur, the criterion to be used in establishing the minimum acceptable performance, and the level at which the behavior is to be accomplished. These three factors constitute the specifications for a behavioral objectivae.

A specification, in the sense that it is used here, is a statement of essential conditions that must be met. Other conditions may be varied, but not these. Ammerman and Melching, in support of this position have noted: "The feasibility of instruction should be made only <u>after</u> decisions have been made with regard to relevancy and criticalness." (1966, 30)

In making instructional specifications, it is important that a differentiation be made between important conditions that must be fulfilled, and less important factors that may be permitted to vary. Unfortunately, throughout the educational enterprise, there is a strong tendency to choose as specifications those factors about which it is most easy to be specific. Consequently the most visible components of the instructional process such as the length of time to be spent on a unit of instruction, or the number of pages to be read, or the number of problems to be solved, tend to become specifications, and the resulting learning effects are permitted to vary.

Despite what is predominantly evident in current educational practice, the means are not more important than the end. If education's primary commitment is to helping students to achieve, rather than to present instruc-

ERIC

tional operations, then it must hold to the achievement of learning outcomes and vary the instructional procedures until these specifications are met, rather than holding fast to our mode of instruction and accepting whatever achievement we observe. To this end the objectives established serve an especially important function.

1.4
1.5
1.6
1.7

Objectives and Evaluation

There has been a tacit inference to the close association between clear, precise statements of objectives and the evaluation function to be performed in the educational setting. Due to the wide discrepancy that frequently exists between what is taught and what is evaluated, the evaluative function of objectives needs to be made more explicit.

To a large extent, this discrepancy between the desired outcomes and evaluation of these outcomes, is caused by objectives not being clearly formulated. The often referred to intangibility of some objectives is nothing but a smoke screen for a lack of clarity in the thinking of those stating them. Objectives of this type, those which are claimed to represent intangible behaviors, describe neither the behavior not the content clearly enough to make adequate analysis possible. Therefore, evaluation of those behaviors tend to be concentrated on the most obvious but not always the most important outcomes, such as remembering information rather than thinking with it.

It is well known that those things which are most clearly evaluated are also most effectively taught. Since intended changes do not become effective changes unless there are clear cut provisions for them in the instructional process, then the instruction is also narrowed and weakened in its effect when some important outcomes are unclearly defined. The process of evaluation

ERIC Full Text Provided by ERIC

therefore, must begin with a clear statement of the objectives of the educational program. Until there is a clear conception of the behavioral intent of the objective, there is no way of determining what kind of behavior to look for in the students to assess the degree these objectives are being realized.

The following points made by Taba (1962, p.312) are most appropriate in cementing this direct relationship between objectives and evaluation; education is a process which seeks to change the behavior of students. These changes are the objectives of education. Evaluation is the process of determining what these changes are and of appraising them against the values represented in objectives to find out how far the objectives of education are being achieved.

The reader is referred to the chapter by Paulson for a further discussion of this particular function of objectives.

Writing Behavioral Objectives

It requires more than a little skill to construct objectives in such a way that they will provide the instructor with information required to make instructional decisions effectively. The formal statement of objectives constructed, no matter how adequate, is only a part of the total platform of objectives. It would be presumptuous to attempt to incorporate all of the necessary qualifications and specifications in such a statement. This would lead only to a cumbersome complexity which would communicate little to anyone except those who formulated it. The important consideration is to generate statements that will facilitate structuring the environment such that learners will respond in a given manner. There is really not much difficulty in getting students to behave in an overt and observable manner, as most teachers will testify. The task is more one of structuring situations so that behaviors

of interest can be observed, and then determining whether such behaviors indicate the attainment of the instructional objectives.

Typical shortcomings in objectives include those which are stated as instructor activities, for example, to introduce the concept of signed numbers. This type of objective is behavioral in the sense that it describes what the instructor will be doing, but it does not provide any information about desired instructional outcomes. The purpose of education is not to have the instructor perform certain activities, but to bring about significant changes in students' patterns of behavior. It is imperative to recognize that any statement of objectives must be statements of changes to take place in the students' behavior. Another major problem with objectives stated as teacher behavior is that there is no way of judging whether these activities should really be carried on. They are not the ultimate purposes of the educational program and are not, therefore, the true objectives of that program.

In as much as most clearly formulated objectives should possess both behavioral and content dimensions, it may be helpful to employ a two dimensional matrix to express one's objectives clearly and concisely, concomitantly insuring incorporation of all important aspects of learner behavior and content in your planning. An example of such a matrix is presented in Figure 2.

This matrix has been developed for a biological science course, and includes seven types of desired learner behavior. In planning and formulating objectives for instruction, the intersections of the behavioral columns and content rows are marked when the behavioral dimension applies to a particular area of content. For example, in this case, the student is expected to develop an understanding of important facts and principles in connection with

Content Aspect of the Objectives	Behavioral Aspect of the Objective: 1. Understanding of important facts and principles.	2. Familiarity with dependable sources of information.	3. Ability to interpret data.	4. Ability to apply principles.	5. Ability to study and report results of studies.	6. Broad and mature interests.	7. Social attitudes
A. Functions of Human Organisms							
1. Nutrition	X	X	X	X	X	X	X
2. Digestion	X		X	X	X	X	
3. Circulation	X		X	X	X	X	
4. Respiration	X		X	X	X	X	
5. Reproduction	X	X	X	X	X	X	X
B. Use of Plant and Animal Resources							
1. Energy relationships	X		X	X	X	X	X
2. Environmental factors conditioning plant and animal growth	X	X	X	X	X	X	X
3. Heredity and genetics	X	X	X	X	X	X	X
4. Land utilization	X	X	X	X	X	X	X
C. Evolution and Development	X	X	X	X	X	X	X

Figure 2. Illustration of the use of two-dimensional chart in stating objectives for a high school course in biological science (Tyler, 1966. p. 32).

ERIC Full Text Provided by ERIC

every one of the content aspects. On the other hand, he is expected to develop familiarity with dependable sources of information only in connection with nutrition, growth, heredity and genetics, land utilization, and evolution, and development. While a chart of this nature does not detail all of the important components of a behavioral objective, it does facilitate the analysis and planning process. From a chart such as this, the instructional planner may readily identify the behavioral and content requirement for objectives.

It should be emphasized again that this matrix does not identify particular objectives but rather is a device for determining whether an objective is appropriate and insuring complete behavioral and content description of the desired outcomes. The chart serves only as an aid in formulating objectives such that their meaning may be more clearly stated. The chart may also serve to identify possible gaps or omissions in the platform of desired objectives.

A statement of instructional objectives should be both comprehensive and explicit but it is not necessary to repeat compulsively elements that a number of objectives have in common. For example, if all of the objectives are appropriate for a given population of students, the audience need be described only once in a prefatory statement. Objectives to be measured under similar conditions can be grouped together, and the conditions described but once. In some instances, it will not be feasible to measure all intended behavior for all students. For example, if one hundred definitions are to be learned, a sample of these definitions can be tested and the proportion of that sample that constitutes an acceptable degree of achievement can be specified.

ERIC Full Text Provided by ERIC

By way of summary, the following points must continuously be considered whenever writing instructional objectives:

1. Be certain that you know what you wish to accomplish with the population of concern. The referent here is to a clear and concise statement of the philosophic orientation and value system which you hold.

2. State operational objectives in a manner that will facilitate derivation of behavioral objectives.

3. Behaviors of intent must be clearly observable and measurable. For example, you can define an objective with sufficient clarity if you can describe or illustrate the type of behavior which the student is expected to acquire such that various observers could agree, on observation, whether or not it occurred.

4. It is important that the behaviors desired be appropriate and consistent with the type of learning and specified content. For example, it is neither appropriate nor consistent with the type of learning to have a learner paint pictures in order to learn to read musical notes.

5. From a design standpoint, your objectives should provide for a careful definition of the entry level of the audience for whom this specific instruction is intended. As mentioned earlier however, you need define the audience only at the beginning of a series of objectives.

6. Other things being equal, more general objectives are desirable rather than less general objectives. To identify appropriate learning experiences, however, it will be helpful to differentiate clearly between types of behavior which are different in their characteristics.

7. The conditions under which you expect the behavior to occur must be clearly identified. This information is essential as you begin to outline learning strategies and evaluation procedures.

8. The criterion you select must relate to the behavior in a relevant way, for example, a time criterion is not relevant to a power objective, such as"...explain the lawmaking processes of the United States, including..." One must also exercise some caution in applying a percentage criterion. For example, one could appropriately expect students to achieve 100 percent correct responses in demonstrating ability to complete multiplication tables but probably something less than 100 percent would be more appropriate in solving more complex mathematical problems, such as story problems. The important concern is to select a criterion level that will facilitate production of success at the level desired.

9. Be sure objectives are written in a future tense, so as to indicate that the behavior specified will be manifest after some instructional event. In other words, be careful in your use of the initial verbs, such as will or should. "Will" implies some behavior after an instructional event, while "should" implies behavior already possessed by the learner. This notion is especially clear when one considers possible entry behavior. For example, in the establishment of an entry level, one might find a statement such as this: "the student should be able to identify the correct use of the following symbols which denote multiplication." In this case, the statement describes prerequisite skills for the learner to enter a given instructional system. The statement is not an objective. The above statement rephrased as an objective might be stated as follows: "All fourth grade students will be able to identify the correct use of the following symbols which denote multiplication."

On Writing Behavioral Objectives.

The intent of this chapter has been to introduce the reader to behavioral or performance objectives. Instruction was implicitly differentiated from teaching, which is a generic term, i.e., it covers numerous functions. It is defined as the process whereby the environment of an individual is deliberately manipulated to enable him to learn to emit, or engage in, specified behaviors under specified conditions or as responses to specified stimuli.

Important considerations in this differentiation were the degree of specifity of the behavior to be learned as well as the conditions for that behavior to occur. Traditionally, the degree of control exercised over the environment of the learner lacks sufficient provision to bring desired behaviors under the control of appropriate stimuli. Unfortunately, for those of us concerned with the design of instruction, the learner can refuse to respond to the instructional stimuli or environmental manipulations, should he so desire. Consequently, in an instructional setting where the learner holds the options of responding, or not responding, an added burden is placed on the instructional program.

ERIC

The basic purpose of instruction, then, is to bring certain learner behaviors under the control of the instructional situation, or to insure that certain behaviors of the learner will be spontaneously emitted under certain conditions. The implication is that all forms of behavior can be specified at some level, and if we cannot specify what the behavior is to be, then we cannot design an instructional program to achieve it reliably.

One might question this notion if he is concerned with statements of attitudinal objectives, such as "derives ideas about the conduct of life," or "begins to develop dominant values." If, in these cases, you will consider your objectives as possessing both behavioral and content components formulation of the objectives will be somewhat easier. For example, by preceding each of the above objectives with the statement "uses reading," then specific content is included, and the instructor has a more definitive notion about where and how to proceed in facilitating attainment of this objective.

In every case, the process of instruction begins with "givens," i.e., the behaviors which the learner brings to the situation and those which the instructional program intends to elicit and imprint. To the degree that the consequences of instruction are defined and observable, the process can continue and become increasingly effective.

Effectiveness in writing behavioral objectives will be achieved only through continued experience. Whatever has gone before is meaningful only if it leads to such experience. The procedures and rationale presented here are worthy of testing, but not blind acceptance.

1.8

References

Ammerman, Harry L., & Melching, William H. *The derivation, analysis, and classification of instructional objectives*. Washington, D.C.: The George Washington University Human Resources Research Office, 1966.

This publication has become a classic to those who write about behavioral objectives. The authors deal with current practices and problems, the use of objectives in instruction, and guidelines for evaluating derivation methods. Perhaps of most use is the analysis and classification system which they develop.

Bloom, Benjamin S. (Ed.) *Taxonomy of educational objectives, Handbook I: cognitive domain*. New York: David McKay Company, Inc., 1956.

In this publication, Bloom et al., develops a taxonomy or hierarchy of objectives within the cognitive domain. The six broad classes of objectives outlined are (1) knowledge, (2) comprehension, (3) application, (4) analysis, (5) synthesis, and (6) evaluation.

Corey, Stephen M. The nature of instruction. *NSSE Yearbook LXVI, 1967*. Chicago: University of Chicago Press.

Gagne, Robert M. *Psychological principles in system development*. New York: Holt, Rinehart, & Winston, 1962.

This is another edited work with the chapters by Miller, Crawford, Glaser, and Klaus being especially useful in outlining objective derivation procedures. Miller's chapter is devoted to task description and analysis, Crawford's to concepts of training, and Glaser and Klaus' to proficiency measurement.

Glaser, Robert (Ed.) *Training research in education*. New York: John Wiley & Sons, Inc., 1965.

This is an edited work with the chapters by Glaser, Miller and Carroll being most useful.

Glaser, Robert (Ed.) Psychology and instructional technology. *Training Research In Education*. Pittsburgh, University of Pittsburgh Press, 1962. (p.8)

Good, Charter V. *Dictionary of education*. New York: McGraw-Hill, 1959.

Gronlund, Norman E. _Readings in measurement and evaluation_. New York: The MacMillan Company, 1968.

This is an edited work with the chapters by Dressal and Krathwohl being most useful. Dressal's article is addressed to measurement and evaluation of instructional objectives, and Krathwohl's chapter is devoted to use of objectives in the cognitive and affective domains.

Krathwohl, David R. _Defining educational objectives, the taxonomy of educational objectives-- its use in curriculum building_. Pittsburgh: University of Pittsburgh Press, 1964.

In this article, Krathwohl outlines the rationale for several levels of objectives and then briefly details how objectives might be used at each of these levels.

Krathwohl, David R., Bloom, Benjamin R., & Masia, Bertram D. _Taxonomy of educational objectives, Handbook II: affective domain_. New York: David McKay Company, Inc., 1964.

In this publication the author has developed a taxonomy or hierarchy of educational objectives in the affective domain. The five levels of objectives established include (1) receiving (attending) (2) responding (3) valuing, (4) organization, and (5) characterization by a value or value complex.

Lindvall, C. M. (Ed.) _Defining educational objectives, a report of the regional commission on educational coordination and the learning research and development center_. Pittsburgh: University of Pittsburgh Press, 1964.

This useful work focuses on the derivation and utilization of objectives in instructional design. The chapters by Gagne and Glaser and Reynolds are especially useful in associating objectives with the learning processes and in specific derivation procedures respectively.

Mager, Robert F. _Preparing Objectives for Programmed Instruction_. San Franciso: Fearon Publishers, 1962.

Smith, Robert G., Jr. _The design of instructional systems_. Washington, D.C.: The George Washington University Human Resources Research Office, 1966.

This publication, in a very succinct way, relates training procedures and behavioral objectives. The author speaks directly to the use of objectives in the design of a total instructional system, especially systems which are devoted to training performance.

Smith, Robert G., Jr. The development of training objectives. Washington, D.C.: The George Washington University Human Resources Research Office, 1964.

The publication can be of most use in the objective derivation process. Chapter 4 is devoted primarily to deciding which tasks to teach and includes such topics as rationale for deciding what to teach, the integration of rationales, levels of training, etc.

Taba, Hilda. Curriculum development, theory and practice. New York: Harcourt, Brace & World, Inc., 1962.

This book outlines Taba's theory of curriculum development. Her notions have been widely accepted in education, especially in the social studies fields.

Tyler, Ralph W. Basic principles of curriculum and instruction. Chicago: University of Illinois Press, 1950.

This entire book details Tyler's concept of instructional design. He is one of the pioneers in stating objectives behaviorally and many of his ideas are deeply rooted in all current writings about behavioral objectives.

ERIC

Section II: Designing Instructional Systems

Editorial Foreword

This chapter by Paul Twelker continues the emphasis placed upon clear and measurable learner objectives set by the preceding chapter. Whereas the previous chapter focused more on the problem of discovering and defining objectives, the present chapter turns to the problem of building a sequence of objectives. Some objectives in the sequence are simply statements of the required en route or facilitating skills. When this sequence of objectives, both ends and means, is specified, usually termed "objective analysis," it presents a coherent description of where the instruction is to take the learner and what major steps are required.

The chapter is comprised of two sections:

1. The first section grapples with the problem of sequencing terminal, or end point objectives, and enabling, or en route objectives. Several techniques of educing networks of such objectives are explained. Twelker emphasizes that any such hierarchy of instructional objectives rests upon the available knowledge of the entering learner. The relevant skills the learner already possesses form an anchoring pier at one end of the sequence, the terminal objectives form a pier at the other end.

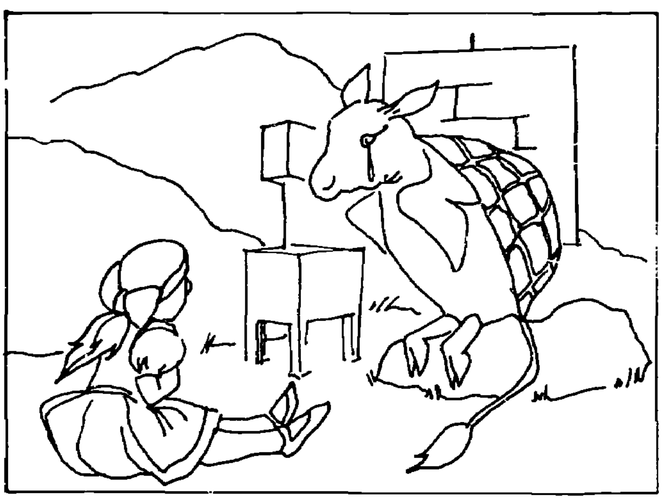
2. The second, and lengthier, section of the chapter is a description of aspects of instruction that need to be considered to construct the network of objectives identified in the first section. These aspects of instruction include: characteristics of the learner, content and form of presentations, classes of learning involved, kind of feedback to the learner, and the setting or contextual factors of instruction. This section could be viewed

as a thirteen-step, exhaustive check list. Unfortunately, the reader will find no readily available knowledge to cope with some of the thirteen steps. Twelker attempts to present some useful working strategies for those steps in which empirical evidence seems lacking.

For all its length, the chapter is a relatively straightforward presentation. To the editor, the author reacts over-sensitively to perceived limitations of usable and useful knowledge for the design of instruction. The editor feels like the aged farmer who said he already knew more about farming than he could apply now.

Readers in a trial audience found it helpful to make their own check list as they perused this chapter, often by simply paraphrasing the thirteen major steps discussed. In fact, for initial reading, such a list could be coupled with scanning a few descriptive sentences under each step. This may prevent the noviate reader from being overcome by sheer volume.

One refrain throughout this chapter haunts and torments the editor. It goes, "When in doubt or difficulty, consult with a competent instructional technologist." Sound advice maybe, as soon as someone discovers or creates a supply of competent, and communicative, instructional technologists.



To design lessons that lessen, an Instructional Tortoise would specify:

Reeling and Writhing Objectives,
Conditions of Derision and Uglification,
Laughing and Grief Settings...

Designing Instructional Systems

Paul A. Twelker

Overview

When we talk about the design of instructional systems, we are thinking in terms of the tasks involved in specifying in a systematic fashion a series of learning experiences that will produce consistently and predictably a desired or stated behavior on the part of the learner when implemented. A parallel may be drawn between an architect and an instructional system designer. The architect specifies guidelines and plans for each step of the construction of a building--the end product is a set of blueprints that when translated by a contractor result in a building. The instructional system designer is in one sense an architect. He specifies various components of instruction: media, content, instructional strategies, and so forth. The result is a blueprint for instruction--something that an instructional system developer might use to build a prototype from the specifications. Needless to say, the specifications should be in sufficient detail so that a person other than the designer could take them and develop the instructional system.

How does one go about the complex job of designing an instructional system? Basically, two major steps are involved. Given a clear statement of terminal objectives, the designer must determine the sequence of instruction. That is, he must determine what enabling objectives or <u>en route</u> objectives are required and in what order they should be taught. Then he must specify the instructional conditions that best "fit" the objectives.

In order to orient the reader as to what follows, the major portions of this chapter are listed here.

Specifying Instructional Sequence - an examination of a method called objective analysis that lets a designer determine what enabling objectives must be covered prerequisite to the terminal objective, and in what order these should be taught.

Specifying Instructional Conditions - an examination of a series of guidelines that helps the designer think about this task.

Relationship Between Research and the Instructional Systems Approach - a brief look at how the sytems approach relates to the planning and conduct of basic research.

Package Summary.

Specifying Instructional Sequence

> One of the major contributions coming from the work of the training psychologists has been the explication of the basic notion that a learning objective or target outcome can be performed only to the extent that all of the skills and/or knowledges subordinate to it are also in the repertoire of the learner. This requires that in order to guarantee that a target outcome will, in fact, be realized there must be a careful hierarchical analysis of the skills and/or knowledges prerequisite to it and the development of effective instructional systems to bring them about. (Schalock, 1968)

The sequencing arrangement of instructional experiences in accordance with the order in which competencies[1] should be learned is an important task indeed. Briggs (1968, p.114) suggests that the matter of sequencing of units of instruction is a more powerful influence determining criterion performance than other variables such as number and type of examples that are concerned

with how competencies are taught. He states that "in a hierarchically structured course, if the units are arranged in the wrong order it may not matter how skillfully the instruction is programmed in the frames comprising the unit." This is based on the assumption that the various competencies a learner must acquire during instruction are dependent one upon the other. That is, the learning of one competency transfers to another, thus facilitating the learning of the other competency. (Often, the term "enabling objective" is used to describe these prerequisite competencies.) If this assumption is true, then the competencies must be sequenced so that transfer is optimized.

The step of specifying the sequence of instruction is actually twofold. First, the instructor must determine what is to be sequenced and then secondly, he must determine in what order a sequence should occur. In the first case, he asks the question, "What competencies are required of the learner that are prerequisite to his satisfactorily completing the terminal objective?" In the second case, the question the instructor asks is "How do I arrange the units of instruction so that the learner achieves the terminal objective?" Fortunately for the instructor, one procedure that he may follow helps answer both questions. This procedure is based upon techniques used by Gagne (1962; 1965) and has often been referred to as

---

[1] A competency in simplest terms refers to a skill or a class of skills. However, as used in this paper, the statement of a competency includes some sort of statement of adequacy or quality of response. Competence may be defined as the possession of sufficient quantity or adequacy. One is competent when he is sufficiently prepared for the purpose. To simply state a skill without some indication of sufficiency is short-changing the term.

ERIC Full Text Provided by ERIC

hierarchical analysis or learning set analysis. In this report, it is referred to simply as objective analysis.

A primary purpose of objective analysis is to discover what competencies that a learner is to acquire during a given course of instruction are independent from one another and what competencies are dependent upon one another. If competencies are independent from one another, the learning of one competency should not facilitate the learning of another, and the different competencies could be taught in any sequence. On the other hand, if the competencies are dependent upon each other, the learning of one competency may facilitate the learning of the other. In fact, Gagne suggests that the learning of a competency may not only be facilitating for another, but may be mandatory for the learning of another competency. If competencies are dependent one upon another for learning, a careful sequencing of competencies is imperative if positive transfer is to be achieved.

In the case where competencies are independent one from another, objective analysis does not help the instructional engineer in determining what competencies should be taught. On the other hand, when objectives are dependent one upon another, objective analysis may be a useful tool in determining what should be taught to reach a given terminal objective. The procedure by which this is done is simple. Starting at any competency, usually a terminal objective, the following question is asked: "What kind of capability would an individual have to possess if he were able to perform this competency successfully, were we to give him only instructions to do so?" The phrase regarding instructions needs some clarification. The

learner should be:

1) told the form of the answer (e.g., numerical, symbolic, or enacted);
2) informed of any definitions that would clarify the meaning of the competency; and
3) provided with guidance suggesting the application of previously acquired information to the objective under consideration.

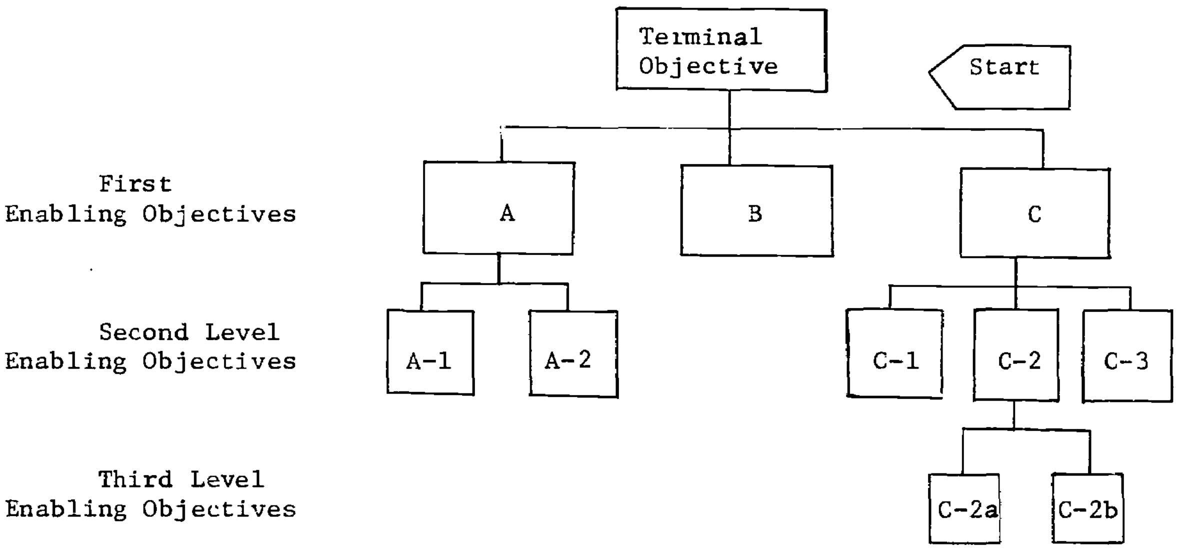


Figure 1. A hypothetical hierarchy of enabling objectives for a single terminal objective.

How does an objective analysis help the designer determine what competencies should be taught? In the case of the hypothetical example in Figure 1, the answer to the question, "What would the individual have to know?" it turns out, identifies a small list of major items of knowledge or skill. These are represented by A, B and C on the figure. When

the same question is asked for each of these items in turn, another list is identified for each, and so on until each item of knowledge and skill derived from the objective analysis has been added to the structure. Note that the enabling objectives thus gained are not the same competency as the final task from which they were derived. They are in some sense simpler and in some sense more general. In the case of the hypothetical example, by using the procedure of objective analysis, we find out that what we are defining is a hierarchy or suborder of competencies growing increasingly simple. In such a structure, options exist within layers as to sequence of acquisition, but acquisition of one layer in its entirety is called for before the presentation of the next higher or more complex layer. For example, A-1 might be taught before or after A-2 is taught, but each of these must be taught before A is taught. Through this procedure, extraneous competencies not required for the completion of the terminal objective are not included in the hierarchy. In answer to the question, "What would the individual have to know?" extraneous or irrelevant competencies would quickly be tested against the terminal objective and discarded. They would be found to be unrelated to the individual's adequate performance on the terminal objective.

Three things should be noted. First, the terminal objective in the hypothetical example, as well as in other examples with which you may be dealing, may or may not be actual terminal objectives of the instructional unit. The designer may start at any point in the to-be-developed hierarchy when he asks the question, "What kind of capability would an individual have to possess?" Building a hierarchy is not usually a simple matter of proceeding from a stated terminal objective to the simplest competency required.

Rather, the instructional engineer must often "mess around" in the hierarchy, slowly building it piece by piece, and unit by unit.

Second, some courses involve instruction where the structure of content does not assume a hierarchical form. Sometimes it is flat, sometimes vertical, and sometimes mixed. Different types of course structures are discussed in detail by Briggs (1968, pp 11-13). A number of examples that illustrate typical course structures are presented in Tables 1-9. Examine these tables carefully and trace the competencies that are dependent one upon another for learning. For example, take the hierarchy of competencies for the topic in elementary non-metric geometry as shown in Table 8. The topic consists of knowledge which a student can use to specify sets, intersections of sets, and separations of sets. Note that at the lowest level (V and VI), the concepts of separation of <u>entities into groups</u>, <u>point and set of points</u> must be known or acquired before the learner is ready to attain the higher order enabling objectives. Take a few minutes to study this hierarchy to prove to yourself that the lower order objectives must be obtained prior to the higher order objectives. If you are not too mathematically inclined, you may wish to examine alternate hierarchies, such as those given in Tables 1-7.

Third, where does the designer stop in objective analysis? Needless to say, he could go on to extremely simple behaviors such as "be able to write words and sentences" when the objective involves identifying procedures, for example. The answer is obvious. Objective analysis is carried out until the competencies determined reach the entry level of the learner. In designing an instructional system, it is useful to think of a gap between the learners initial capability and his capability after instruction. The objective analysis is limited by these constraining "end points."

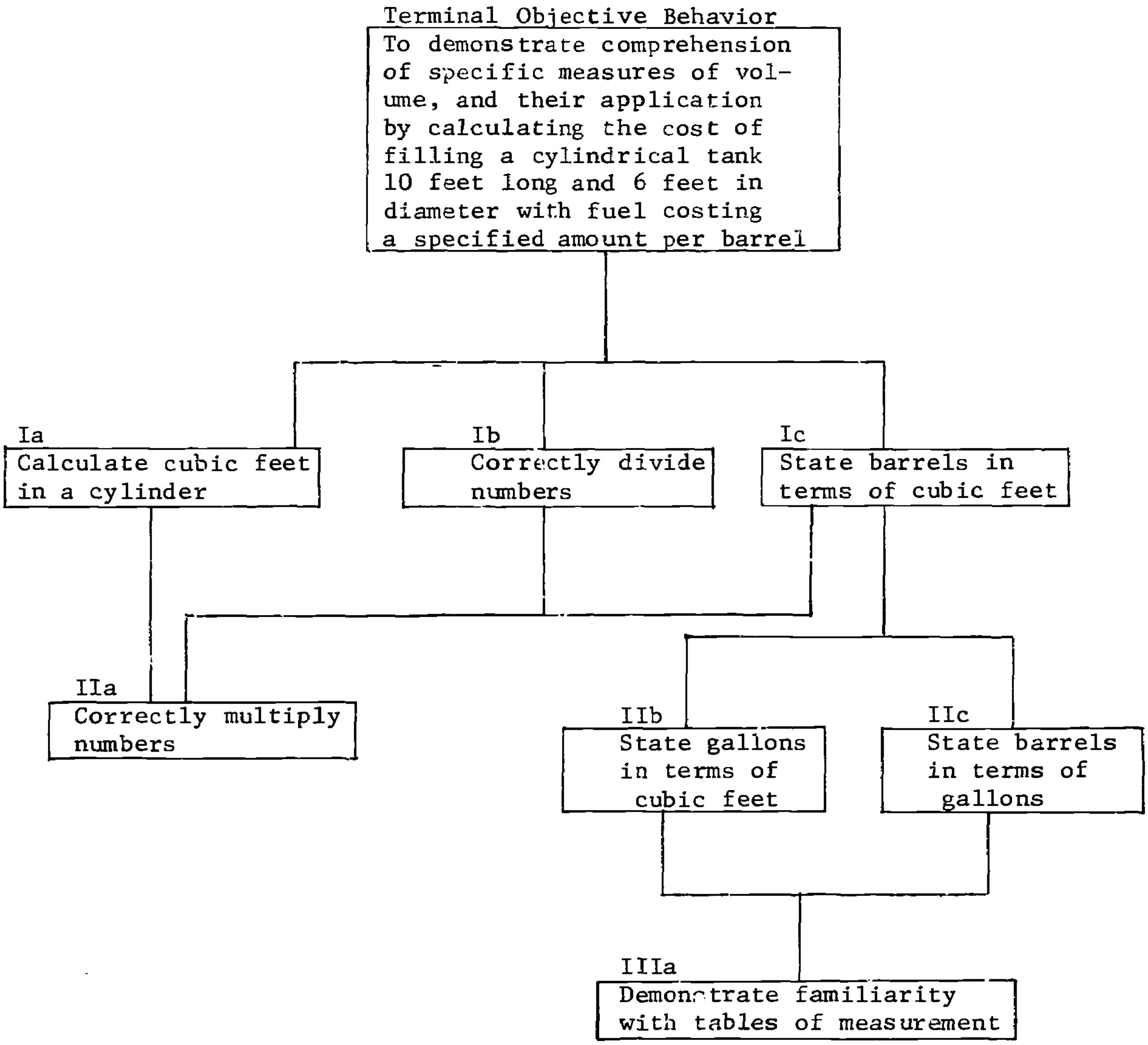


Table 1. Hierarchy of enabling objectives in mathematics

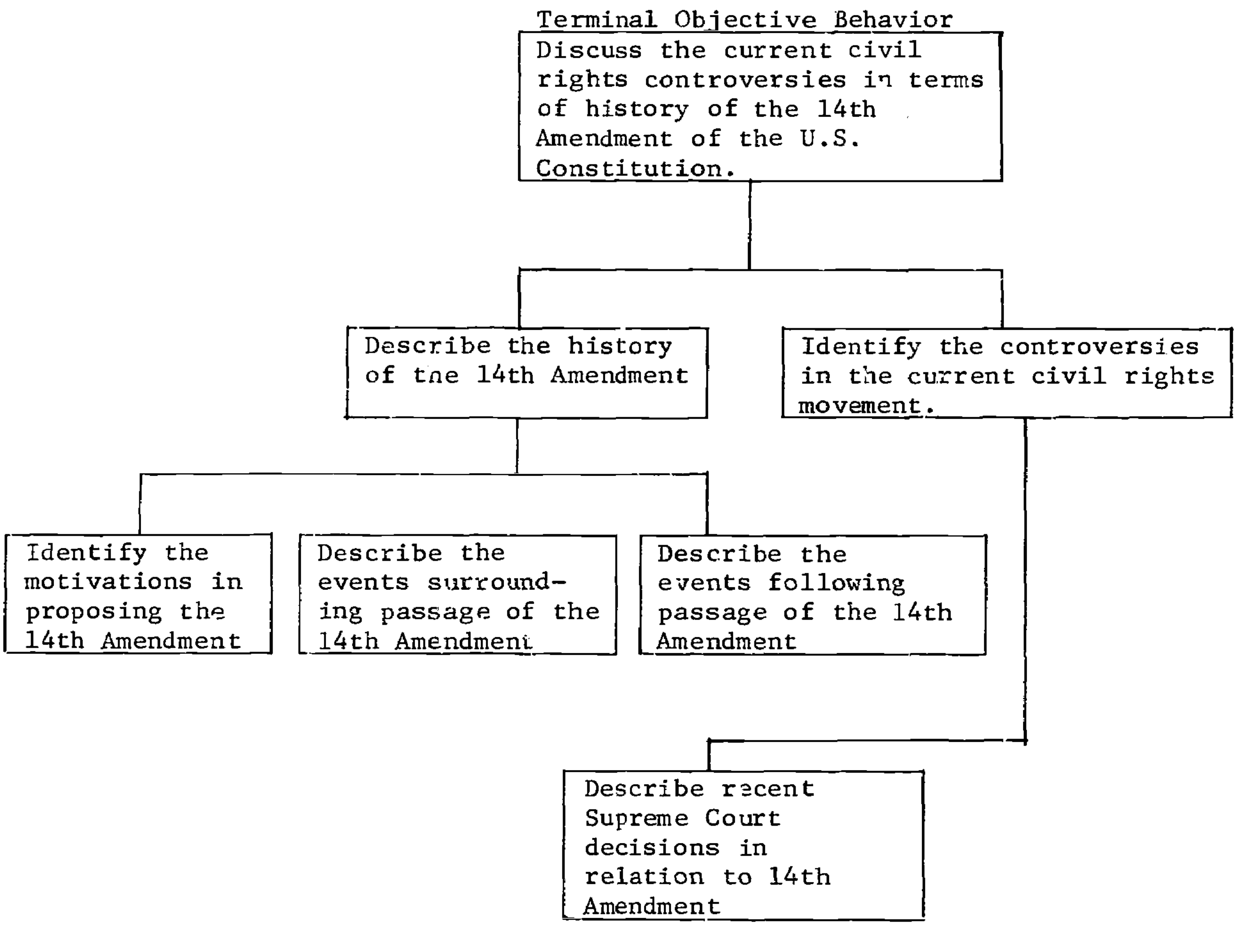


Table 2. Hierarchy of enabling objectives in social studies.

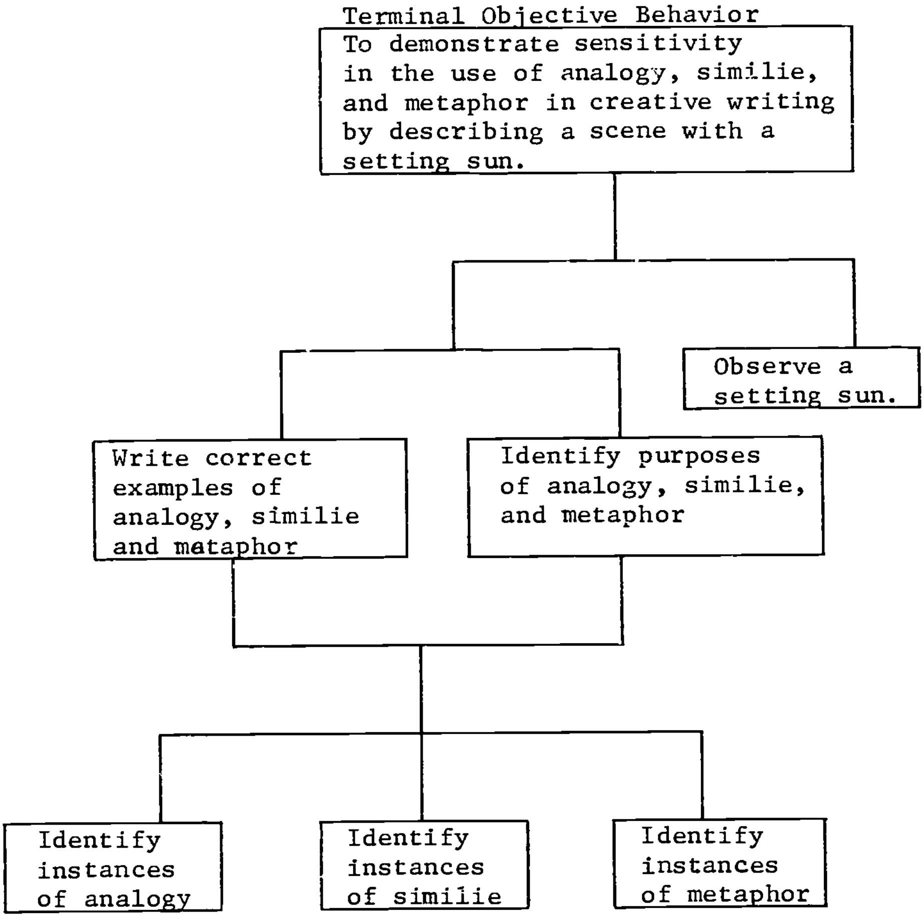


Table 3. Hierarchy enabling objectives in English.

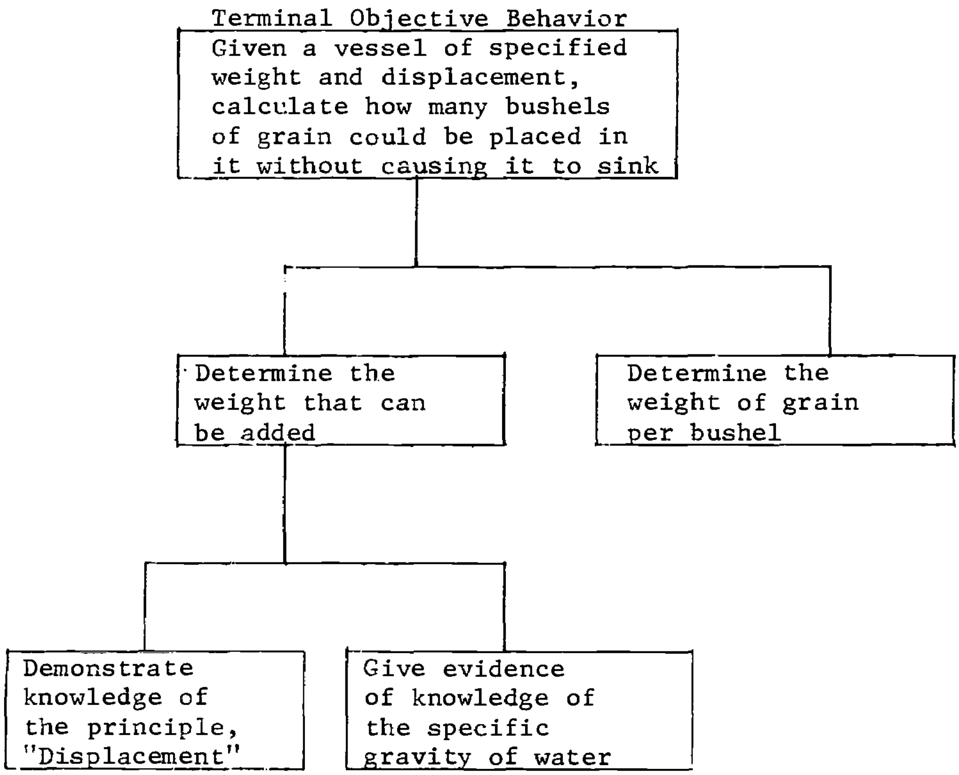


Table 4. Hierarchy of enabling objectives in science.

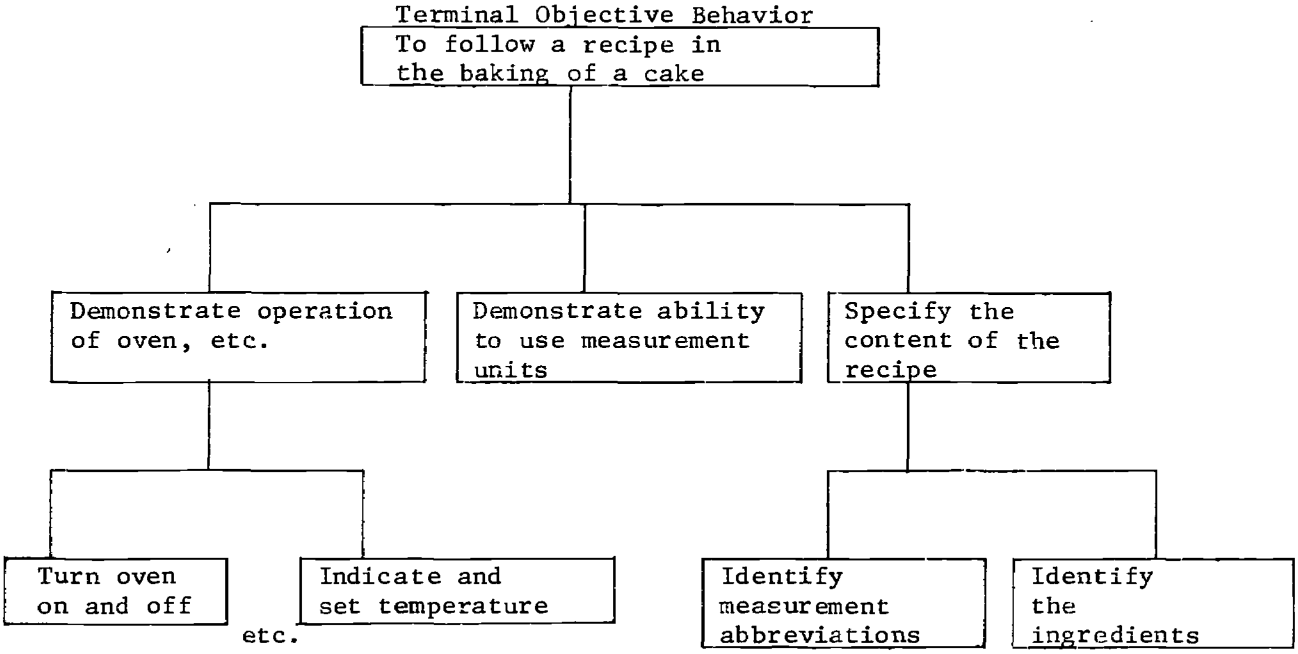


Table 5. Hierarchy of enabling objectives in home economics.

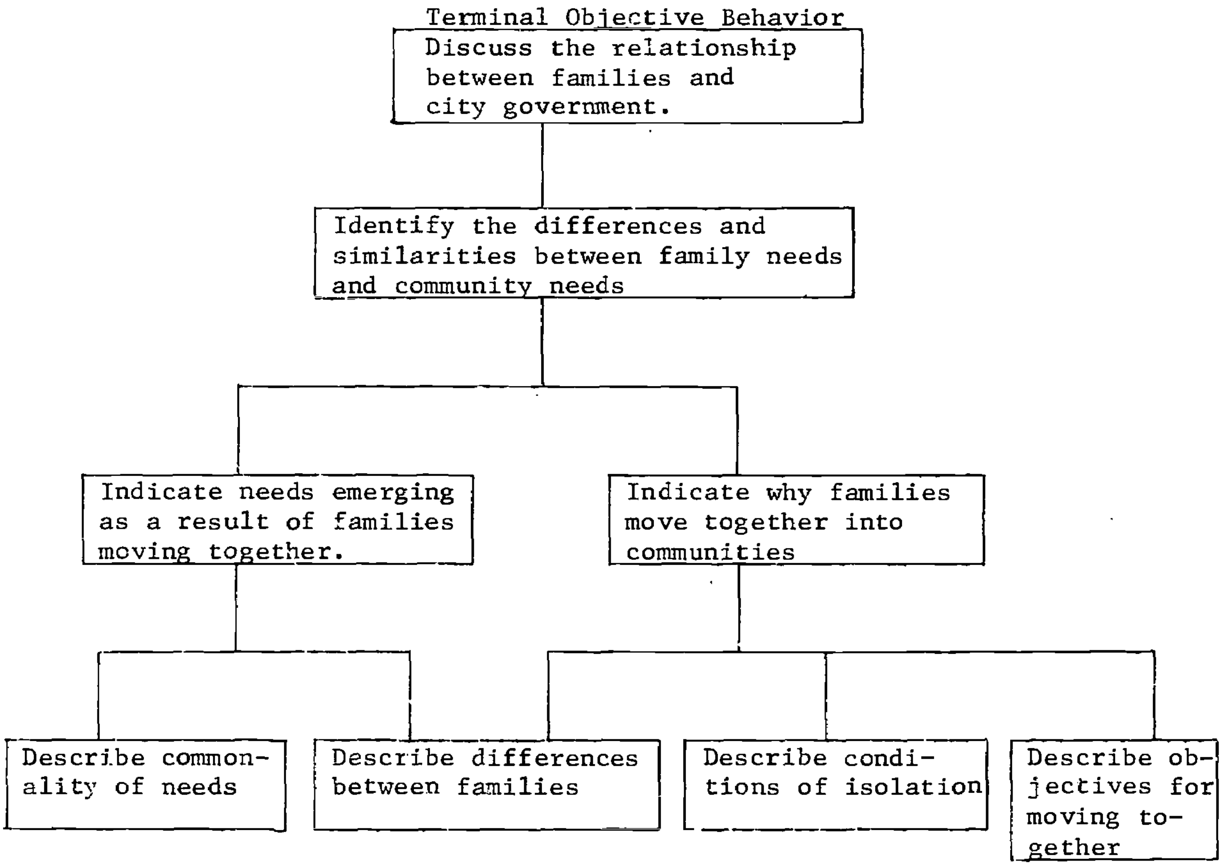


Table 6. Hierarchy of enabling objectives in social studies.

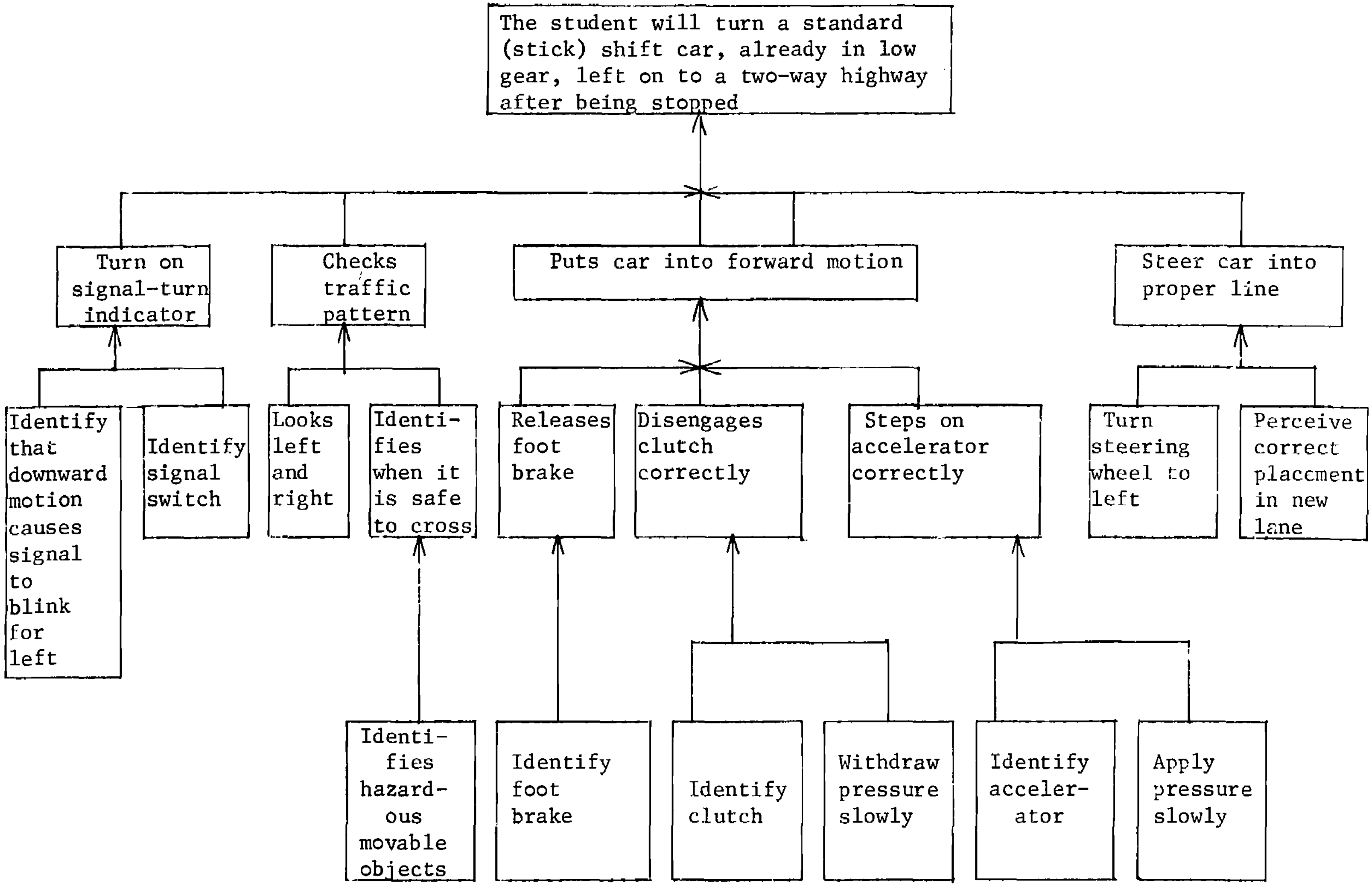


Table 7. Hierarchy of enabling objectives for a psychomotor skill.

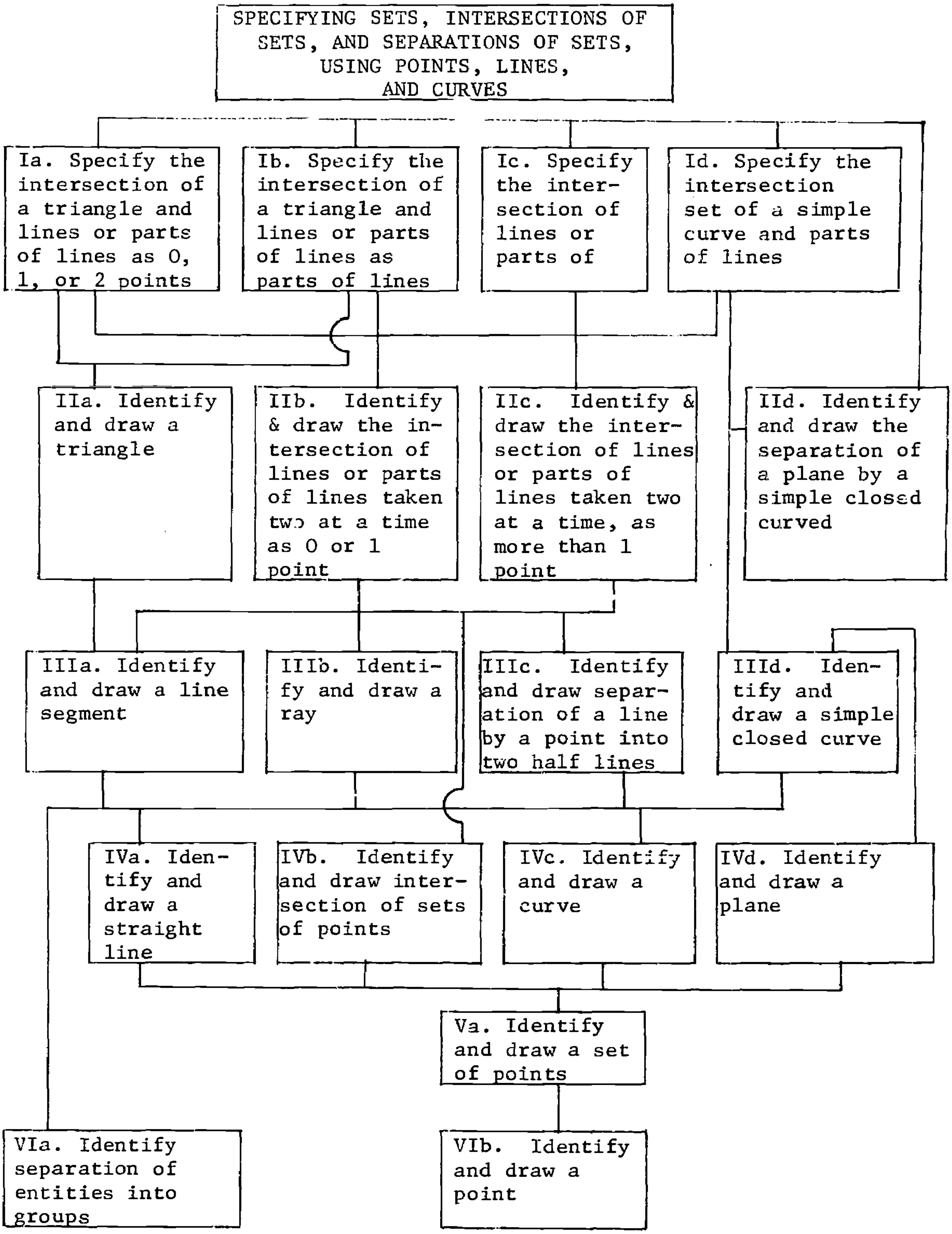


Table 8. Hierarchy of enabling objectives for a topic in elementary non-metric geometry. Adapted from Gagne (1965).

ERIC Full Text Provided by ERIC

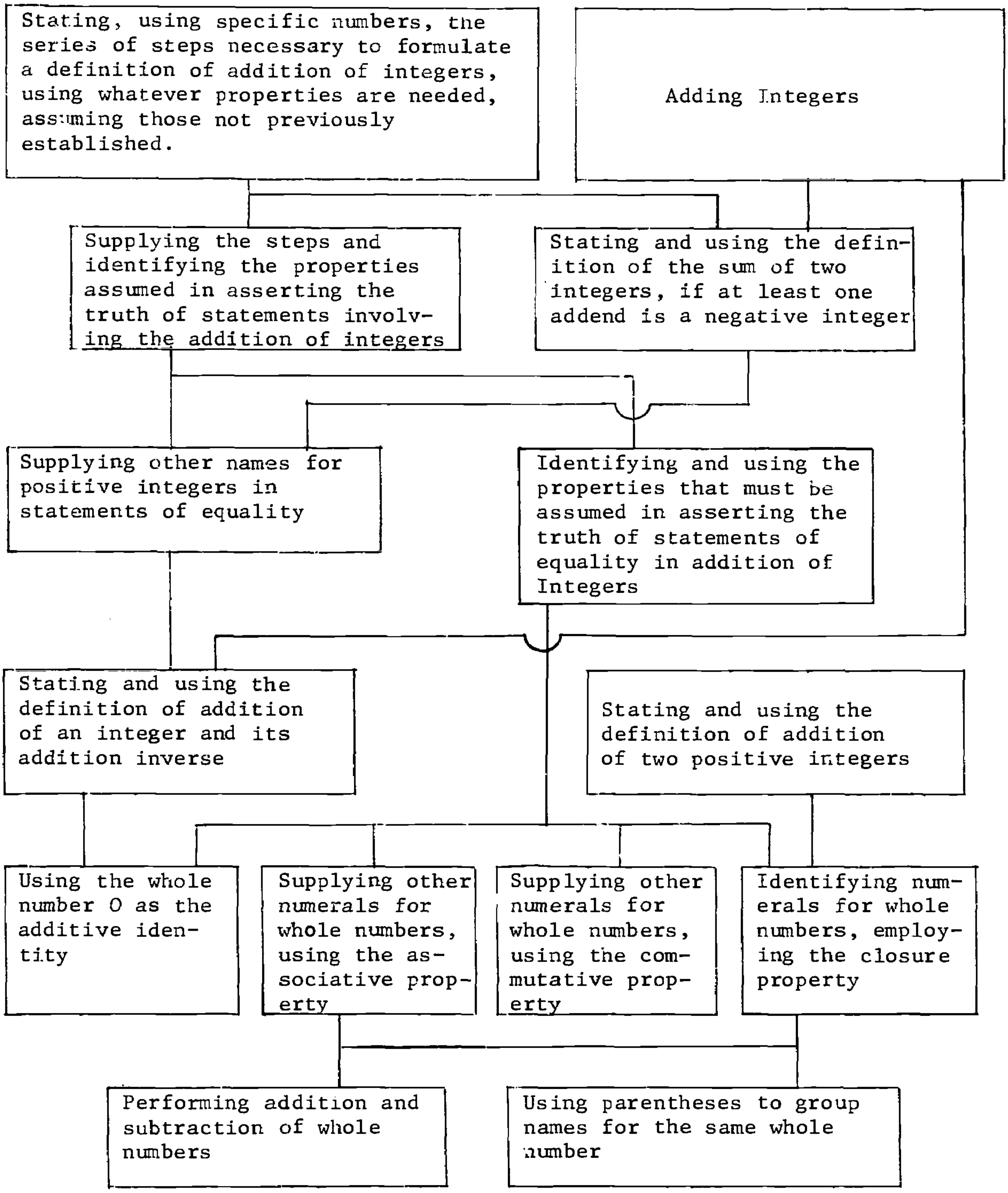


Table 9. Hierarchy of enabling objectives for a topic in mathematics. Adapted from Gagne, et al (1965).

Desired Capability of Learner

↑

OBJECTIVE ANALYSIS

↓

Initial Capability of Learner

Is there evidence available that shows what happens when learners actually undertake to acquire a set of principles that appear to have a hierarchical structure determined through these procedures? A few studies have been undertaken, but the analyses have been restricted to competencies in mathematics and science. You may wish to examine one of these studies. The study sought to answer the question, "Does the mastery of a lower order objective actually affect the learning of the next "higher" objective as would be expected?" The results of the study are summarized below:

> The results showed that the learning of 'higher-level' principles was dependent on the mastery of prerequisite 'lower-level' principles in a highly predictable fashion. For example, of the 72 students who performed correctly on principle IIa, only one did not perform principle IIIa correctly on the test. Of the 18 students who did Principle IIa incorrectly, all 18 did principle IIIa incorrectly. The prediction that learning IIa depends on knowing IIIa was borne out, therefore, with a frequency of 99 percent. For all the other possible comparisons,...the frequency of correspondence between predictions and findings ranged from 95 to 100 percent. The learning of organized knowledge, according to these results, appears to be predictable from the pattern of prerequisite principles that make up the hierarchy of knowledge to be acquired. (Gagne, 1965, p.153).

The conclusion cited above has been verified in a number of studies using topics in mathematics (Gagne, 1962; Gagne et al., 1962; Gagne and Paradise, 1961).

It is a relatively simple task to <u>order</u> enabling objectives, asking the question, "What would the learner have to know?" It is a more difficult task to <u>derive</u> the enabling objectives in your own objective analysis. Rarely will you be fortunate enough to have the enabling objectives already

formulated and diagramed for your particular terminal objective. One thing is certain. In deriving the competencies in an objective analysis, you will progress in a rather haphazard fashion through the analysis. It may be true that you will begin at the terminal objective, but you may find it easier to skip around the developing hierarchy, filling in those parts that you are most familiar with until the analysis is completed. In many cases, you will find that as you proceed down the hierarchy to the more simple level competencies, you may skip some of the required competencies or perhaps fill in a few that are irrelevant. In some cases, you will find yourself stating competencies that are too detailed and should be omitted.

Briggs (1968) suggests that there is a great need to conduct more empirical studies in several subject-matter areas other than mathematics and science. In this way, evidence could be gathered to confirm whether or not students taught by an "optimal" sequencing method learn more efficiently or more effectively than students taught by a random sequence determined in the objective analysis. For readers who would be interested in such experiments, the reading of Briggs' discussion of sequencing of instruction in relation to hierarchies of competence is highly recommended.

Sometimes the question is asked, "If an objective analysis is done correctly, can there be more than one structure obtained?" To quote from Gagne and Paradise (1961, p.5), "...there are perhaps several <u>possible</u> learning set hierarchies which could be worked out..., and it is quite conceivable that some are 'better' than others in a sense of being more efficient or more transferable to later learning." In the long run, any hierarchy that is developed is "good" only to the extent that it aids the instructor in determining what to teach, and in what sequence to teach it.

Another way to describe the organization of enabling objectives is by using an outline. The terminal objective diagramed in Figure 1 could be outlined as follows:

1. Terminal Objective
    - A. First level enabling objective
        - 1. Second level enabling objective
        - 2. Second level enabling objective
    - B. First level enabling objective
    - C. First level enabling objective
        - 1. Second level enabling objective
            - a. Third level enabling objective
            - b. Third level enabling objective
        - 2. Second level enabling objective
        - 3. Second level enabling objective

However, as shown by Tables 1-9, some lower-order objectives "tie-in" to more than one higher order objective. The hierarchy that is diagramed in Figure 1 shows these relationships more easily. It is more useful in organizing objectives to type or print each component on a 3 x 5 inch card and sort these cards on a large table. You may find the chalkboard too inflexible to be of much value. When the objectives have been arranged, they may be diagramed.

In summary, let us review the steps involved in objective analysis.

Step 1 Identify terminal objective.

Step 2 Identify learner entry level.

Step 3 Starting at the terminal objective, ask the question, "What kind of capability would an individual have to

possess to accomplish this objective?"

Step 4 Repeat procedure for each competency that is determined.

Step 5 As prerequisite competencies are determined, begin to arrange in hierarchical fashion. (Using 3 x 5 cards are helpful here.)

Step 6 Continually check efforts to assess relevance of competencies thus determined and to weed out competencies that are inappropriately stated, too detailed for the entry level of the learner, etc.; asking the question, "Do I really need this competency?" is useful here.

2.1

Specifying Instructional Conditions

How can the instructional systems designer best specify instructional conditions? Are some instructional methods more effective than others in certain learning activities? It is probably safe to say that our present state of knowledge exhibits more of the characteristics of an art than a technology. However, we cannot go so far as to say that there are no principles or procedures that might guide our efforts. To be certain, there are some general rules that seem to hold in a variety of conditions. Such principles, if they be that, that might seem to hold in a variety of conditions, include the provision for a proper feedback, active participation, spaced practice, and so forth. But, it is clear that these rules of thumb do not lead us far enough down the road of instructional specification to be of much help at this stage in our technology. Many of the decision questions

being asked in the course of specifying instructional conditions cannot be answered by examining past research, theory, or intuition. For example, Saul et al., (1954) prepared annotated reviews of the literature in specified areas pertinent to the problem of developing standards, criteria and utilization of effective graphic training aids. These reviews served as a preliminary basis for the formulation of principles of design and utilization of graphic training aids. Results of the study revealed that there was a wide variety of information available, but much was contradictory and inconsistent. "Extensive and detailed study and evaluation of these data are necessary prior to their application as criteria in the design, preparation and use of visual training aids."

We cannot be dogmatic when it comes to specifying instructional conditions. Empirical trials and revision of the system based on the evaluative data gathered must be used in conjunction with procedures that help the designer specify the best possible specifications initially. Procedures for developing and revising instructional prototypes are discussed elsewhere in this manual.

A Prerequisite for Specifying Instructional Conditions

Most instructors have had ample training in substantive areas. Yet, when it comes to designing instructional systems, their training is of little benefit without the catalystic effect of another individual trained to identify the optimal means for developing behavior. This second party, an instructional technologist, has the difficult task of relating processes implicit in learning to the requirements of the particular instructional situation.

ERIC

Our purpose here is not to make instructional technologists of substantive individuals. Neither is our purpose to train individuals who might wish to become instructional technologists. Such an undertaking would certainly require more than what is contained in this manual. Rather, our purpose is to sensitize the individual who is attempting to develop an instructional system in his own particular substantive area to the whole matter of specifying instructional conditions. An attempt will be made to demonstrate the crucial steps involved in specifying instructional conditions so that the individual may be better able to work with the instructional technologist in a meaningful and productive relationship. In a real way, the guidelines presented should be thought of as benchmarks. A benchmark, in surveying terms, refers to a reference point from which further measurements may be taken. The guidelines should be taken as springboards to further discussion and inquiry, and are not meant to be answers in themselves. They represent a process, not a product. The usefulness of these guidelines lies in the fact that other points may be established from them, points that lead the instructor directly to the specification of instructional conditions. They furnish the instructor with a starting point, a foundation upon which can be built and developed an instructional system that accomplishes prespecified outcomes.

All of the guidelines to be presented may not be new to you. Certainly, a competent instructor utilizes many of the procedures listed as he attempts to prescribe a learning experience for a course. Further, these guidelines do not at all inhibit an instructor's creativity in

specifying instruction. Rather, it is hoped that the guidelines will enable an instructor to be optimally creative. In other words, what are presented below are not answers for the instructor who is grappling with the problem of instructional specification; but are only suggestive of appropriate questions an instructor should ask as he sits down to determine how best competencies might be taught.

Variables to be Considered in Specifying Instructional Conditions[2]

When one wishes to specify instructional conditions, it is imperative that he keep in mind all of the inter-relationships between the various outcomes or competencies he wishes the learner to exhibit and the factors which affect these outcomes. Figure 2 illustrates the various factors to be considered in specifying instructional conditions to achieve one competency, i.e., one enabling objective, hereafter referred to as a "unit of instruction." In the context of this paper, the term "enabling objective" refers to those competencies that a student must learn in order to arrive at a more general and more complex terminal objective. That is, enabling objectives state prerequisite behaviors to more complex objectives. Typically, enabling objectives might involve competencies such as, "identify and draw a set of points," and "indicate and set temperature." The factors that the designer of an instructional unit must consider are:

1) Learner characteristics,

2) The context in which instruction takes place,

3) The stimulus situation(s) that serve as instructional events,

---

[2] The discussion of instructional design variables, and subsequent discussions of specifying instructional conditions draws heavily on previous work at Teaching Research (Schalock, 1967, Schalock et al., 1969).

ERIC

UNIT "A"

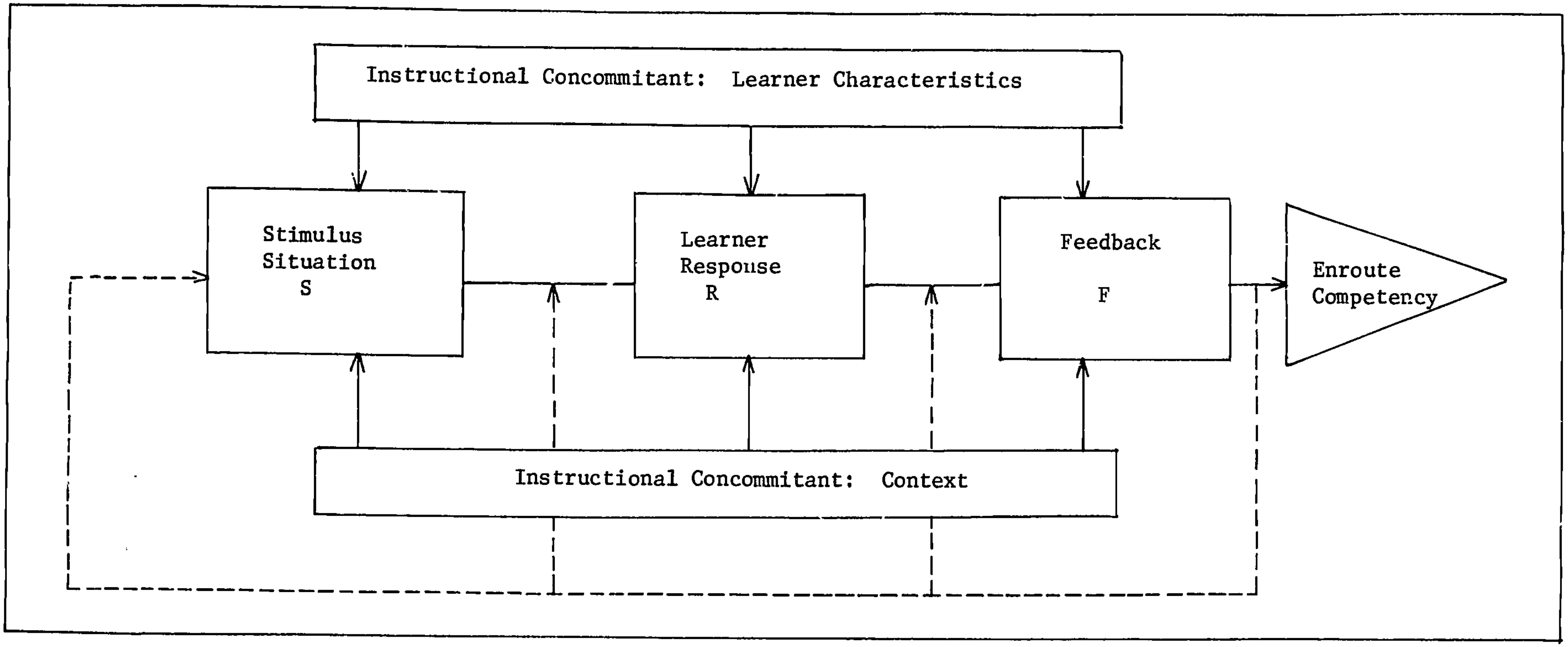


Figure 2. Variable to be considered in specifying instructional conditions for one enroute competency -- a "unit" of instruction.

4) The learner responses required for the competency, and
5) Feedback situations that again represent instructional events given to the learner.

A crucial assumption is that the instructional conditions required to bring about the particular learner competency desired will vary according to what types of learners are involved as well as the context in which the instruction will be given. By context of instruction is meant such factors as the organization of the learning space, the number of learners, male-female ratio, learner grouping, and physical characteristics of the learning space. The fact that little regard historically has been given these factors does not make them any less important.

The inclusion in the model of the three factors, the stimulus situations, the learner response, and feedback, should not be surprising. The stimulus situation represents various types of instructional messages presented to the learner. These messages may be rather short and concise, e.g., "Pick up the chemical and place it in the flask," to rather lengthy discourses. Somewhere along the line, however, the instructor must know if the messages presented are effective. This requires that learner responses be elicited to assess performance. Again, instructional messages may be given to the learner in the form of feedback.

All instruction need not follow the stimulus-response-feedback cycle. Lectures represent the almost total dependency on the presentation of instructional messages without the evocation of responses or the presentation of feedback. At this point, no value judgment is being attached to this mode of instruction, but it will probably become clear in subsequent

ERIC

discussions that this technique has both advantages and limitations. Whereas lectures almost exclusively represent presentation of instructional messages to the learner, here labeled stimulus situations, other techniques such as the programed instruction and instructional simulation exercises represent a class of techniques that emphasize the learner responses and feedback factors as well.

It is unlikely that the instructor wishes to build an instructional system that simply involves one enabling competency. Usually, several enabling competencies must be combined in some way to arrive at a terminal performance such as discussed within objective analysis. Hereafter, this combination of instructional conditions to teach several enabling objectives will be referred to as an instructional "sub-system." When a sub-system of instruction is specified the relationship between the manner in which each enabling competency is taught must be examined. The designer must consider the sequence of instruction as well as the general context in which instruction takes place. This is graphically illustrated in Figure 3.

Just as it is probably a rare case that a designer wishes only to consider the teaching of one particular objective, it is also rare that an instructor limits his instruction to but one terminal objective. Usually, the designer wishes his instructional system to incorporate a number of terminal objectives. It should be recognized that semantics may be confusing at this point in that really none of the objectives might properly be considered "terminal" objectives. The designer could build a system that involved objectives for the entire general education

SUB-SYSTEM 1

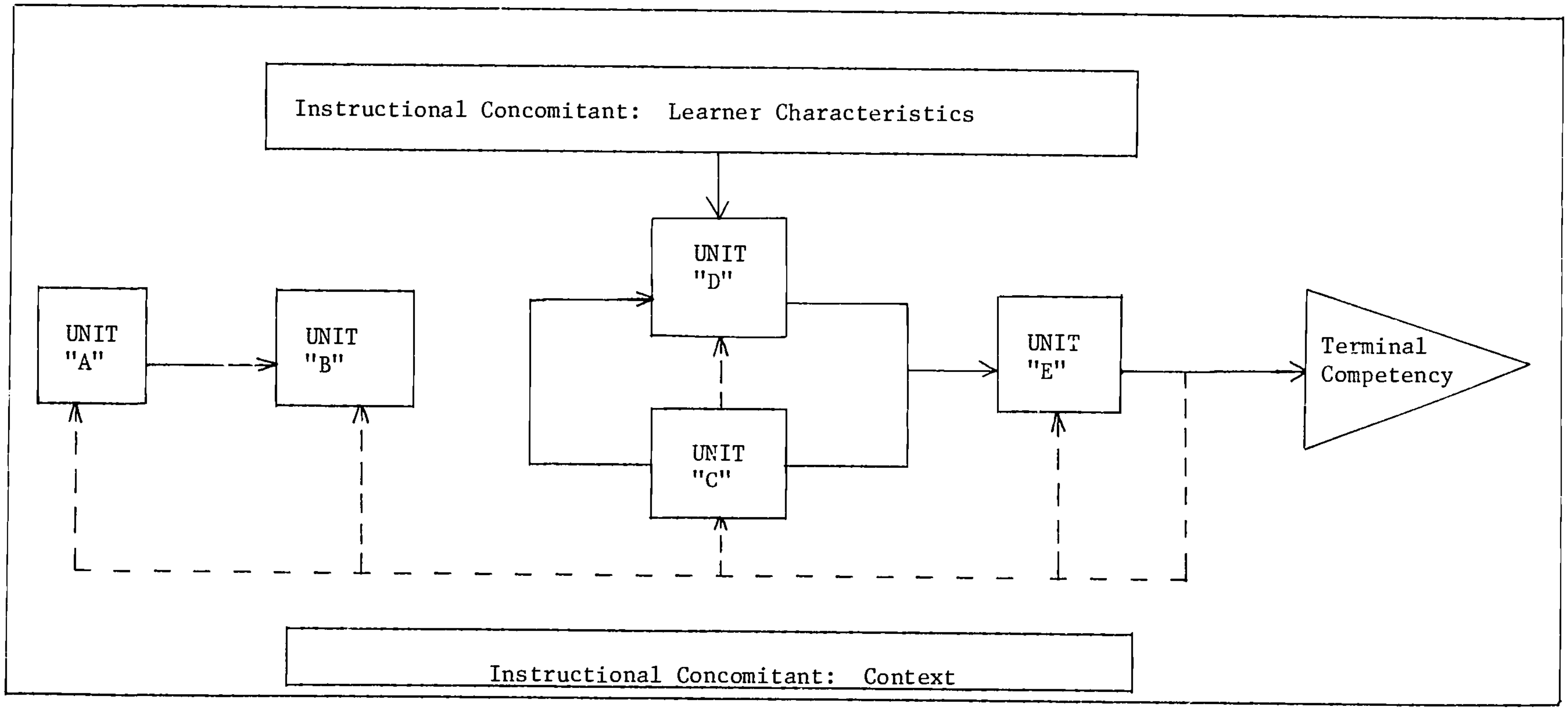


Figure 3. Graphical representation of variables to be considered in specifying instructional conditions for a terminal competency--a "sub-system" of instruction.

curriculum. If this were the case, we could go on and on in terms of specifying how each component was related to the other, and the context in which instruction would take place across all components. However, for the present paper, simply consider an instructional system of being composed of a combination of instructional units, each teaching an en route competency, and a series of sub-systems (each teaching toward one or more terminal competencies). These relationships are graphically represented in Figure 4.

From Figures 2, 3, and 4, twelve steps break out as important for an individual to consider as he specifies instructional conditions. They are:

(1) Identify learner characteristics. This maximizes the adaption of instruction to the learner's personal needs, and reduces the chance of the system simply becoming a slick "Madison Avenue" package.

(2) Identify tentatively the general characteristics of the instructional system to be used to achieve the terminal objective(s). That is, attempt to look at the whole system and outline its characteristics.

(3) Identify tentatively the relationship between, and general characteristics of, the way in which en route objectives in the instructional system will be taught. This lets the designer match his tentative overall specifications for the system with each objective in the system.

An Instructional System

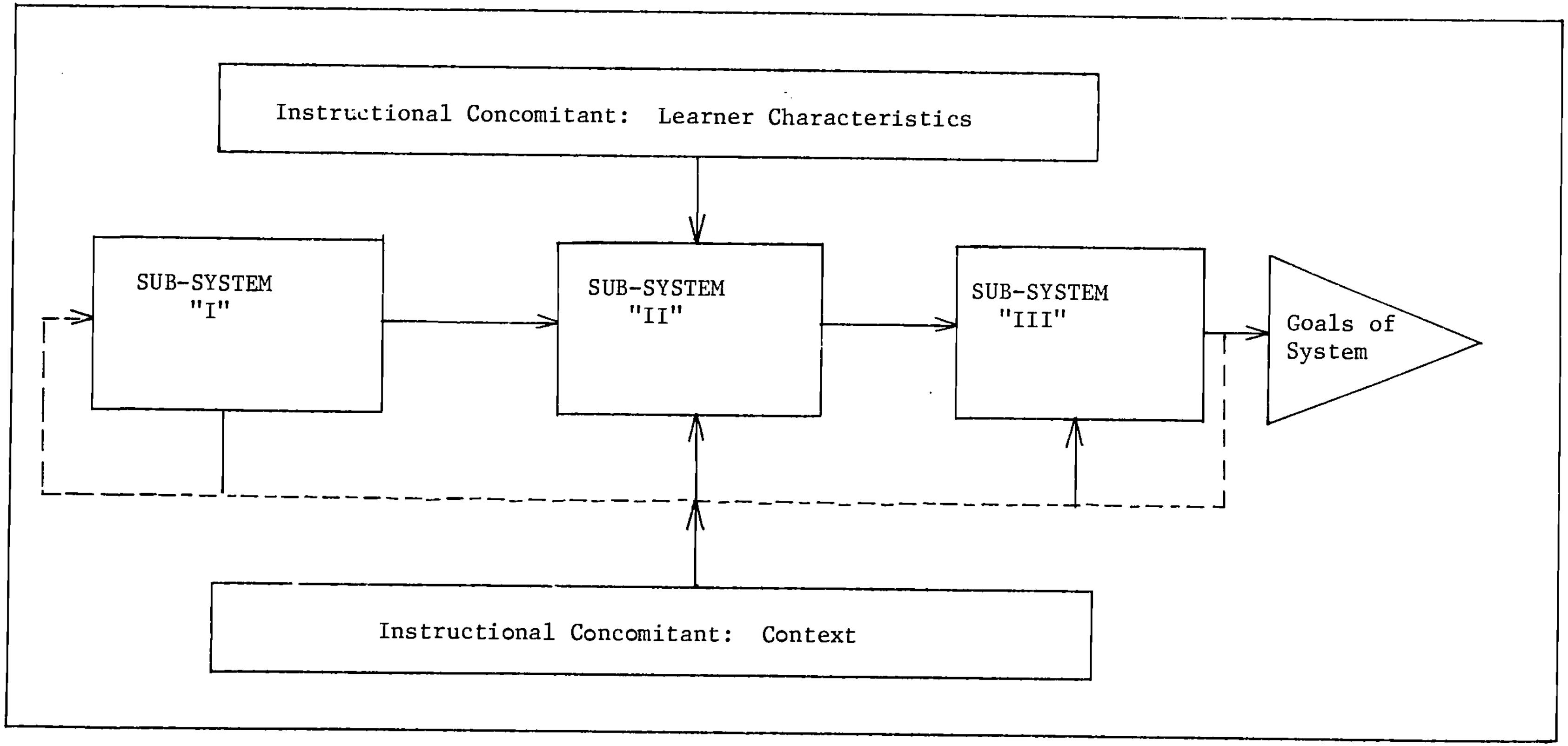


Figure 4. Graphical representation of variables to be considered in specifying instructional conditions for an instructional system.

Then, for each <u>en route competency</u>,

(4) Identify the type of learning function represented. Is it problem-solving or multiple discrimination or something else?

(5) Identify the instructional strategies that provide general conditions of learning. If general principles govern learning of one type or another, they must be identified.

(6) Specify the learner response(s). What is the learner supposed to do? What is the form of the response? What media are required?

(7) Specify the stimulus situation. What is the occasion for the response, or what precedes the response in way of exposure to information or orientation? What form does it take? What media are required?

(8) Specify feedback for each instructional event. How are you going to tell the learner that his response is correct? What will you tell him?

(9) Specify the required or permissible context of instruction. In what environment does all this take place?

(10) Specify the appropriate sequence for each instructional component to ensure optimal mediational effects from one component (i.e., response, stimulus, or feedback) to another. How responses, stimuli, and feedback are put together can't be ignored.

For each instructional <u>sub-system</u> and <u>instructional system</u> as a whole,

(11) Specify the required or permissible context of instruction, considering the relationships and specifications previously identified. Now that the instructional conditions for each objective have been

specified, take a new look at the tentative specifications listed from Step 2 and 3. Adjust accordingly.

(12) Specify the appropriate sequence of all instructional units to ensure optimal mediational effects from one unit and one sub-system to another. Once this is done, sit back and relax. If you have done your job well, your blueprints can now be used by someone else or yourself to build an instructional system.

C. Step One: Identify Learner Characteristics

Ideally, an instructional system must have the instructional events (the content and the operation of presentation) well matched with the characteristics of each learner--his capabilities, his personality, his learning style, his previous experience, and so forth. This is in direct contrast to most instruction that takes a standardized form, i.e., instructional methods employed that are common across all learners, regardless of their individual characteristics. Although practical and economical factors may preclude the use of materials designed with individual differences in mind, educationally such instructional systems geared to the norm may be far less effective than those designed to capitalize on and take into account known relationships between learner characteristics and instructional methods.

This means that when instructional specifications are developed, the designer must, to the best of his ability, know his audience. That is, he must identify what factors of individuality are most likely to interact with the instructional events in a given instructional context. Note that in Figure 3 the learner characteristics are concomitant not only to the stimulus situation but also the learner response, the feedback, and the context of

instruction as well. Too often individual differences are taken into account only when the stimulus situation is specified. Yet, in a very real way, the learner's unique set of characteristics may interact with the response required and feedback given, and certainly with the context in which instruction will take place.

What learner characteristics might be important in the specification of instructional conditions? Are some learner characteristics more crucial than others? In answer to the first question, it is not too difficult to at least list several factors that may be measured that would tell us something about the learner. These factors are:

(1) Personality characteristics

(2) Capabilities or aptitude

(3) Learning style

(4) Age

(5) Sex

(6) Race/Culture

(7) Social class

(8) Motivation

(9) Previous experience

(10) Perceptual sets, and

(11) Personal history

It is more difficult to determine which of these factors may be more important than others when one wishes to relate them to the specification of instructional conditions. Empirical evidence is sketchy. This writer knows of no research that has compared the importance of including one or another type of learner variable in the consideration of the design of the system and its contribution to the overall success of the instructional system.

Are there data to support the formulation of any unified theory regarding interaction between instructional methods and learner characteristics? In other words, are there principles which will allow an instructor to specify instructional conditions given knowledge of the learner? In their review of studies that examine the interaction between individual differences and instructional methods, Tallmadge and Shearer (1967) and Tallmadge, et al., (1968) found that the individual difference variables which occurred most frequently in significant interactions were related to such factors as intelligence, ability, and academic achievement. Relatively little success was obtained when other individual difference variables such as personality and motivation were studied. It was found that studies designed to measure these latter variables were often characterized by relatively lower reliability and validity than measures of aptitude. They were more difficult to define and measure reliably and thus seemed to be less often used in studies. However, in their review of the studies dealing with aptitude measures, Tallmadge and Shearer (1967) concluded that no particular method or combination of methods have been shown to be most effective in instructing the students at various ability levels. Many of the results were conflicting, others were ambiguous or inconclusive, and those that seemed to be reliable did not fit into any recognizable pattern. In fact, from findings from their own project, a third factor, the type of material being learned seemed to be an important variable that needed to be considered in the study of instructional methods as they relate to individual differences.

From a follow-up study, Tallmadge et al., (1968) obtained further data to substantiate the notion that a large number of variables exist that not

only influence effectiveness but also interact with individual differences and each other in such complex ways that a thorough understanding of all of the interrelationships is impossible to achieve in any given study. However, their results strongly support the existence of learning <u>styles</u> and suggest that multi-track instruction based on learning styles is feasible. Incidentally, the individual differences which interacted with subject matter and instructional methods <u>were all shown to be non-cognitive in nature</u>.

From this cursory review it would seem, then, that the instructor or designer of instructional systems is practically on his own in regard to the wedding of individual differences with instructional requirements. Indeed, the way open to the instructor for matching students with common attributes to appropriate curricula or units of instruction are very gross in nature. Tallmadge, et al., (1968) suggests several ways:

(1) Classify students on the basis of administered aptitude and interest tests. Assign students to courses (or instructional systems) which differ in terms of intellectual demand and difficulty level.

(2) Classify students in regard to differential aptitude patterns and interest data. Assign students to courses (or instructional systems) encompassing the appropriate subject matter content.

(3) Classify students according to their learning style. Assign them to courses or instructional systems employing appropriate instructional methods.

The first strategy seems to be successful when cost and effectiveness are considered. The second method is less widely used, but seems to be

ERIC

appropriate. In regard to the third method, learning style is defined by Tallmadge as a characteristic of a learner which interacts with the instructional situation in such a way as to produce differential learning performance as a function of the situation. This matching technique has not been shown to be of practical value yet.

What a dismal state of affairs! In our opening remarks, it was suggested that a requirement of an ideal instructional system is that the instructional events be well matched to the characteristics of the learners. Then, come to find out, there are not any empirical data that lend themselves to the formulation of any unified theory regarding interaction between training methods and learner characteristics. Once again, the reader is directed to the opening remarks regarding the relation between research and the instructional systems approach, and the handicap in designing instructional systems. Identified here is a large gap in our knowledge. With the aid of an instructional technologist, the instructor can examine what available data there are, and tentatively use this knowledge in the specification of instructional conditions. Perhaps in the tryout and revision cycle information can be gained that will better enable the instructor to match techniques with learners. For the reader interested in this problem, an excellent discussion of learner variables and the instructional technologist has been prepared by Beaird (1968).

D. Step Two: Identify Tentatively the General Characteristics of the Instructional System to be Used to Achieve the Terminal Objective(s)

Before proceeding further it is imperative that the instructional designer examine once again the instructional problem that has led him to

tamper with the status quo and to examine the numerous solutions which were proposed in his initial efforts. The "way in which the instructional system will look" after completion will depend to a great deal on the scope of the problem and the proposed solution. That is, the designer should assess whether the instructional system is one that will cover only one or possibly several terminal objectives (a "micro-system") or one that covers many objectives that are related in some complex manner (a "macro-system"). Before the designer gets down to the nuts-and-bolts operation of specifying the instructional conditions for each enabling objective, he must consider all of the enabling objectives and terminal objectives as a totality. It is inconceivable that each objective that the designer has identified as important to teach will be taught in a way that is isolated and unrelated to other objectives. Indeed, the analysis of objectives discussed earlier implies that certain objectives are closely related to each other, and might best be taught in one or another manner that capitalizes on their inter relatedness. In this step, the instructional designer has a chance to identify the general characteristics of the instructional system that he is about to build in very general terms. As an example, from the analysis of the instructional problem and the thinking about proposed solutions to the problem, he might want to investigate the use of a simulation exercise or programed instruction or some type of independent study or the audio-tutorial approach. Note that the specification of these general characteristics is tentative. After proceeding through the model for specifying instructional conditions, the designer may find that the requirements for teaching one or another component are

not at all matched to the type of instructional system he has identified. Yet, to examine how each individual component (usually an en route competency) is taught without first looking at the total system in which the component is taught may result in wasted effort and an ineffectual instructional system.

How does one identify the general characteristics of an instructional system? How does he determine which instructional system may be more appropriate than another? Unfortunately, there exists no manual or set of guidelines that would allow an individual designer to choose between one or another type of instructional system. In some respects, the individual designer must set his own criteria for acceptance or rejection of a particular technique. Yet, there are available in isolated instances some general guidelines that may be of benefit to the designer. For example, Crawford and Twelker (1969) give several possibilities for determining the appropriateness of simulation exercises as one useful and cost-justifying alternative. They go one step further in that if the designer decides that simulation is appropriate, the authors present guidelines to help the designer determine what type of simulation is best suited for his individual needs. These guidelines have been revised into chart form by Twelker (1965, Chapter 10) and A Department of the Army manual (U.S. Army, 1967). The guidelines by Gagne appear in Appendix A as well.

Another example might be cited. Should the designer include any (live) instruction in the system or rely on all-machine components? Kersh (1964; 1967) makes the case for the inclusion of an instructor

in discovery teaching. He argues that there are some instructional objectives that are amenable to automated or self-instruction and there are other instructional objectives which are most readily developed through human instruction. It is suggested that involvement from an instructor may be required in the attainment of instructional objectives for one or more of the following reasons:

"(1) The required behavior cannot be identified by machine processes presently available or by the learner himself without previous instruction.

(2) The required behavior cannot be readily elicited through direct or indirect intercommunication with another person who is capable of identifying the required behavior once it has been elicited.

(3) The learner cannot determine that he is making progress toward the instructional objective independently comparing his own behavior against a behavioral standard or model."

Kersh further suggests that instructional objectives that involve the attainment of factual knowledge are amenable to automated instruction while objectives which involve patterns of behavior occurring at unpredictable intervals and reflecting "mediational" processes will be more readily attained through human instruction. Thus it follows that the processes involved in learning by discovery probably are most readily attained through the use of human instructor. In studies conducted by Kersh (1964) and Twelker (1967), only an instructor could identify approximations of the class of complex behaviors called "searching behaviors" among a variety of other behaviors shown by the students during instruction.

When this behavior did not appear, the instructor was expected to interact with the student in an attempt to elicit approximations of the desired behavior. Finally, only an instructor was able to present feedback to the students, either in the form of knowledge of results or approval of the students' efforts.

In addition, Kersh also suggests that the provision of feedback may be a crucial factor in our choosing between automated or human instruction. When factual material is being learned, knowledge of results must be provided to the learner as he practices. Studies have shown that it is usually desirable to give an explanation for the correct answer, if practical, rather than simply telling the student that he is correct or incorrect. In the attainment of complex objectives, e.g., those that involve problem solving, hypothesis formation, searching for patterns, and so forth feedback may be delayed for some time until the student arrives at an answer to the problem or finally formulates a hypothesis. Kersh suggests that the instructor might interact with the learner during his problem-solving activity and offer encouragement such as "Keep up the good work" or "You are doing very well" without interrupting the learner (cf., Kersh, 1964). The instructor would probably offer such encouragement only while the learner was exhibiting approximations to the behavior that was desired. Of course, if the learner was not behaving in the appropriate manner, the instructor could prompt the learner with suggestions that would lead him toward the use of a correct strategy without giving the answer to him directly.

Easley (1966) argues that a teacher is in a unique position to not only choose among several alternative learning branches of a previously developed lesson plan on the basis of an on-the-spot evaluation of learning, but he may choose to create an entirely different branch at a moment's notice. This characteristic of discovery-teaching is what Easley calls "provocative feedback." "The feedback has the quality of provoking the teacher into abandoning his current teaching tactic (and often his strategy as well) and striking out in search of some more attractive possibilities" (Easley, 1966, p.11).

Other investigators have reported interesting examples of the importance of the human instructor in an instructional system. Silber (1968) reports that in an experiment where children were isolated from each other and from the teacher in a dimly lit room and instruction was presented by a computerized system, the children would reach out to touch the experimenter as he walked about the room.[3] Apparently, physical contact played an important role for the children in this man-machine instructional system where communications with the computer were impersonal and non-rewarding, at least in terms of interpersonal relations.

It is not the purpose of this discussion to discuss the merits of completely automated instructional systems as compared with augmented instructional systems. Rather, the example is presented merely to illustrate the types of decisions the designer must make at this step.

---

3 Personal Communication

Once the designer has tentatively identified the general characteristics of the total instructional system, he must then examine the relationship between, and the general characteristics of, the way in which each objective (enabling and terminal) in the instructional system will be taught. This is discussed below.

E. Step Three: Identify Tentatively the Relationship Between, and General Characteristics of, the Way in Which Each Objective (Enabling and Terminal) Will Be Taught

The purpose of this examination is to find a match between the general characteristics of the instructional system previously identified and each component in the system. Where possible, the instructor should identify how each objective relates to the general characteristics of the instructional system and its components. That is, he should ascertain if the teaching of some competencies conflict with the specification of the characteristics of the overall system. He should determine if certain competencies do not match well his tentative plan, however nice it is in the abstract.

Consider Figure 5. In Step Two, tentative general characteristics of an instructional system are specified (A-H in the figure). In Step Three, the designer determines if these characteristics "fit" for each sub-system and unit of the instructional system. In the hypothetical example, the designer decides that for some sub-systems (and its objectives) the characteristics fit (Component 5, for example). For others, only some of the characteristics make sense (Component 3, for example). For still others, new characteristics must be specified (Component 2, for example).

Step Two

Instructional

System

General Characteristics

of

the System

A

B

C

D

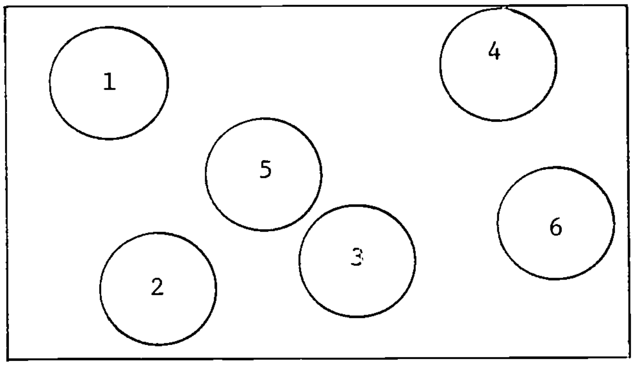
E

F

G

H

Step Three



General Characteristics
of
Each Component Objective

1	2	3	4	5
A	A	A		A
B		B	B	B
C	C	C		C
D	D	D	D	D
E	E			E
	F	F	F	F
G	G			G
	H	H		H
I	I		I	
J	J		J	

Figure 5. Hypothetical example illustrating Steps Two and Three

Let's consider another example--one that has been simplified to illustrate a point. Let's suppose that a designer, upon review of his terminal objectives, decides to incorporate four general characteristics into his instructional system. (We could name many more, but these will do for now.) They are:

(1) Learners will be responding actively during instruction.

(2) Immediate feedback will be given through peers, and not the teacher.

(3) Learners will receive a positive affective experience through interacting with each other.

(4) Due to cost, media will not be used.

(Any similarity between this list of characteristics of the system and a simulation game are definitely not coincidental.)

Now in Step Three, the designer takes a look at each enabling objective to determine the "fit" between the manner in which it might be best taught and the previously specified list of characteristics. This analysis might show, for instance, that one particular objective cannot be taught adequately in a simulation game. Perhaps media might be required. In this way, Step Three serves as a check against faults in characteristics specified in Step Two.

Another point should be made. The relationship between objectives and the components of the instructional system should be examined. For example, in an instructional simulation system, at least five general phases of instruction may be identified:

(1) Pre-simulation system activities

(2) Briefing

(3) Conduct of the simulation exercise itself

(4) Debriefing

(5) Follow-up activities

In this case, the simulation designer should ask himself where a particular objective best fits in terms of these five phases. It might be that a particular objective is best covered in the briefing session, reinforced in the simulation exercise and expanded in the debriefing or post-exercise evaluation. Another objective might best be handled in the simulation exercise itself. Still a third objective might be covered solely in the exercise briefing. Only until these decisions are made can the designer specify instructional conditions for each enabling and terminal objective which make any sense and which have any relation to the instructional system that will be emerging. This is why it is crucial to identify early in the "game" the general characteristics of the total instructional system and the way in which the objectives previously identified relate to these general characteristics.

It should be emphasized again that there is nothing wrong with having a "bias" or "prejudice" about how the content is best taught before the designer actually sits down and specifies the instructional conditions to teach the content. Indeed, in the very early stages of instructional system design, the designer is encouraged to think of the educational problem he is confronted with in terms of possible alternate solutions. The designer, by following Step Two, examines the system as a whole and tests the possible alternatives against what he now knows about the system. In Step Three, the designer relates the whole to each component objective. It should be emphasized that what has been identified in Steps Two and Three

are tentative characteristics and relationships and subject to revision as the designer goes on with the task of specifying instructional conditions. Surely, what will be discovered in the next seven steps may have important ramifications as to the designers choice and the type of instructional system he wishes to build. The purpose of these seven steps is to give the designer further information upon which to build the instructional system and assess its appropriateness. Note that the next seven steps relate most directly to individual enabling objectives. Each is considered separately at first and then related, all in the context of the tentative identification of the general characteristics of the instructional system.

F. Step Four: Specify Type of Learning Function Represented for each En Route Competency[4]

It would seem that an important basis for matching instructional objectives with instructional conditions would be to draw as much as possible upon the large amount of information of human learning that has accumulated through the years (within limits described in Appendix B, and discussed in a later section). If each objective could be classified into a category which is homogeneous with respect to the conditions fostering learning of that

[4] This step, as well as the next six steps, all involve the specification of instructional conditions for individual enabling objectives. If the designer is considering just one competency, he would clearly go through the seven steps successively. But what if he were working with 5 or 10 or even 20 objectives? Should he take one objective and complete all seven steps before moving on to the next objective, or should he complete the first step, for example, for all objectives before moving on to the next step? It would seem that both procedures would possess advantages, but this writer suspects that the latter method of completing each step for all objectives (or as many as practical) would have the advantage of allowing the designer to better integrate the instruction of all objectives. There is certainly litte empirical evidence to support this hunch, however.

ERIC

type, the instructor would have at his command a very powerful tool with which to specify instructional conditions. Eckstrand (1964) has suggested that such a taxonomy would provide "a system for organizing existing information in a way which would facilitate its application to particular training problems" as well as provide "a most useful tool in determining deficiencies in our knowledge and thereby serve to guide future research." Several attempts have been made in specifying this learning taxonomy (e.g., Cotterman, 1959; Gagne, 1965; Miller, 1963; Altman, 1966).

A taxonomy is given below that is but one of many possible taxonomies that may be appropriate. Indeed, any of the taxonomies given by the authors mentioned above could be used. The crucial test for any taxonomy is whether or not the designer can specify with any degree of preciseness the unique characteristics of the conditions of learning for each category within the taxonomy. The taxonomy given below was adapted from Altman (1966). Briefly, the taxonomy is as follows:

(1) Chaining or rote sequencing

(2) Discriminating or identifying

(3) Coding

(4) Classifying

(5) Discrete estimating

(6) Continuous estimating

(7) Logical manipulation

(8) Rule-using

(9) Decision-making

(10) Problem-solving

As each learning function is identified below, keep in mind that the purpose of this taxonomy is to provide a system for specifying as best we can the different set of conditions necessary for the learning of each of the objectives that is previously identified by the instructor. It is impossible in this paper to discuss each of the learning functions in detail, or to expect the reader to master each type efficiently so as to recognize examples or to generate specific examples. Instructors faced with the task will usually secure the services of an instructional technologist who is well acquainted with the various types of learning and the conditions necessary to bring the learning about. For purposes of this paper, it seems appropriate to at least look at the more important learning functions, so that you will have a better understanding of the processes involved in designing an instructional system.

(1) Chaining or rote sequencing. Chaining or rote sequencing is the following of a pre-specified order of verbal motor acts. It is the connecting together in a sequence two or more previously learned stimulus-response connections. Language is filled with these chains of verbal sequences such as "scotch and soda" and "over and above." In the psycho-motor domain, checking all gauges in an aircraft before taking off is an example. Other examples include: wetting the end of a thread before inserting it through a needle, checking all receipts before entering them on the books, introducing people before starting a conference (cf. Altman, 1966). Chaining or rote sequencing includes those categories described by Gagne (1965) as Type 3 (chaining) and Type 4 (verbal association).

(2) Discriminating or identifying. Discriminating or identifying is the type of learning that a teacher undertakes in order to be able to call

each of her students by his or her correct name. It is the perceiving of the appearance of an object that is distinct from other objects. This learning function applies when students learn to distinguish plants, animals, cars and so forth. In the psychomotor realm, it refers to the identifications of various parts of a motor or the identification of parts of a clock. Other examples include discriminating the English from the Greek alphabet. This learning function is identical with Gagne's Type 5 (multiple discrimination) category.

(3) Coding. Coding is defined as the translation of the perceived stimulus into another form, locus, or language (cf. Altman, 1966). Coding does not necessarily involve the application of the sequence of logical rules. It is a special case of Gagne's Type 5 (multiple discrimination) learning category. Examples include the writing of the name of a part found to be faulty, recording a voltage measurement, and rewriting ten as 10.

(4) Classifying. Classifying is a learning function that is critically dependent on internal neutral processes of representation that are served by language. It is perceiving an object as representative of a particular class, where the objective characteristics of objects within the class may be widely dissimilar. Classifying, similar to Gagne's Type 6 (concept-learning) category, requires a learner to respond to stimulation in terms of an abstracted property like color, shape, or size, as opposed to concrete physical properties like specific wave lengths or particular intensities. Examples include: distinguishing red and green lights, differentiating acids from bases.

(5) Discrete estimating. Discrete estimating refers to the perceiving of distance, size, or rate with discrete recording or responding. It

ERIC

is identical with Brigg's continuous case of perceptual motor skills learning. Examples include the keeping of a moving vehicle on the road, handwriting and maintaining a desired distance from a dancing partner.

(7) Logical manipulation. Altman (1966) defines logical manipulation as the application of formal rules of logic or computation of an input as a basis for determining the appropriate output. It is a case of Gagne's Type 7 (principle-learning) category. Examples include the working out of the horsepower of an engine from standard formulas, scaling a map, and computing one's personal income tax.

(8) Rule-using. Rule-using is the executing of a course of action including one or more contingencies by the application of a rule or principle. Again, it is an example of Gagne's Type 7 learning category. It should be noted that rules such as "salt is composed of the elements Na and Cl" may be learned as verbal chains by memorization. If this were the case, terminal performance would be limited to having the learner repeat the statement. This is not what is meant by rule-using. When a student uses a rule, he is able to use it in a variety of instances. Gagne refers to this process of generalization as lateral transfer.

(9) Decision-making. Decision-making is the choosing from a field of alternative actions in a probabilistic situation, one particular action. Decision-making also includes the following of optimum strategies in non-rote behavioral sequencing. Again, it is a case of Gagne's Type 7 (principle-learning) category. Examples include: selecting the type of engine for a new product, and deciding on what type of statistical analysis to use in a given instance.

(10) Problem-solving. Altman defines problem-solving as the resolving of courses of action where routine application of rules for logical manipulation and decision-making would be inadequate for making the optimum choice. It is a case of Gagne's Type 8 (problem-solving) category. Problem-solving implies that the learner integrates and adapts existing principles into novel, specialized, or higher order rules. The learner combines old principles into new ones. He thinks out a new principle that combines old principles into new ones. He thinks out a new principle that combines previously learned old ones. Examples include: the developing of a new design for a new type of engine, the developing of a new statistical analysis technique, and the developing of an improved approach to customer service.

G. Step Five: Identify Instructional Strategies that Provide General Conditions of Learning for Each En Route Competency

After identifying the type of learning function represented by each enabling or en route objective, the task is to devel p an instructional environment which will transform learners into graduates who can perform at the specified levels after instruction is terminated. To accomplish this, instructional sequences should be designed that reflect the instructional strategies that provide the general conditions of learning for each en route objective. In other words, whatever is known about the means or methods of obtaining the most efficient and effective learning possible in each category must be brought to bear upon the problem the instructor faces in such a way as to increase the probability of his designing the optimal instructional system.

As stated above, the instructor is not expected to be able to identify these strategies without help from an adequately trained individual. However, to illustrate what is meant by this step, a summary of the conditions

ERIC

for principle-learning as cited by Gagne (1965) are listed below. Recall that principle-learning is the category for the learning functions identified as logical manipulation and rule-using. If the reader wishes to study the conditions of principle-learning in greater detail, he is directed to the book, The Conditions of Learning (Gagne, 1965).

> "The requirements for instruction of principles whether practiced by a teacher, a film, or a textbook (are):
>
> Step One: Inform the learner about the form of the performance to be expected when learning is completed.
>
> Step Two: Question the learner in a way that requires the reinstatement (recall) of the previously learned concepts that make up the principle.
>
> Step Three: Use verbal statements (cues) that will lead the learner to put the principle together as a chain of concepts in the proper order.
>
> Step Four: By means of a question, ask the learner to 'demonstrate' one or more concrete instances of the principle.
>
> Step Five: (Optional, but useful for later instruction). By suitable question require the learner to make a verbal statement of the principle." (Gagne, 1965, p. 149)

Often, instructors do not teach principles in isolation as described above. Instead, they teach a number of principles often without ever stopping for a breath in between. Consider, for example, the principle that is stated in Step One above: "Inform the learner about the form of the performance to be expected when learning is completed." Rarely would an instructor question the learner to make sure that he knew the concepts of "inform," "learner," "form," "performance," and so forth. Indeed, the instructor usually does not even assume that the learner would not know these concepts. But this is precisely where misunderstanding occurs in instruction. If the student does not understand even one concept included

in a principle, the probability of his successfully understanding that principle (demonstrating one or more concrete instances of the principle) is lower than when he knows that concept. This would imply that lecture method of instruction is woefully inadequate in setting the conditions of principle-learning.

Examination of the conditions necessary to bring about the learning of principles has been presented simply as an example of what an instructor, together with an instructional specialist, might go through in specifying instructional conditions. The reader is encouraged to study Gagne (1965) for a detailed discussion of the conditions of learning for most of the types discussed. In addition, the document by Altman summarizes error classes in each of the ten learning functions identified above.

H. Step Six: Specify the Learner Response(s) for Each En Route Competency

Why specify the learner response before the stimulus situation? Some individuals might say, "I've always specified what I'm going to say in my instruction. I don't particularly care what the learner response is. My calling in life is to teach something. So why all this emphasis on response?" Although this statement is probably an overstatement, it does point up a point of view commonly taken by instructors. Witness the detailed lecture notes collected over a period of years and presented to learners time after time after time with little or no revision. Witness the innumerable books of readings and textbooks that reside on your own bookshelf. These "encyclopedias of knowledge" are all aimed at presenting information. Rarely do they elicit responses on the part of the learner that help him to practice his newly learned knowledge or skill. So in

this chapter, at the risk of going too far in the opposite direction, the tack is taken that says in effect, "Let's spend some time thinking about what the learner is going to be doing during instruction (other than acting as a sponge) and what the learner will be doing after instruction." For example, if a learner is given textual material, what responses are expected of him while he reads this material? Research by Rothkopt (1963) has shown the value of the insertion of questions in textual material. These questions lead to some learner response, and it is this type of response that we are interested in now. As another example, Maccoby and Sheffield (1961), in a review of research on combining practice with demonstration in teaching complex sequences, note that pure demonstration rarely if ever suffices to provide complete learning. Demonstration is usually considered as only part of the teaching process where ability to perform is the ultimate criterion of learning. The step of specifying learner responses for each en route competency is simply the specification of responses such as mentioned above. In specifying learner responses, two things are noted: (1) content, and (2) operation.

(1) Content. Content is specified by taking a look at the competency you wish the student to perform. For example, if a competency involves the student listing seven causes of the Civil War, the response content is going to involve those causes. If you wish the student to construct an equilateral polygon, using pencil and paper and a straightedge and compass, then the content involves the drawing of a particular arc or the measuring of a particular distance. Needless to say, in rather complex units of instruction, many learner responses may be elicited.

ERIC

(2) Operation: Form. In addition to specifying content of the response, the instructional designer should specify the form of the response. There are many ways to classify responses with respect to form. The taxonomy presented at this point is useful only as it helps the reader think in terms of the ways in which learners can be actively engaged in the learning process.

If practice of knowledge (e.g., memorizing steps and procedures) is desired, what form should it take?

Overt	vs.	Covert
(An observable response)		(An unobservable response)
Button pushing, Verbal response, etc.		Instructor: "Answer this question in your mind." Learner: "Mental" response
Selective		Motor
Making a multiple-choice response on an answer sheet.		Pushing a button.

Vocal

Saying or writing, "The answer is ten" or "3 times 3 is 9."

Needless to say, the designer is free to use any of these forms if they are consistent with empirical evidence.

If practice of performance (e.g., application of knowledge, analysis, synthesis, psychomotor skills) is desired, the designer has several options open to him:

(1) A real-life response to real-life stimuli (doing the task in the operational setting);

(2) An enacted response (doing the task in a non-real life setting);

(3) An iconic response (drawing what would be done in a real-life setting);

(4) An analogue response (giving a non-corresponding response in a non-real life setting); or

(5) A symbolic response (saying, writing, or choosing among given alternatives what would be done in non-real life setting).

Let us examine these in a bit more detail.

A real-life response must be made to a real-life stimulus. For example, the inserting of a cartridge into a tape recorder would be considered a real-life response. If the stimulus is simulated, that is, if the tape recorder is a simulation of a real-life recorder, then the response should be considered as representative of real life since it is not made to a real-life stimulus (the actual tape recorder) and transfer is involved from the instructional situation to reality. Confusion will arise if this point is not clear. The basis for labeling a response "enactive," "iconic," "analogue," or "symbolic" is not whether or not the real-life response involved doing, drawing, writing, etc. The basis of labeling is whether or not the response is representative of real life in the instructional context, and how real life is represented. A real-life response may take any form, but we are not distinguishing these forms now. The above mentioned labels of enacted, iconic, analogue, and symbolic only serve to identify the type of representation of the real-life response, whatever form it may take. If a simulation, the learner may do the real, live activity (enactive response), draw the activity (iconic response), tell about it (symbolic response), or do something that is an analogue to the real, live behavior (analogue response). It makes no difference in this taxonomy what kind of real-life response it is that is represented.

With this in mind, we can examine each of the four types of response representation in greater detail. The enactive response is essentially doing

what is done in real life with the exception that it is elicited by a <u>simulated</u> stimulus <u>situation</u>. As an example, a student may insert a cartridge into a simulated tape recorder that is a mock-up of the real thing. Or, in a learning game, a student may play the role of a senator. He may make speeches, he may lobby, and attend conferences. In either case, this enactive behavior is elicited, not by the real life legislature of the state or by the tape recorder, but by a series of rules and instructions for playing the game that simulate real life, or, in the other case, by a mock-up of a tape recorder. The responses are enactive since he does what is usually done in real life. Enactive responses usually are characterized by an interaction with the stimulus, and are most realistic in terms of the four categories mentioned above. Enactive responses require the least interpretation by observers witnessing the behavior.

The <u>iconic response</u> is essentially drawing what would be done in real life. In designing an instructional simulation, this type of response would probably not be used to any great extent. Yet, it definitely represents a class of responses that should be recognized. As an example, a student might draw how he might insert a cartridge into a tape recorder. This would be an iconic response.

The <u>analogue response</u> is elicited by an analogue stimulus. Although we have not talked about the representation of stimuli by analogue (see page 56), let it suffice to say that an analogue stimulus is one in which one property is used to represent another. Non-correspondence is the rule here. For example, the flow of electricity may be represented in analogue fashion using water flowing through a pipe. So, a response that is used to

represent another in analogue fashion will be termed an analogue response. This requires that the learner must transfer from the analogue response in the instructional situation to a real-life response. Theoretically, transfer will be more difficult than when transferring from an enactive response to a real-life response. Taking the example of water flowing through a pipe to represent electrical current, we consider a faucet to represent a switch. In real life, the response would be an up or down motion with a lever, while in the simulated situation, the response would probably be a turning motion with the hand. Not that in both instances the response is not thought of separately from the stimulus, since in psychological terms the stimulus elicits some response. This points up the fact that in the step of specifying learner responses to each en route competency, the stimulus situation which elicits the response is not forgotten. In fact, in some cases the stimulus situation will be specified along with the response. This is discussed in greater detail below.

The last type of response representation that we will consider is the symbolic response. Saying what would be done in a given situation, writing what would be done, or choosing among the given alternatives, are all types of symbolic responses. Actually, symbolic responses could be considered a form of analogue response. An example of the symbolic response is classroom simulation training (Twelker, 1967) where problematic classroom episodes appear on a screen in front of the learner, and the learner is expected to respond to them, telling what he would do in the problem situation. Essentially, the student would "armchair" a real-life response. If he chose to act out the response, it would be classified as an enactive

response since he was doing what would be done in real life except to a simulated stimulus. Conceivably, the student could even choose to draw his responses,; in this case it would be termed an iconic response.

(3) Operation: Strategy. We have talked about the content and form of the learner response. Another operational specification that must be considered is the strategy to be used stated in terms of learning principles. Needless to say, when the specification of content and operation remains hazy. recognition of available teaching strategies expressed in terms of learning principles (or intuition as a last resort) may provide clarification of content and what must be done operationally. The instructional technologist at this point should help the substantive person determine what type of operational response is most appropriate for his objective, based on whatever evidence is available in the learning research literature. In one example, a designer might be called upon to specify whether learner responses should be interspersed at the end of sub-units of a film, obtained only at the end of the film, or required at the end of some predetermined sub-unit based either on natural units or units stated in terms of the amount of demonstration material which could be assimulated and translated immediately into adequate or perfect performance (cf., Weiss, Maccoby, and Sheffield, 1961). Actually, a great deal of rather practical research has been done relating to this problem, and it would seem to make some sense that such knowledge should be brought to bear upon the instructional system being designed, if at all appropriate.

(4) Operation: Machine requirements. A final factor to consider in specifying the learner response is whether or not machine components are

required. This decision will relate directly to the specification of the operations involved. For example, if the practice of performance is desired, and button-pushing or other motor acts are specified, machine components may be required. These components must be outlined and anticipated. Consider another example. If the instructional engineer specifies practice of performance as essential and enacted responses (doing a task in a non-real life setting) are indicated, machine components may be more advantageous although not mandatory. For example, consider the teaching of certain decision-making skills, discrimination skills, and problem solving skills that are inherent in an electronic trouble shooting task. Kristy (1967) describes a training environment that staggers the imagination of those who spend most of their time in front of a chalkboard. The Simutech Trainer was conceived to train and reinforce electronics technicians in a manner that provided both teaching capabilities and realistic on-the-job experience. Specification of the system which Kristy describes calls for computer-controlled programed-learning linked with a simulation of an electronic system which a student is responsible for "maintaining." This system, linked with the computer, senses and responds to the student making his decisions. What if the Air Force had only a few dollars instead of $130 million needed to develop this system? An alternative non-machine system could be worked out. In fact one has been developed that is a low cost system that exercises many of the same skills, but without the use of machine components. A class of auto-instructional devices termed Trainer-Tester Simulators illustrate this non-machine system. Pictures are used in place of actual or simulated electronic hardware. Specially designed

worksheets on which the trainee can carry out maintenance and trouble-shooting functions are used in lieu of machine components. Data are concealed by a silver overlay which is quickly removed by using a pencil eraser. The data uncovered reflects deviations from normal operation with the equipment operating under an indicated trouble symptom. The trainee must be familiar with the equipment to select those check points from which he wishes to obtain data so he may arrive at a correct solution. Indiscriminate erasures indicate that the trainee has analyzed the problem incorrectly. The Trainer Tester Simulator as well as the Simutech Trainer have been discussed in detail by Twelker (1968; 1969).

I. Step 7: Specify the Stimulus Situation for each En route Competency.

The discussion below will again focus on two important factors: 1) content and 2) operation. We will first examine content.

1. Content. Recall that the stimulus situation represents various types of instructional messages or acts presented to the learner. The stimulus may be rather short and to the point, for example, "Mark an 'X' in the space to the right of the answer" to rather lengthy discourses. Several functions may be served by these types of instructional messages. These are listed below:

1) Orient the learner to the behavior desired (sensitize the task);
2) Shape behavior;
3) Assess whether the learning has occured.

Orientation. How many times do instructors ask learners to dive headlong into an instructional unit without first informing them of the end behavior that is desired? Often, instructors assume that the student is all

set and ready to go at the beginning of the course or a unit of instruction, and little is done in the way of sensitization. It seems appropriate to list at least six functions that are served by sensitization to the task:

1) Student interest is increased;

2) Student acceptance and commitment to the task is increased;

3) The student is oriented to policies, rules, procedures, purposes, and learning resources;

4) A topic is introduced, it's importance is indicated, and an overview of the scope is presented;

5) Directions on procedures for use in subsequent learning activities are given;

6) Recall or prerequisite competencies are stimulated.

In summary, orienting the student can serve a useful purpose as a prelude to the shaping of behavior. It lets the student in on what the instructor feels is important, and if nothing else, serves to provide the student with an affective experience that allows him to say, in effect, "This instruction is worthwhile and I'm ready to give my full attention to it."

Shaping. It goes without saying that the bulk of instruction is usually concerned with this category of instructional message. Several purposes are served by instructional shaping messages or acts;

1) Knowledge is presented;

2) Information about skills, e.g., the process of arriving at a solution to a problem, the process of forming a psychomotor skill, etc., is provided;

3) The learner is guided to subsequent steps in the performance of any response;

4) Closure to or a solution to a problem is provided;

5) Information required to solve a problem or make responses is provided;

ERIC

6) A review or summary of information is provided.

7) Practice on a task is provided.

<u>Assessment</u>. Assessment gives the instructor an opportunity to monitor the learner's progress. In some cases, it also gives the learner an opportunity to employ that which is already known for purposes of practice or motivation. For example, many classrooms are being equipped with what are called, "student responses systems" that gives learners an opportunity to answer, by means of individual student responders, multiple choice pre-planned questions presented through an audio-visual system. Such an installation serving 500 students has been in operation since the summer of 1964 at Chicago Teachers College North. Similar installations have been installed at Florida Atlantic University and Foothill College in California. Some of these systems are coupled with automatic data handling systems so that the instructor can collect information on each student and assess his performance during and after the instructional presentation. In this way, the instructor can modify the presentation on the basis of total student response. Of course, assessment need not rely on such complex systems as that mentioned above. But it should be recognized that assessment should be carried out systematically and frequently throughout instruction, not only to benefit students but to benefit the instructor as well.

Now that the functions served by the stimulus situation have been discussed, let us now turn our attention to specific strategies and tactics to accomplish the various functions mentioned above. Two general strategies are: (1) exposure to information; (2) precipitation of performance. That is, the instructor may wish to orient the student to the task by giving him

some information or showing a demonstration and precipitating some performance such as questions about the task. In the same way, the instructor may use these two strategies to shape behavior. And again, the instructor may use these strategies to assess performance by giving the learner some information and then asking a question or two. The specific tactics and instructor moves that may be used are discussed in detail below.

Exposure to Information. Two tactics may be used by the instructor for exposure to information. They are listed below with examples.

1) Exposition: Exposition is a generic term that refers to a statement by the instructor that may be verbal or presented in writing. Visual types of communications are not used in pure exposition. Several specific aspects of exposition may be listed:

   a) Description--a detailed statement, lecture, reading, etc.

   b) Conceptualization--drawing conclusions from statements (synthesis)

   c) Explanation--explaining meaning in other terms of language (analysis)

   d) Evaluation--making judgments about the value of that presented.

2) Illustration and Demonstration: Illustration refers to an iconic representation but is intended to supplement exposition. For example, a person may talk about an imbricate flower bud, and use a verbal description slowly. On the other hand, he illustrates an imbricate flower bed with a picture. Demonstration serves the same function as illustration, but is given with the help of specimens or by experiments. Concrete representation may be used in demonstration and presents in a clear way that which is being talked about in exposition. The functioning of an automotive engine by means of a mock-up is an example of demonstration.

Precipitation of Performance. Generally, two tactics may be used in the precipitation of performance: inquiry and demand.

1) Inquiry. When an instructor chooses to use the inquiry tactic to precipitate performance, he may give one of six types of questions

(cf., Bloom et al., 1956). Each type of question with examples is shown below.

a) Knowledge (recall) question: Questions of this sort ask for "the remembering, either by recognition or recall, of ideas, materials, or phenomenon." "Generally, the recall of any specific bits of information, whether concrete or abstract, belongs in this category. For example, "When did Columbus discover America?"; "What is a synapse?"; "What is the population of Oregon?"

b) Comprehension question: Comprehension includes objectives which represent an understnading of the literal message contained in a communication. Comprehension demands translation, interpretation, and extrapolation. Translation refers to an individual putting a communication into other language, into other terms or into another form of communication. For example, "Can you give me an illustration of a metaphor?" Interpretation involves dealing with a communication as a configuration of ideas the comprehension of which may require a reordering of the ideas into a new configuration. Examples of interpretation include "What does the boy mean by 'fuzz'?" Extrapolation not only includes the skills of translation and interpretation, but also the skills of extending trends or tendencies beyond given data and the findings of documents to determine implications, consequences, etc., which are in accordance with the conditions as described in the original communication. For example, "What follows three, five, seven?" is an extrapolation question.

c) Application question: An application causes the learner to apply abstractions to explain concrete situations. The abstractions may be in the form of principles, models, or propositions. For example, "An electric iron has been used for some time and the plug contacts have become burned, thus introducing additional resistance. (The iron is a 110 volt, 1000 watt model). How will this affect the amount of heat which the iron produces?"

d) Analysis question: Analysis is the breaking down of a whole into subsequent parts. When relevant information is separated into its elements or parts so that the relations between the parts is explicit, analysis has taken place. An analysis question is a relational question. For example, "Why do automobile manufacturers produce only three body sizes for the many models produced each year?"

e) Synthesis question: Synthesis is the combining together of components in some organized way as to constitute a whole. Synthesis is building up, as compared with tearing down. An

example, "Why the decisions of Robert E. Lee to commit treason and David Farragut to remain loyal are both admired by Americans today?"

f) Evaluation question: An evaluation question calls for the making of judgments for some purpose about the value of ideas, work, solutions, etc. It involves the learner in using criteria as well as standards for appraising the extent to which particulars are accurate, effective, or satisfying.

2) Demand. If the instructor chooses to elicit or to precipitate performance by direction, he has four tactics that may be used. These are listed below with examples.

a) Demand for performance with suggestion, e.g., "I think that it would be a good idea to finish the exercises on page 1."

b) Straightforward demand for performance accompanied by a cushion, e.g., "It may seem like hard work, but I'd like for you to complete the exercises on page 10."

c) Straightforward demand for performance, accompanied by an explanation, e.g., "If it hadn't been for the fact that we didn't get yesterday's work done, the assignment wouldn't be given. However, for tomorrow, complete the exercise on page 10."

d) Straightforward demand for performance without qualification, e.g., "Complete the exercise on page 10."

2. Operation. Now that we have briefly examined the content of the stimulus situation, let us take a look at operation, that is, how the content is presented. Four things seem important here:

1) the form of each message;

2) anticipation of interactions between learners and/or teachers required;

3) strategy used in terms of learning principles;

4) machine requirements.

Let's take these one at a time.

a. Form of each message. There is no simple way of classifying stimuli in instructional messages as to form. The categories mentioned here are

better thought of as possible ways in which a message might be presented, and do not necessarily represent mutually exclusive categories. The message could be classified into more than one category. Note that as these forms are discussed, each of the types of messages discussed above should be reviewed in light of the present discussion to assure yourself that the suggestions given here are valid.

A stimulus that is given to the learner might be real life and as such, nothing is simulated. In real life a stimulus that is presented to the learner may involve one or more senses: vision, hearing, touch, taste, or smell. Most often, only the audio and visual modes are used in instruction, and our instruction will be limited to these modes.

If the instructional designer does not wish to represent real life stimuli, he may then simulate real life. Real life may be simulated in a variety of ways. The stimulus may be a concrete representation. A model of an apple that is made up of wax and possesses many of the attributes of an apple, such as shape, color, size, and texture, may be termed a concrete representation. It is clear that the concrete representation may very closely approximate the real-life apple. It may even use real seeds and a real stem. But in some way reality has been altered by omission of the representation of one or more elements.

There are degrees of concrete representation. At one extreme is a nearly perfect representation of an apple as noted above. But the other extreme may be a gross caricature of a real apple--a red, hollow shell that emphasizes certain features for the sake of instruction. In this case, it might be called a "mock-up" (in media terms). At the middle of the continuum may be a three dimensional representation of the apple that is rather

ERIC

similar to the real life object. The important thing to know is that concrete representation allows the learner to experience directly the phenomena. The learner may see, and in some cases hear, touch, taste, and smell the simulation. Stimuli that may be classified as concrete representations are three dimensional, thus allowing the learners to interct with the stimulus. Edling (1966) suggests that "all senses can be employed to provide cues to the learner" in the case of three-dimensional stimuli.

Besides direct experience with objects, the learner may also have direct experience with people. Indeed this experience may serve the functions of the stimuli situations described above. Direct experience with people might include discussion debates, politics, tableaus, pantomimes, plays, or pageants, simulations, and authentic situations.

The stimulus may be an <u>iconic representation</u>. Iconic representation, as described by Bruner (1966) "depends upon visual or other sensory organizations and upon the use of summarizing images."

> "Iconic representation is principally governed by principles of perceptual organization and by the economical transformations in perceptual organization that Attneave has described--techniques for filling in, completing, extrapolating." (Bruner, 1966, p.11)

Iconic representation may deal with pictures--with sense of vision. Edling (1966) notes that representations that are iconic are "objective" because elements in the representation (the picture or drawing) correspond to specific elements in the reality. The key in iconic representation is <u>correspondence</u>. The elements in the model (in the general sense of the word) "contain cues that make it possible for a learner to associate an object with visual representation of the object <u>without</u> prior association with the object itself" (Edling, 1966, p. 38).

Edling (1966, p. 39) presents a list of "objective" visual stimuli that is helpful in realizing the wide range of iconic representations.

Motion pictures, with illusions of 3-D in color
Motion picutres, with illusion of 3-D minus color
Motion picutes, 2-D, in color
Motion pictues, 2-D, minus color
Still pictures with illusion of 3-D, in color
Still pictures with illusion of 3-D, minus color
Still pictures, two dimensional, in color
Still pictures, two dimensional, minus color
Painting (realistic), in color
Photograph of painting (realistic) minus color
Sketch (with shading)
Representational color cartoon (with animation)
Representational color cartoon (without animation)
Representational cartoon (minus color)

Note that as the iconic representation moves from the motion picture to the still photograph and to the cartoon, the number of cues that are available to the learner about real life decrease.

Visual stimuli are not the only stimuli that may be iconically represented. Sounds may also be classified as iconic since they depend on perceptual organization. Sounds are "objective" in that elements in the representation (the recording of a bullfrog, for example) corresponds to specific elements in real life. Here again, there are degrees of representativeness. A stereophonic recording of the mating call of the bullfrog has more cues that correspond to reality than a recorded sound effect that sounds <u>like</u> a bullfrog but is produced by other means, for example.

So far, we have considered the representation of stimuli by concrete and iconic means. A third type of stimulus representation is <u>analogue representation</u>. By that, we mean most simply that the property of correspondence changes to non-correspondence. One property is used to represent another. Whereas iconic representation <u>models</u> relevant properties of real

ERIC

life through the properties themselves, analogue representation models relevant properties of real life through <u>other</u> properties, so that a code or legend is required in order to learn. "This condition requires that a learner have associations with the visual stimulus and the object it represents if the visual stimuli is to be associated with the object" (Edling, 1966, p. 40). The learner must know the code or legend in order to associate the representation with real life. For example, the flow of electricity may be represented in analogue fashion using water flowing through a pipe. The well-known cartoon figure, Uncle Sam, is an analogue represenation, since it stands for a country. In both instances, the learner must be told what the model stands for in order to learn. Examples of stimuli that are represented by analogue include:

Symbolic cartoons
Diagrams
Maps
Charts
Graphs

A final type of stimulus representation that shall be considered is <u>symbolic representation</u>. For example, numbers and words are symbolic models of real life. In one sense, it is a form of analogue representation in that the property of non-correspondence is still operative, and one property (e.g., a word, "Plato") is used to represent another property (the man, Plato). Yet, in terms of a continuum of realism, it seems that a large gap exists between maps and similar models, and symbolic representations, such as words. To this end, representation of reality by symbols is best thought of as a separate category.

b. Anticipation of interactions required. In addition to specifying each message, the designer should take some time to outline the interactions that may be involved between learners and the teacher. These interactions are required to support the presentation of the instructional messages of the types outlined above (precipitation of performance, exposure to information, and centralization to task). These interactions should be looked upon as supporting messages.

c. Strategy used in terms of learning principles. Again, at this point, the services of an instructional technologist are required to help the substantive person determine what type of stimulus situation is most appropriate for his objective, based upon whatever evidence is available from the learning psychologist. Certainly, the work of Travers (1964) would be of interest to the instructional engineer concerned with efficient presentation mode. Also, mention made in the introductory remarks of the work of Saul et al (1954) would be pertinent to the specification of stimulus situations involving graphic training aids. Again, the great amount of research which has been done on the particular problem of stimulus presentation should be brought to bear upon the instructional system being designed.

d. Machine components required. Given specifications of the stimulus situations as outlined above, it is an easy task to consider the machine components might be specified, each with their advantage and disadvantages. Two types of machines may fit the general requirements of the system, but each might exhibit different limitations that should be considered. Often,

ERIC

trade-offs between one or another factor will determine what machine component will be used. Needless to say, the consideration of machine components is closely related to the particular presentation form used for the message. Yet, it should be emphasized that these considerataions must be separated so that the consideration of machine components is secondary to the consideration of presentation forms. 2.2

J. Step Eight: Specify Feedback for each Instructional Event.

Feedback must be specified whenver responses are called for on the part of the learner. It makes little sense to elicit a response from a learner without giving him some indication of the correctness of his response. Again, it will be helpful to look at the specification of feedback in terms of content as well as operation. Let us discuss the content of feedback first.

1. <u>Content</u>. The instructional message contained in the feedback may be of four types:

1) information presented to the learner about the appropriateness of his response;

2) information presented to the learner that is primarily intended for an affective response from the learner;

3) information about the actual correct response;

4) information about the rationale or reasons for the correct response;

a. <u>Information presented to learner about the appropriateness of response</u>. Two functions may be served by this type of feedback. First, it serves to increase the probability of a similar behavior in the future. For example, if a learner is asked a question and he responds with an answer that he is not sure of, and the teacher says "right", the probability of his responding in a similar fashion on the next occasion is increased. There is

another function served by this type of information, that of decreasing the probability of a similar behavior in the future. For example in the same situation of a teacher asking a question, if the response "wrong" is given to an incorrect answer, the chances of his answering the question in a similar fashion on the next occasion is decreased.

Tactics that maybe used to present this information are:

1) Signal. The teacher may use gestures such as a smile, a frown, a nodding of the head, and other non-verbal types of communication to inform the learner about the appropriateness of the response.

2) Word. The instructor may choose to give verbal feedback in the form of statements such as "yes", "no", "correct", "good", "right" and so forth.

3) Objects. Peanuts, M & M's, and Cracker Jacks all qualify as objects which for most students represent information about the appropriateness of their response. These objects increase the probability of similar behavior in the future. Ingenious use of not-so-desirable objects such as candied grasshoppers and other unpleasnat stimuli might serve to inform the learner that his response was less than appropriate.

b. <u>Information presented to learner that is primarily intended for affective response from the learner</u>. Often, the instructor wishes to present information in the form of feedback to the student to simply "keep him going in the system" or perhaps to take a new tack in his direction of learning. The main purpose is to elicit an attitudinal change on the part of the learner, rather than present information. The two functions served by this type of feedback are similar to those listed above: 1) increase probability of a similar behavior in the future. Consider the list of these statements below. These statements are taken from Frase (1963) and used by this writer to reinforce searching behavior in subjects in an experiment on learning by discovery (Twelker, 1967).

Very good! You're starting off very well.

Excellent, keep up the good work.

That's fine, you seem to understand the materials.

Yes, very good.

Very good.

Right! Keep it up.

Right on the button!

Right, and the material is fairly difficult.

These statements all serve primarily to promote an affective response on the part of the learner. In some cases, they serve a dual purpose of informing him about the appropriateness of his response. It should be noted, however, that this is not necessarily the case since searching behavior was being reinforced, and wrong answers as well as right answers were permitted in the experiment. It should also be noted that the same tactics used to present the information are the same as those listed above: 1) signal, 2) words, 3) objectives.

The reader may wish to think of examples of each of these tactics to demonstrate the feasibility or utility of this type of information being presented to the learner.

c. <u>Information about the actual correct response</u>. Again, the functions served and tactics used to present this information are the same as those discussed above. But in this case, the instructor actually gives the correct response rather than simply an indication whether the response was correct or incorrect. Again, the function served and the tactics used are similar to those discussed above.

d. Information about the rationale or reasons for correct response. Finally, the instructor has the option of not only providing the information discussed above, but also providng the reason or rationale for the correct response. Explanations may be given as to why one response is better than another. If the rationale given is lengthy, it will probably be difficult to separate this instructional message from the information given as stimulus situations. Indeed, the feedback given often serves as the stimulus situation for the response. Once again, the functions served and the tactics used are the same as those discussed under the three previous sections.

2. Operation. Now that the content has been discussed, it would be well to take a look at the operations involved in specifying feedback. Note that the operations largely parallel those given for specifying the stimulus situation.

a. Form of feedback message. Since a feedback message parallels closely the stimulus situation presentation, the taxonomy discussed above in terms of stimulus situation applies here.

b. Interactions between learners and/or teachers required. The discussion above on feedback has centered on instructor-presented feedback. Perhaps over-due emphasis has been placed on this particular form of the message. This tends to prevent an instructor from helping the learner to take over the role of being his own reinforcer and corrector. Indeed, the emphasis on learner response advocated in the early sections of this paper are every bit as important in the matter of feedback and reinforcement. An instructor is not a "perennial source of information" and should definitely not interfere

with learner ability in correcting himself. (cf., Bruner, 1966, p. 70). Therefore, interactions between learner and teacher where the emphasis is placed on the learner, is crucial. It has been pointed out that an advantage of simulation exercises is the maximum interpersonal interaction that is utilized. A concomitant of this interaction is the way in which students appear free to talk among themselves and to correct each other. If the teacher does not wish to use this type of learner-learner interaction to present feedback, the situation may be teacher-controlled with an emphasis upon having the student correct himself. In such cases, the instructor may use a message such as, "Do you think that you were correct?" or "What do you think about your performance?" to stimulate the learner in self-corrected behavior.

c. Strategy used in terms of principles of instruction. Feedback is crucial to learning. This fact cannot be argued. But when and how and what type and in what quantity are questions that are not easily determined. Consider these examples.

> Werewolfe is undergoing obedience training. He is learning to come to his owner at command. He has just performed successfully.
>
> Nick, a boy in the primary grades, is learning arithmetic. He is being shown that two sets each of three oranges count to six. Later in a test situation he is asked to differentiate the correct answer 2 x 3 = 7, 5, or 6. He has just answered the question correctly.
>
> Melvin, a fifth grader, is engaged in a unit of instruction on a distributive principle of arithmetic. Instead of simply being told the principle, he is being asked to "discover" it. Currently, he is being asked to distinguish open math sentences that represent the distributive principle from open math sentences that do not. The instructor wishes to reinforce searching behavior on the part of Melvin, that is, behavior which benefits Melvin in problem solving tasks. He wants Melvin to check for patterns in problems, shift strategies to a solution rather than persevering with a single strategy, and so forth. Melvin is currently trying to solve a series of problems, so far without success.

Each of these sitations is unique. One involves a dog, one involves a small boy, and one involves a fifth grader. They are different in terms of the learning functions represented. Each situation represents conditions that require different amounts of motivation. Each situation represents conditions that require different amounts of information in the form of feedback. Each situation is unique in terms of timing of feedback.

In terms of the four types of feedback discussed above, which would be most appropriate for the instructor (or a machine proxy) to present in each case? Certainly, if keeping Werewolfe or Nick or Melvin at the task of learning is crucial, the timing and type of feedback would be different than if motivation were not of primary concern. If keeping Werewolfe learning was of concern, giving him candy would be appropriate when he came to his master at command. But note that this also may function to let the dog in on something which is a source of pride for his master--he performed appropriately. What if a pat on the head had been given in place of candy? Would it function any less as a motivation and a source of information? Consider the case of Melvin. He's searching for patterns in a series of problems, and not getting anywhere fast. The instructor decides to give him some feedback: "Keep it up, Melvin. You're on the right track." Clearly, this may function to keep Melvin at the task. It may or may not tell him that his searching behaviors are appropriate. If it in fact reinforces his searching behavior and his heuristics are inappropriate for the task, the instructor has just possibly slowed down Melvin's progress to the solution. On the other hand, if Melvin was using heuristics that were ap-

ERIC

propriate the instructor in one simple statement has helped him. What if the intructor in the progress of monitoring Melvin decided to present something like:

> "You're doing fine, Melvin, but I think that you should consider a couple of points. First of all, I note that you are not using your pencil to mark the math symbols as you check those patterns. Now if I were you, I'd check off on paper what you're doing in your head. Then another thing you don't seem to be systematic in your checking. You're here one minute, there the next. Stick with one . . ." (and on and on).

Certainly, this is good informative feedback, is it not? Certainly this would tell Melvin about the appropriateness of his heuristics he's using or the rationale behind the correct responses, or would it? It might function to interfere with learning because poor Melvin can't use all that feedback. He may become confused as he attempts to use all of the information the well-meaning instructor presented. (cf., Miller, 1953, p. 93). These illustrations are given simply to point out the fact that the type, quantity, and timing of feedback used during instruction depends a great deal on strategies related to sound principles of instruction. The substantive person might well lean heavily on an instructional technologist at this point to identify these appropriate strategies.

K. Step Nine: Specify the Required or Permissible Context of Instruction for Each Enroute Competency.

If the reader has persisted to this point, it should become evident that the task of specifying instructional conditions <u>for each enabling objective</u> is drawing to a close. Only two steps remain: Specifying the

ERIC

context of instruction or the setting within which both teacher and learner behavior occurs, and arranging all of the instructional events in a sequence to ensure optimal mediational effects from one component to another. Before that last step is taken, the context of instruction must be specified as it in part determines how the instructional events are coupled together into some reasonable order.

In step 3, attention was given to the matter of identifying tentative relationships between, and general characteristics of, the way in which en route objectives would be taught. Using these tentative specifications as guides, the designer is now in a position to specify in greater detail and in a more positive way the context in which instruction will take place.

Four factors should be specified when context is examined.

1. Organization and physical characteristics of the learning space.[4] What organizational arrangement is required or best suited for obtaining the objective being considered? Should the learning space be organized to encourage small study and conference groups, individual study, large group instruction, or tutoring? Is one alternative as attractive as another? When possible, these advantages and disadvantages should be listed to facilitate comparison. Of course, it is realized that when enabling objectives are interrelated, specifying the context of instruction for one relates to the others. In addition, the designer should specify the physical characteristics of the learning space especially as it relates to the learning space organi-

---

[4] This term "learning space" is used rather than "classroom" since the latter carries some connotations that may be limiting, e.g., a classroom "contains" one instructor, thirty students, desks arranged in rows facing the chalkboard, etc.

ERIC

zation discussed above. For example, the following factors should be noted:

a. Space available per learner--how much physical space is necessary for the learner.

b. Type of furniture--e.g., desks, study carrels, laboratory benches, etc.

c. Air conditioning--includes heat, cooling, humidifying and dehumidifying, air purifying, etc.

d. Ventilation

e. Lighting

f. Proximity to other locations that are deemed desirable and necessary to support instruction.

2. Hardware. Hardware may refer to materials and devices that store instructional messages, transmit instructional messages, or store and transmit instructional messages (cf., Hamreus, 1969). Hamreus has listed examples of these three types of hardware:

Storage	Transmission	Storage and Transmission
Overhead Transparency	Overhead Projector	Book
Slide	Slide Projector	Magazines
Filmstrip	Filmstrip Projector	Newspaper
16mm Film	16mm Film Projector	Encyclopedia
8mm Film	8mm Film Projector	Bulletin Board Display
Videotape	Videotape Recorder	Chalk Board Display
	Television Set	
Record	Record Player/Radio	
Audio Tape	Tape Recorder/Radio	
Phone System/ Computer Memory	Teletype System	Chart
Computer Memory	Computer Input/Output System	Poster
		Cartoon
		Flat Picture
Instructional Program	Teaching Machine	
		Globe
		Programed Text

In addition to these types of hardware, specialized training equipment might also be required, such as typewriters for clerical courses, physics laboratory set-ups, simulator trainers, radio receivers and test equipment for electronics training, and so forth.

The designer should summarize the hardware required as specified in Steps 6-8. In some cases, he'll find requirements for monitoring or facilitating learner response, for example, to be closely related to requirements for presenting stimulus situations or feedback. One type of machine might suffice for all requirements. Note that when hardware is considered in this phase of instructional system design, specific types or brands of hardware may be considered. In Steps 6-8, only general requirements of machine components were noted.

3. Number of learners. How many learners are considered an optimal number to instruct? Also, the instructor might wish to specify, where possible minimum limits on the number of learners. This deceptively simple statement is often quite difficult to accomplish as the designer must take into consideration factors such as cost-benefit ratios and so forth. But where possible, the designer should make some effort to determine this factor.

4. Learner grouping. Should learners be grouped homogeneously or heterogeneously? Should learners be randomly assigned to groups or assigned in a purposeful manner?

L. Step Ten: Specify the Appropriate Sequence for each Instructional Component

To this point, the designer has concerned himself with details--for example:

What stimulus situations should be presented?

What form should they take?

What responses should be elicited from the learner?

What form should these responses take?

What kind of feedback should be presented?

What should the learning space look like?

Now the designer must put all these components together into a meaningful set of specifications that reflect their sequence. Only in this way can he insure optimal mediational effects from one component to another. Unless this is done, the instructional system developer hasn't the slightest chance to develop the various units of the system.

In many cases, the designer will rely on a logical analysis of the learning task to help him determine the sequence of instructional events. For example, if a model is used for demonstration, it is clear the sensitization to the task must occur before the model is shown. Likewise, the stimulus situation, in this case a model, must be presented prior to the eliciting of a learner response. In addition, the designer should specify how the components should relate to any recycling, if required. Does exposition have to be repeated on subsequent recycles, or can only a portion of the stimulus situation be repeated, e.g., that necessary to precipitate performance?

ERIC

M. Step Eleven: Specify the Context of Instruction for each Sub-System and Instructional System as a Whole.

Only two tasks remain for the instructional engineer before he may proceed to build the prototype unit based on the specifications he has developed. These two tasks parallel the last two steps discussed. First, he must specify the required or permissible context of instruction, considering all of the relationships and specifications previously identified. Then he must specify the appropriate sequence of all of the instructional units and sub-systems to insure optimal mediation affects from one unit in a sub-system to another. In regard to the step of specifying context of instruction, the designer at this point considers the tentative general characteristics of the instructional system that he identified previously. He also considers his tentative specification of the relationship between the various enabling objectives in the instructional system and the ways in which these objectives might best be taught. The designer must compare the tentative specifications with the results of the various analyses and development of specifications described above. If a "match" is achieved, he may go on and complete his work. If there seems to be large discrepancies between what he thought was appropriate and what turned out to be appropriate, he must take a new look at the overall strategies that he will use in the instructional system.

Again, in specifying the context of instruction the designer should pay attention to such factors as:

1) Organization and physical characteristics of the learning space
2) Hardware
3) Number of learners
4) Learner grouping

These factors were mentioned above and need not be discussed here.

N. Step Twelve: Specify the Appropriate Sequence of Instructional Units and Sub-Systems.

One of the most important things that should be kept in mind at this point is the relationship between various enabling objectives and terminal objectives. As discussed above, the learning of one competency may be prerequisite to the learning of another. This is why stress was made on performing what was called an "objective analysis" so that the designer could determine what competencies were related. This analysis is now referred to by the instructional engineer as he specifies the appropriate sequence of instruction. If certain competencies are independent from other competencies, they may be taught in random order as compared with the situation where competencies must be taught in an appropriate order so that one builds upon the other.

With the completion of this step of specifying sequence, the designer is now in a position to use this information in the actual building of a prototype instructional system. Many refer to this phase as the "fun" part of development. The hard work of specifying content and operation results in a "paper" product--a lengthy list of specifications which may now be transformed into a prototype product. These specifications are in essence tested against empirically derived evidence of whether or not the system is working. The better the specification of instructional components, the more likely the chance of success in the building of the prototype and its subsequent tryout.

V. <u>The Relation Between Research and the "Instructional Systems Approach"</u>

Possibly the educator's greatest handicap in designing an instructional system is in the area of applying in a practical manner findings from learning research. To be blunt, the translation of empirical research to practical application is an exercise in futility.

> Both academic and practically oriented psychologists agree that a very small percentage of findings from learning research is useful, in any direct sense, for the improvement of training or educational practices. Some contend that they are applicable to learning <u>processes</u> that occur in the training environment but admit that these processes have not been adequately identified (Mackie and Christensen, 1967, p.5).

The fault, Mackie and Christensen argue, seems to be both at the door of the experimental psychologist, with his research philosophy, <u>and</u> the potential user, who evidences a reluctance to make the effort to use research findings. The finding of the Mackie and Christensen study was that only a small proportion of the research studies on learning that were reviewed could be applied to instructional technology, and that the translation of learning research to education is possible today on only a limited scale. Reasons for this state of affairs are summarized in Appendix B of this paper. The Summary and Conclusions sections of the report are reproduced in whole to encourage the reader's attention.

What relation do the above tasks--determining what competencies shall be taught, determining in what order they shall be taught, and determining how best they might be taught--have to the development of research hypotheses? How can the procedures for instructional specification help the instructor in deciding what to research in terms of his own instruction?

ERIC

Although the translation from research to practical application may be difficult, the translation of procedures for instructional specification to research hypotheses may be more fruitful. To paraphrase Gustafson (1960, p. 10), neither the determining of what to teach and in what order to teach it, determining how best competencies might be taught, or even the determining of training objectives can significantly advance the state of the art and education. Genuine improvements can evolve only through empirical research conducted in accordance with sound experimental practices. However, the instructional systems approach can help clarify a problem and bring the experience and ingenuity of instructional specialists to bear more directly on them. The process of designing pretested, revised, and evaluated instructional experiences which must be repeated anew for each course in each research project, has the dual effect of: (1) limiting the application of subjectivity to only those areas where the state of the art makes subjectivity necessary; and (2) clarifying the issues so that the state of the art can be brought into the play most effectively.

With the aid of an instructional specialist and a media specialist, the instructor brings to bear on the question of the design of instructional materials all that he knows, and all that is known by others (within personal limitations) about the subject. Only then may he conduct research in a meaningful manner. The manner in which research hypotheses may be derived is taught to every graduate student. Confronted with a problem, he lists all of the alternative hypotheses he can think of, and then tries to systematically disprove each hypothesis, either by citing theory, empirical evidence, or common sense. Those alternative hypotheses that are not discarded in this

manner are examined more closely as probable hypotheses to research. This procedure, in effect, provides a simple tool for the instructor who is grappling with the problem of instructional specification. With the aid of other experts, he attempts to discard as best he can all of the alternative hypotheses he has specified concerning an instructional problem he faces in his courses. Thus it is that research procedures when applied to instructional system design, sensitizes the instructor to those areas of instruction which require his greatest attention.

2.3

VII. A Concluding Statement

From the discussion of the specifications of instructional conditions given above, the reader may be lulled into a false sense of security. So far, the cognitive domain has been given primary emphasis. If the suggestions given above are followed to the letter, and an "excellent" instructional system is produced, does that guarantee that learning will take place? Not at all. Unless the learner's needs and affects toward learning are considered, the best designed instructional system may fall flat on its face. Silber (1968) goes so far as to say that the decisions of which media to use when should be based not on the cognitive learning desired, but on the affective learning desired. Some experienced teachers have suggested that it doesn't matter what kind of instruction is given if the learner's attitude toward instruction is satisfactory. Are these overstatements? The reader is left to judge for himself. But what is clear is that attention must be given to the interaction between emotions and learning. To date, little is available in the research literature to aid the instructional designer in

his task. Research is needed to investigate the use of emotions as facilitators of educational goals. What has been suggested on the preceding pages concerning the use of feedback and reward, and the importance of considering individual differences and even sensitizing an individual to the learning task is but a start in the direction of determining how the instructional events should (will) interact with the learner's affect in a facilitating manner.

This is related to a second point that should be made. The logical analysis of behavior which has been stressed might serve as "blinders" to the designer. Could it be the case that while behavioral objectives state in precise terms competencies expected on the part of the learner, they state only a selective or representative number of behaviors that are components of a more general goal or objective that may be difficult to state in behavioral terms? Is it possible that following the guidelines presented above may result in an instructional system that is limited in some way since only competencies stated in behavioral terms were considered? Consider this example. An instructional systems designer specifies and builds a system to teach principles in social studies. He designs a simulation exercise that involves, among other things, the exchange of money. He tries out his prototype game, and finds that students cheated during the game. In fact, dishonesty and cheating were so prominent that a complaint was registered by some of the more honest souls after the game. The designer in fact held the values of honest dealing with others. Why did he give students an opportunity to rob each other blind during the game? One answer might be that his over-concern with cognitive outcomes blinded him to sev-

ERIC

eral important "unstated" objectives, e.g., learners will make correct change, learners will not steal money, etc. While stated cognitive objectives are being achieved, other unstated objectives are not being achieved. What if the designer did in fact state an objective involving honesty or even truthfulness among the learners? Clearly, his prototype game gave learners an opportunity to cheat and be dishonest. This illustrates the importance of trying out the system and seeing if it works, making revisions on the basis of the tryout, and trying it out again. What has been presented in this paper are guidelines for specifying instructional conditions. They are not guaranteed to result in a perfect instructional system. But they are a beginning--a series of benchmarks for examining alternatives.

References

Altman, J.W. Research on general vocational capabilities, skills and knowledges. Pittsburgh: American Institutes for Research, 1966.

Beaird, J. H. Learner variables and the instructional technologist. In The contribution of behavioral science to instructional technology. Monmouth, Oregon: Teaching Research, Oregon State System of Higher Education, 1969.

Bloom, Benjamin S. (Ed.) Taxonomy of educational objectives: Handbook I, cognitive domain. New York: McKay, 1956.

Briggs, Leslie J. Sequencing of instruction in relation to hierarchies of competence. Pittsburgh: American Institutes for Research, 1968.

Bruner, Jerome Seymour. Toward a theory of instruction. Cambridge, Mass.: Belknap Press of Harvard University, 1966.

Cotterman, T. E. Task classification: An approach to partially ordering information on human learning. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, January, 1959.

Crawford, Jack, & Twelker, Paul A. The design of instructional simulation systems. In P. A. Twelker (Ed.), Instructional simulation: A research development and dissemination activity. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, February, 1969. (U.S. Office of Education Final Report, Project No. 7-I-045)

Easley, J. S., Jr. Remarks on exposition vs. teaching by discovery. (Unpublished report) Urbana, Illinois: University of Illinois Curriculum Laboratory, Working Paper, No. 7, October, 1966.

Eckstrand, G. A. Current status of the technology of training. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Division, 1964.

Edling, Jack V. Media technology and learning processes. In J. W. Loughary (Ed.), Man-machine systems in education. New York: Harper & Row, 1966, Pp. 31-44.

Frase, L. T. The effect of social reinforcers in a programed learning task. Urbana, Illinois: University of Illinois, Training Research Laboratory Technical Report No. 11, 1963.

Gagne, R. M., Mayor, J. R., Garstens, H. C., & Paradise, N. E. Factors in acquiring knowledge of a mathematical task. Psychol. Monogr., 1962, 76, No. 7 (Whole No. 526).

Gagne, Robert M. The conditions of learning. New York: Holt, Rinehart & Winston, Inc., 1965.

Gagne, Robert M. The acquisition of knowledge. Psychol. Rev., 1962, 69, 335-365.

Gagne, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. Psychol. Monogr., 1961, 75, No. 14 (Whole No. 518).

Gustafson, H. W., Honsberger, W. D., & Michelson, S. Determination of task analysis content. In WADD Technical Report 60-593, Uses of task analysis in deriving training and training equipment requirements. Wright-Patterson Air Force Base, Ohio: Wright Air Development Division, 1960.

Hamreus, Dale. The systems approach to instructional development. In Contribution of behavioral science to instructional technology. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1968.

Hamreus, D. G. The domain of media. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1969. (mimeo).

Kersh, B. Y. Directed discovery vs. programed instruction: A test of a theoretical position involving educational technology. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1964. (U.S. Office of Education, NDEA Title VII, Project No. 907)

Kersh, B. Y. Engineering instructional sequences for the mathematics classroom. In Joseph M. Scandura (Ed.), Research in mathematics education. Washington, D.C.: The National Council of Teachers of Mathematics, Inc., 1967.

Krathwohl, David R., Bloom, Benjamin S., & Masia, Bertram B. Taxonomy of educational objectives: Handbook II, affective domain. New York: David McKay Co., 1964.

Kristy, Norton F. The simutech trainer for teachnical and vocational training. In Werner Z. Hirsch, et al., Inventing education for the future. San Francisco: Chandler Publishing Co., 1967, Pp. 114-122.

ERIC Full Text Provided by ERIC

Maccoby, N., & Sheffield, F. D. Combining practice with demonstration in teaching complex sequences: Summary and interpretation. In A. A. Lumsdaine (Ed.), Student response in programmed instruction. Washington, D.C.: National Academy of Sciences, National Research Council, 1961.

Mackie, R. R., & Christensen, P. R. Translation and application of psychological research. Goleta, California: Human Factors Research, Inc., 1967.

Miller, R. B. Handbook on training and training equipment design. Wright-Patterson Air Force Base, Ohio: American Institute for Research, June, 1953.

Miller, E. E. A classification of learning tasks in conventional language. Chicago: University of Illinois, 1963. (AD 419 122)

Rothkopf, E. Z. Some observations on predicting instructional effectiveness by simple inspection. Journal of programed instruction, 1963, 7.

Saul, E. V., et al. A review of the literature pertinent of the design and use of effective graphic training aids. Port Washington, New York: Special Devices Center, 1954.

Schalock H. D. An overview of the teaching research system for the description of behavior in context. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1967.

Schalock H. D., Smith, C., Micek, S. & Vogel, F. The teaching research system for describing teaching behavior in context: A training manual. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1969. (in preparation)

Silber, Kenneth. An experimental study of the degree of affective response illicited by several mediated and non-mediated instructional methods. Unpublished doctoral dissertation, University of Southern California, 1968.

Tallmadge, G. K., & Shearer, J. W. Study of training equipment and individual differences. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1967. (NAVTRADECEN 66 C0043 1. CONTRA CT N 61339 66 C 0043)

Tallmadge, G. Kasten, Shearer, James W. & Greenberg, Anne M. Study of training equipment and individual differences: The effect of subject matter variables. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, May 1968, 61 p. (Technical report NAVTRADEVCEN 67-C-0114-1, AD 671 842)

Travers, Robert M. W. (Ed.) Research and theory related to audio-visual information transmission. Salt Lake City: University of Utah, Bureau of Educational Research, July, 1964.

Twelker, P. A. Two types of teacher-learner interaction in learning by discovery. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1967.

Twelker, Paul A. Simulation: Status of the field. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, October, 1968.

Twelker, Paul A. Designing Simulation Systems. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, February, 1968, (Presented at American Educational Research Association, Los Angeles, California).

Twelker, Paul A. (Ed.) Instructional simulation: A research development and dissemination activity. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1969. (U.S. Office of Education, Final Report No. 7-I-045)

U.S. Army. The development of instructional systems: Procedures manual. Fort Devens, Mass.: Training Center and School, Hdq., U.S. Army Security Agency, Department of the Army, December, 1967.

Weiss, W., Sheffield, F. D., Maccoby, N. Combining practice with demonstration in teaching complex sequences: Serial learning of a geometric-construction task. In A. A. Lumsdaine (Ed.), Student response in programmed instruction. Washington, D.C.: National Academy of Sciences, National Research Council, 1961.

APPENDIX A

	Factor	Type of Technique: Interpersonal Ascendent Simulation	Machine/Media Ascendent Simulation	Non-simulation Game
1.	Control (reproducibility) desired	Questionable	Good	Difficult
2.	Control (planned variation) desired	Questionable	Very Good	Good
3.	Input must be machine mediated e.g., visual stimuli required	not easily adopted	Appropriate	Not easily adopted
4.	Psychomotor and perceptual learning involved	Questionable	Very Good	Questionable
5.	Learners possess low entry skills & limited response repertories that prevent an interpersonal-ascendent simulation from functioning	Limited value	Useful	Questionable
6.	Teacher control over class required	Not so good	Very good	Fair
7.	Simple and inexpensive technique desired	Good	Not easily done	Very good
8.	Interaction between learners required	Very good	Possible but costly	Good
9.	Burdern of simulation experience on learner required	Good	Possible	Possible
10.	Model that emphasizes human interaction involved	Very good	Not easily done	Questionable
11.	Individual differences must be considered but expensive to implement	Good	Not easily done	Good
12.	Feedback must be easily designed into system	Good	Sometimes costly and difficult	Good
13.	Feedback must come from peers	Good	Difficult	Good
14.	Easy development process required	Sometimes difficult	Sometimes difficult	Very good
15.	Insertion into curriculum must be easy	Sometimes difficult	Sometimes difficult	Very good
16.	Must be generally acceptable	Sometimes difficult	Good	Very good
17.	Few learners involved	Limited	Good	Very good
18.	Learning objectives congruent or identical to standard course objectives	Limited	Good	Very good

Table 1. Advantages and Limitations of Simulations and Games

ERIC Full Text Provided by ERIC

APPENDIX B

Translation and Application of Psychological Research

Robert R. Mackie and Paul R. Christensen
Human Factors Research, Inc.
Goleta, California

SUMMARY AND CONCLUSIONS

In an effort to enable the Navy to gain more practical applications from the research it sponsors, a study was conducted to describe the processes involved in translating the results of laboratory research in psychology into forms that would be meaningful and useful in operational settings. Because of its obvious importance, both to the Navy and to society in general, the investigation was concentrated on experimental studies of the learning process. The findings, however, are believed to be generalizable to virtually all areas of psychology.

The study had three principal objectives:

(1) To study the communicative processes that do, or should, intervene between the researcher and the eventual users of research;

(2) To determine the characteristics of research studies, or the practices of research personnel, that affect the likelihood of application;

(3) To describe the attitudes and practices of the logical users of research that appear to affect the likelihood of application.

PROCEDURE

A variety of methods was used to achieve the objectives of the investigation.

1. Selected studies of human learning, most of which had been supported by the Navy, were analyzed in detail and their findings reviewed for possible practical applications to Navy training.

2. The apparent impact of research findings on Navy training personnel and training practices, and the channels through which research findings reach those responsible for developing Navy training philosophy, were studied.

3. An attempt was made to formulate useful principles of learning from a sample of studies of learning and to determine the problems of translating these principles into a form suitable for application in Navy schools.

Responsible for the isolation of these two relatively well developed subsystems are two necessary, but insufficiently developed, functions that have been the primary concern of this investigation. At the present time, one of these functions is neglected and the other is virtually non-existent. The first, labeled Collation and Interpretation, refers to activities required for relating and generalizing the findings of the very large number of highly specific basic and applied research studies in which the profession is engaged. This activity is a necessary condition for translation and theory construction. In its simplest form, it may involve little more than reviewing and classifying. It has as an end objective, however, the development of generalizations about learning behavior. Consequently, formulations at this stage are necessarily more abstract than they are at the research level although they are not as devoid of content as formal theory might be.

The next element in the model, and the one that is most conspicuously inactive, is that of Translation. The functions associated with this portion of the system have been variously referred to as "bridging the gap," "operationalizing," "engineering," etc. Whatever the label, the essential activity is one of relating the operations, variables, and functional relationships found to be important in laboratory studies of learning to corresponding operations and variables in some outside educational environment.

The translation function is also essential to communication both from the research community to the world of practice and the reverse, from practice back to research. Direct communication between these two subsystems presently is impossibe because of the lack of a common technical language. Most serious in the breakdown of the system, however, appears to be the positions taken by many research psychologists concerning the role of learning research in relation to educational technology and certain of their viewpoints concerning procedures desirable in conducting learning research. These are explored in detail in the body of this report. They are summarized briefly below, together with other findings of the study.

SPECIFIC FINDINGS AND CONCLUSIONS

A. Limited Impact of Learning Research.

1. Research on learning processes represents perhaps the largest single area of investigation presently being pursued by experimental psychologists. Although this has been true for some time, there has been no systematic effort directed toward practical application of the findings from learning research. As a consequence, modern learning research is producing very little impact on educational technology or training practice.

2. Both educators and scientists (of other disciplines) except psychologists to make contributions to a developing educational technology. Most psychologist studying learning seem to feel that this responsibility lies with others.

3. Both academic and practically oriented psychologists agree that a very small percentage of findings from learning research is useful,

ERIC Full Text Provided by ERIC

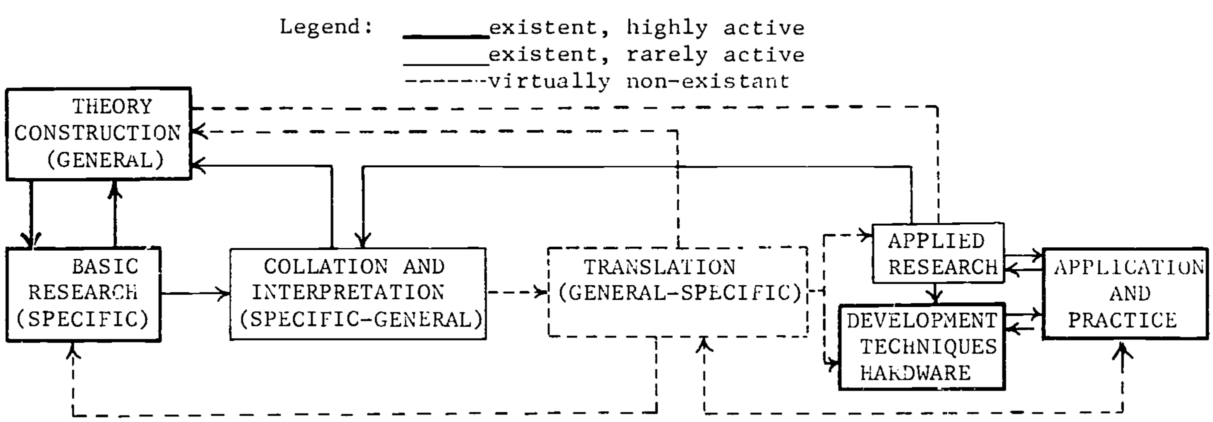
4. A study was made of the type and amount of information that would be required by research translators or educational engineers in assessing the utility and validity of principles derived from research for any specified purpose.

5. Using hypotheses developed from the above procedures, the judgments and attitudes of psychologists, renowned in the field of learning, in educational psychology, or in positions of responsibility for training research, were solicited on issues that were considered vital to the translatability and applicability of research results.

It was found that the research-to-application process never had properly developed for the psychology of learning. Consequently, there have been far fewer applications and much less impact on the educational process than might reasonably be expected in view of the size of the learning research effort. The reasons are believed traceable, in large part, to the research philosophies of experimental psychologists. But it was evident, also, that potential users have been reluctant to make the effort necessary to realize the benefits of research findings.

Figure 1 is a descriptive model of the functions and processes involved between basic research and practical application. Although an attempt has been made to make it particularly descriptive of learning research, it is generally descriptive of the process for all areas of psychological research as it exists today.

An attempt has been made to code Figure 1 so as to reveal the serious deficiencies found to exist in the total process. On the other hand, there is a lively subsystem of theoretical and basic research activity that feeds largely upon itself with an almost total lack of influence by the practical world. In fact, it is rare that even the results of applied research impinge on this subsystem. On the other hand, there is a highly active subsystem concerned with the development of materials, machines and techniques for use in the everyday business of meeting the educational and training needs of this country. This subsystem is almost totally divorced from the theory construction and basic research subsystem although it encounters, periodically, a limited influence by applied research.



in any direct sense, for the improvement of training or educational practices. Some contend that they are applicable to learning processes that occur in the training environment but admit that these processes have not been adequately identified.

B. The Inspiration for Learning Research.

4. The impetus for the great majority of learning research comes from theoretical considerations or previous laboratory research. This leads to an ever increasing narrowness of problem definition. Further, it cuts the researcher off from operational problems as a directional force in formulating his research.

5. Learning theory has been particularly barren of useful predictions about human learning behavior in the educational or operational environment. A major reason for this appears to be that theory development has been nourished by laboratory experiments involving tasks that bear no known relationship to real-life learning requirements.

6. It is generally agreed that fundamental knowledge of learning behavior can come both from research that is theory oriented and that which is stimulated by practical problems. The two sources are regarded as somehow complementary by most psychologists. In spite of this, learning specialists in many universities show little inclination to direct their own research, or that of their students, toward problems associated with educational or training operations.

7. Applied research as a source of fundamental knowledge about learning is often criticized by academic psychologists on the basis that the work is directed toward discrete, unrelated propositions. They are less likely to admit that a great deal of laboratory research is subject to similar criticism because of the narrowness of inquiry and specificity of experimental outcomes.

C. Specificity of Findings and the Experimental Approach.

8. The findings of psychological research in learning, and other behavioral areas as well, are distressingly specific to the stimulus (task) conditions, independent and procedural variables, and dependent measures selected for study by the investigator. Sensing this specificity, many potential users, rightly or wrongly, disregard any report whose title does not directly reflect the particular subject matter or task of their interest. Since most learning research reports fail to meet this criterion, they are conveniently ignored as something important but not relevant.

9. Many learning research studies simply are not translatable because the stimulus (or task) conditions employed by the researcher bear no determinable relationship to stimuli (or tasks) outside the laboratory. The only utility of such studies, in relation to application, is in producing hypotheses that become the subjects of additional research.

in any direct sense, for the improvement of training or educational practices. Some contend that they are applicable to learning _processes_ that occur in the training environment but admit that these processes have not been adequately identified.

B. The Inspiration for Learning Research.

4. The impetus for the great majority of learning research comes from theoretical considerations or previous laboratory research. This leads to an ever increasing narrowness of problem definition. Further, it cuts the researcher off from operational problems as a directional force in formulating his research.

5. Learning theory has been particularly barren of useful predictions about human learning behavior in the educational or operational environment. A major reason for this appears to be that theory development has been nourished by laboratory experiments involving tasks that bear no known relationship to real-life learning requirements.

6. It is generally agreed that fundamental knowledge of learning behavior can come both from research that is theory oriented and that which is stimulated by practical problems. The two sources are regarded as somehow complementary by most psychologists. In spite of this, learning specialists in many universities show little inclination to direct their own research, or that of their students, toward problems associated with educational or training operations.

7. Applied research as a source of fundamental knowledge about learning is often criticized by academic psychologists on the basis that the work is directed toward discrete, unrelated propositions. They are less likely to admit that a great deal of laboratory research is subject to similar criticism because of the narrowness of inquiry and specificity of experimental outcomes.

C. Specificity of Findings and the Experimental Approach.

8. The findings of psychological research in learning, and other behavioral areas as well, are distressingly specific to the stimulus (task) conditions, independent and procedural variables, and dependent measures selected for study by the investigator. Sensing this specificity, many potential users, rightly or wrongly, disregard any report whose title does not directly reflect the particular subject matter or task of their interest. Since most learning research reports fail to meet this criterion, they are conveniently ignored as something important but not relevant.

9. Many learning research studies simply are not translatable because the stimulus (or task) conditions employed by the researcher bear no determinable relationship to stimuli (or tasks) outside the laboratory. The only utility of such studies, in relation to application, is in producing hypotheses that become the subjects of additional research.

ERIC

10. Many learning researchers defend the use of simple or abstract experimental tasks with the assertion that general principles of learning are more readily discernible from such tasks than from more meaningful ones. However, there is abundant evidence from various fields of experimental psychology that the more abstract the task the less generalizable any resulting principles are except to other tasks of a similar, often narrow character.

11. Interpretation and translation of research is complicated by the limited duration of the great majority of learning studies. It seems likely that a good many experimental outcomes are time-specific. Surprisingly little attention seems to be directed toward time as a variable.

12. In general, studies which employ "natural" stimuli (and responses), that is, stimuli that are actually encountered in some operational task, are much more translatable than those which employ artificial or contrived stimuli. The use of natural stimuli, or accurate simulations of natural stimuli, is believed to be characteristic of most psychological research that has achieved application.

13. Translation and application are facilitated by the use of meaningful experimental environments as well as by meaningful experimental tasks. If a variable has an important effect on the learning process, it is likely that this can be demonstrated in the natural environment as well as in the laboratory. The natural environment has been rejected by many researchers on the assertion that it does not permit proper experimental control. This attitude shows some signs of change with the advent of instruction mediated by machine.

14. There is a need for laboratories that do basic research, even develop theory, but which operate in, or take their problem direction from, the operational environment. There should be more emphasis on how to cast operational problems in a form that can be investigated by research rather than on how to relate theoretically inspired research to operational problems.

D. <u>The Need for Translation</u>.

15. Application of learning research findings nearly always requires a process of translation from the variables and context of the experimental laboratory to the variables and context of some educational or training environment. But the translation of learning research is an almost totally neglected function.

16. There presently is no methodology for translating betwen the laboratory and real-world contexts because of the inadequate development of a common task taxonomy and consequent limited understanding of the process variables involved.

17. There are virtually no psychologists who regularly attempt translation. What translation does occur usually takes the form of guidance in designing an applied research study in which hypotheses

ERIC

developed from earlier laboratory investigations are tested in the training or operating environment. This is done rarely and usually only if the laboratory researcher himself is sufficiently interested in application to initiate follow-up investigations.

18. With respect to the difficulties of translation, there appears to be a fundamental difference between theoreticians and practitioners in psychology and their counterparts in the physical sciences. Physical scientists and engineers have the advantage of common technical terms and mathematical laws describing natural phenomena. Psychologists, on the other hand, are inclined to invent or contrive behavioral phenomena through the use of tasks that suit experimental convenience. The result is that it is extremely difficult to identify the learning processes involved in the laboratory as either similar to, or different from, those in a given educational environment.

E. Interpretation is Seriously Neglected.

19. In relation to the volume of learning research being conducted, the processes of reviewing and interpreting research results are being seriously neglected. Consequently, the development of principles of learning or generalizations from the many diverse studies conducted is severely limited.

20. Interpretation and translation of research is inhibited by the enormous volume of published research on very narrow, highly specialized aspects of learning. Regardless of its merits, the audience for most of these studiesis very small and similarly very specialized.

21. Interpretation and translation of research is severly inhibited by the difficulties involved in extracting the critical details of experimental procedure, stimulus and task variable, temporal conditions, and independent variables reported in the great volume of research performed by psychologists. This is particularly true of technical reports prepared for the government but many journal reports reflect the same problem. A standard synopsis of experimental studies, such as that developed during the course of this investigation, would facilitate interpretation and translation.

22. Generalizations from learning research have been inhibited by limited sampling of subjects, independent variables, and tasks. Recognizing this, some psychologists have adopted the position that no generalizations are possible in the absence of long programs of research involving many diverse studies that produce similar results. However, because learning researchers tend to invent tasks, rather than study natural ones, and because of the specificity of human response to these tasks, the prospects for being able to generalize become ever more remote.

23. Applications of research are greatly inhibited by the unwillingness of many psychologists to speculate on the implications of their findings. The attitudes underlying this reluctance have been reinforced by scientific conservatism and by the publication policies of

the technical journals. While this is to some degree defensible, it is clear that many psychologists prefer to live on the conceptual level, to enjoy the security of unassailable theoretical research, to avoid the risks and discomfort of stating implications.

24. Interpretation and translation of psychological research requires the development of a precise, technical language. However, the use of stylized language, esoteric symbols, and technical jargon has seriously limited the learning psychologist's ability to communicate with anyone other than other learning psychologists. Even psychologists are having difficulty understanding psychologists.

This is a major source of the communciation breakdown between psychologists and educators, and particularly between researchers and training personnel in the Navy. One result is that the average Navy instructor, or officer-in-charge of a staff of instructors, has limited knowledge of modern learning principles. He also has little information, and does not know how to get information concerning such things as the real worth of a proposed teaching machine, or the critical factors to be simulated in a new complex trainer. In fairness to the researchers, however, it must be admitted that the training community rarely admits to the existence of problems on which help is needed.

F. The Need for Learning Engineers.

25. To the extent that more translatable research is performed, there will be an increasing need for a corps of professionals who may be described as learning engineers. It is far more important that these individuals be trained as constructive critics of learning research, and as experts in the learning processes, than as educators or teachers. The tasks of relating theoretical, laboratory, and real-world variables, assessing the meaning of research findings, and innovating applications, make it essential that the qualification and training of these engineers be of a very high calibre.

26. The learning engineer will require specialized knowledge of the subject matter, or operational tasks, in the field where application is to be attempted. Without such knowledge, he will not be able to identify the operational counterparts of laboratory variables nor understand the nature of the stimuli operating in the real-world situation.

He will require training in research, but as a generalist rather than a specialist. He must avoid the trap of research specialization but at the same time develop specialization in the field of application.

27. There presently are no satisfactory channels for formulating and communicating meaningful problems from the training environment back to the research community. This is potentially an additional role for the translator or learning engineer. At the present time the only effective procedure is for the researcher to get first-

hand experience in the operating or training environment itself. The proportion of psychologists willing to "get their hands dirty" in this way is small.

28. Educators and instructor personnel are generally incapable of formulating their problems in research terms. Therefore, training in problem definition will be an essential and critical part of the curriculum for educational engineers. It is likely that this can be achieved meaningfully only through internships in agencies active in training or training research.

29. At the present time, university curricula, especially in psychology departments, are woefully inadequate for the development of learning engineers. Because of the depreciation of applied science in many universities, this is likely to remain the case for some time. However, the strong interest developing in computer aided instruction, and the emergence of centers devoted to the development of eduational technology in several major universities, may signify the beginnings of a beneficial change.

The findings and observations that led to these conclusions are presented in detail in the chapters that follow. Chapter II is concerned with the problem of the limited impact of learning research and the attitudes of psychologists concerning its applicability to education and training. Chapter III explores the viewpoints of psychologists on a number of issues concerning research procedures that affect the translatability and application of research. Chapter IV is devoted to a discussion of the factors inhibiting the processes of collating, integrating and interpreting research results. Chapter V is concerned with the missing function of translation and describes the qualifications and training necessary in a new type of specialist, the learning engineer. Finally, in Chapter VI, a number of specific recommendations are made to ONR regarding actions desirable if it is to realize greater application from psychological research.

ERIC

Section III: Instructional Systems Development

Editorial Foreword

Dale Hamreus presents a nine-stage outline of the major steps in developing an instructional system. The entire chapter is essentially a verbal elaboration of these nine stages, summarized in a diagram on the third page of the chapter. The first four of the stages represent capsule summaries of the previous chapters (I and II) in this volume. The other five stages describe (in the editor's words):

5. build a prototype of the instruction
6. try out the prototype
7. obtain evidence from the tryout
8. modify the prototype where necessary, based on the tryout
9. repeat the revision-test-modification cycle, if possible.

These five stages present a plan for useful trial runs. The nine-step outline is straightforward, and most of the explanatory narrative is clear. Unfortunately, for concrete-minded readers like the editor, no example of the model in action is presented in the chapter. However, its application to a study skill guide is illustrated in the workbook.

Dale Hamreus admits the frequent lack of knowledge on the part of the developer of the instructional system, which necessitates a resort to hunches, and trial and error tactics. Perhaps Dale admits to more limitations than the present state of our knowledge of man requires.

Stage Five, the production of the original prototype, is encumbered by an extensive digression into classifications of instructional media in terms

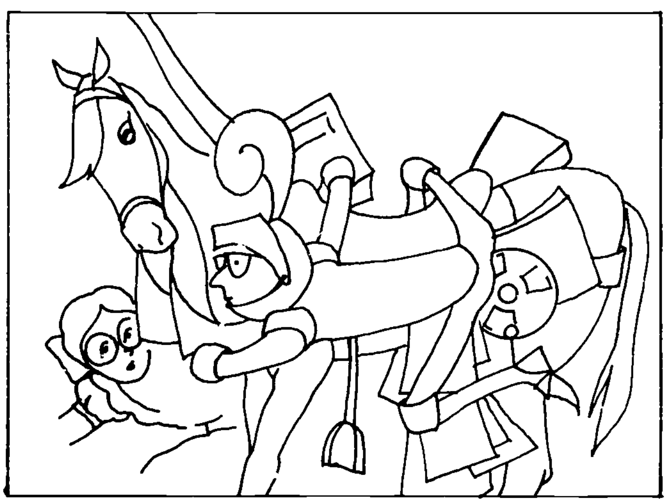
of the complexity of cues impinging upon the learner. This six-page section can be skipped without impairing the reader's grasp of the development model.

The author's portrait of instructional development as necessarily rough and ready may be comforting both to the novice and the harried developer. However, there is a correlated tendency to regard any measurement in such trial runs as necessarily imprecise. The editor feels that measurement of such behavior can be as precise in instructional development as in research efforts--depending on the kind of behavior to be measured and the value of increased precision.

The nine steps Hamreus presents can be expanded or condensed. We have seen versions with twenty or more steps. But, the present plan is sufficiently detailed to be useful, it is simple to grasp, and adaptable to almost any kind of instruction.

III

Try out--revise--try out--recycle...



Alice: "I'm afraid you've not had much practice in riding."

Knight: "I've had plenty of practice...plenty of practice...plenty of practice."

SECTION III

Instructional Systems Development

Dale G. Hamreus

Thus far in the institute you have learned (1) how to state behavioral objectives that clearly indicate what terminal or outcome performances are expected of the learner following instruction; and (2) how to determine the specifications for achieving the terminal behaviors in (1). The next step to be presented in this section, is to develop, test, and modify the prototype instructional materials and methods that are required to comply with (2) in fulfilling (1).

Why should we be concerned with all the time and energy required to develop an instructional system? Is not our teaching good? Frankly, no, at least not good enough! For example, there are other ways to teach than just the lecture method. Yet, it can be readily documented that by far the most frequently employed instructional methodology at any level is that of lecturing. Why is this? At least two reasons readily come to mind: (1) lecturing is probably the most natural methodology for teachers to employ, particularly since it has been the principal model confronting them throught their own education and formal teacher training; and (2) contingent upon the first reason, continuing increases in enrollments force the media-unsophisticated to persist in the lecture. Yet, the very fact of continuing increases in numbers of students results in proportionately less instructor time available to the individual student and argues strongly for the need of more systematically organized instruction. The

only way to break out of the lecture lock step and improve instruction is to systematically develop instructional systems that are valid for the purposes intended. This requires that all methods, materials and procedures be clearly identified, systematically developed, and refined until they effectively and efficiently achieve what they were intended to achieve.

Our present knowledge and skill in the "science" of developing instructional systems is admittedly limited; however, we have moved past the crude groping and floundering stage and are able to at least prescribe a modified trial-and-error process that works. As Dr. Gabriel Ofiesh, Director of the Center for Educational Technology at the Catholic University of America, recently said, we have developed learning technology to the point "where we can start shooting with a bullet instead of a shotgun." (1967, p. 26)

In this section of the manual discussion will center on the important stages and procedures required for developing and validating an instructional system prototype. These stages have emerged from developmental research over the past five years at Teaching Research and other research centers. The general process calls for the application of a systems approach and requires specific identification and specification of all elements of the instructional system along with systematic try-out, analysis and modifications until the desired outcomes are achieved. The flow chart presented in Figure 1 contains the major stages in the system.

Notice that stages 1 and 2, to identify behavioral objectives and determine enabling objectives, were discussed in Section I of this manual and stage 4 was presented in Section II. Stages 3, 8a, and 9a, dealing with evaluation and measurement are discussed in Sections IV and V.

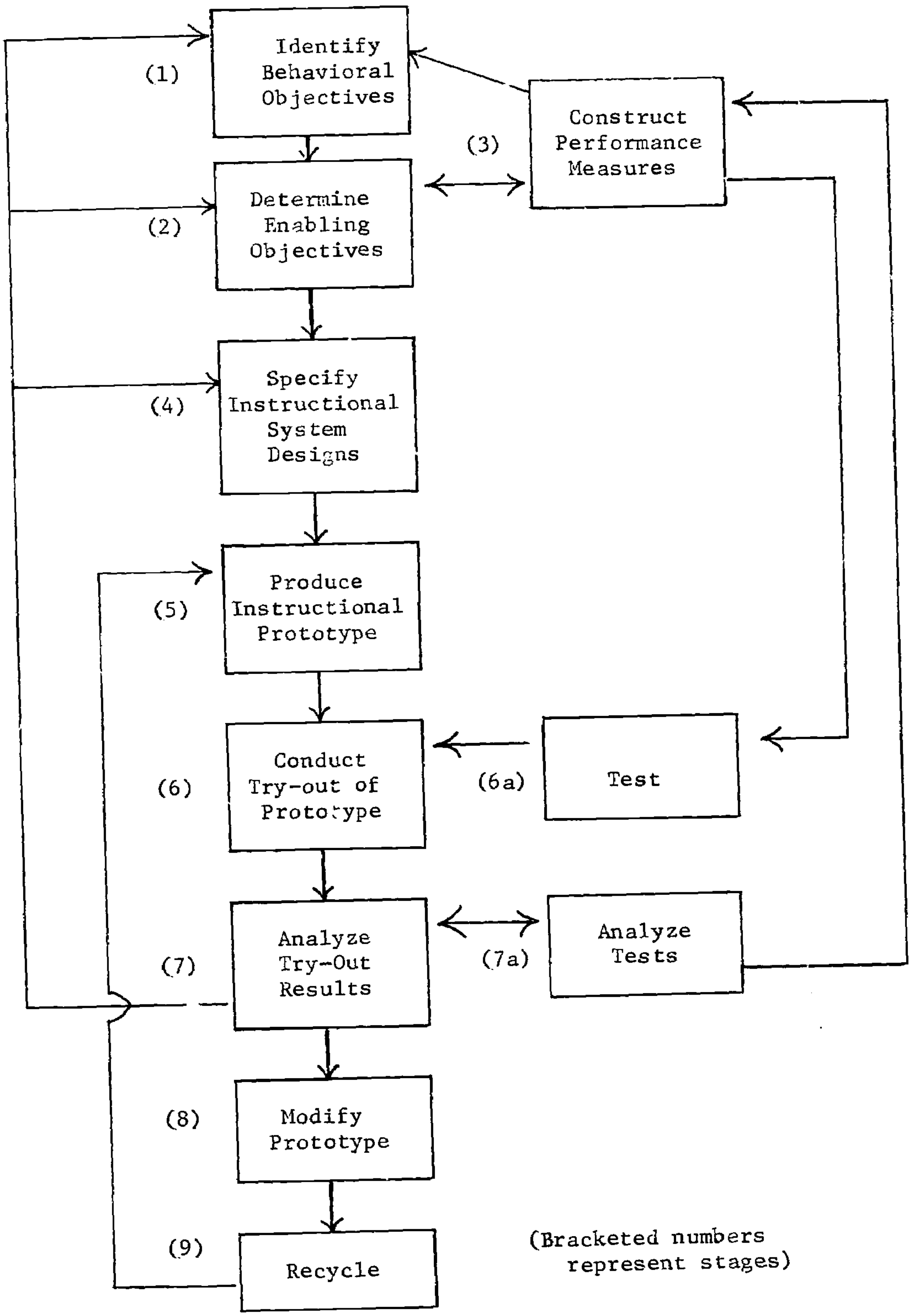


Figure 1. Major stages in system design for developing validated instructional systems.

Stages 5, 6, 7, 8, and 9 are specifically concerned with the process of building and validating an instructional system and are discussed in detail in this section. Stage 5 is concerned with the actual translation of specifications into the instructional prototype and involves content specialists, behavioral scientists and media experts. Try-out of the instructional prototype to determine the effectiveness of the new system is the central issue of stage 6; in stage 7, analysis of feedback from the try-out is discussed. Stage 8 concerns the use of the analysis for making necessary modifications to the new instructional prototype system. Recycling of this whole process to achieve continued refinement of the product occurs in stage 9.

The building and validating stages can be considered culminating activities to the total instructional system development. In order for the reader to more fully understand the relationships among all stages in the Systems Design in Figure 1, a brief review of the other stages and their importance to the stages in this section will be given.

Stage (1): The first stage in the development of an instructional system is to specify the behavioral objectives. First, check your memory on what you learned in Section I. In writing behavioral objectives, what are the principal elements that each statement should contain? You should have identified the following elements: 1) Audience, 2) Behavior, 3) Conditions, and 4) Degree. Can you reconstruct to what these elements refer?

If one is to achieve predictable learning outcomes from the instructional system being developed, it is crucial to know very precisely what

changes in behavior are expected of the learner and to have defined standards against which his behavior can be checked to determine whether indeed he has actually gotten there. This is the place of behavioral objectives. Behavioral objectives are the very cornerstone upon which the systematic development of an instructional system rests; without them one need go no further.

Stage (2): The next stage in the development of an instructional system is to determine the enabling objectives. Test your recall again and review, in your mind, what enabling objectives are. Yes, enabling objectives consist of component actions, knowledges, skills, etc., that enable the student to attain the specified terminal objectives.

Certain assumptions must be made regarding what sub-behaviors the learner must acquire if he is to achieve the stated terminal objectives. These assumptions are based on empirical evidence, research, theory, and many times a seat-of-the-pants logic. They become the basis for designing the specifications that guide the instructional system prototype development.

Stage (3): At this point in the development of an instructional system it is necessary to construct performance measures. Pertinent discussion on this concept is presented in Sections IV and V. However, a few important points will be made here to stress the important place that performance measurement holds in instructional system development.

Criterion measures are necessary for determining whether the learner meets or exceeds the level of performance expected for each behavioral objective. If this was the only use for performance measures there would be only minimal

need for this stage to be included in the system design and no need at all for it to be located opposite stages (1) and (2): behavioral objectives and enabling objectives. By developing measures for assessing criterion performance at the same time as behavioral objectives are determined, rather than after development has been completed, eliminates the pitfall of assessing that which has been developed rather than changes which were desired in the learner. It also requires that close scrutiny be made of the behavioral objectives which has the advantage of uncovering ambiguities or gaps in the objectives.

Two very important additional functions are served by performance measures that argue strongly for performance test construction to be tied to stage (2), determining enabling objectives. First, a diagnostic means is needed for determining the validity of various steps of the instructional system. By developing performance tests from enabling objectives a grid system of tests is produced for assessing all parts and points of the instructional system to determine where weaknesses exist.

Second, at our present level of sophistication, one of the weakest links in instructional systems development concerns the assumptions underlying the determination of enabling objectives. For example, as subject-matter properties to be generalized and discriminated shift from simple to more complex such as the musical concept of early to late Beethoven, the instructional process becomes more complicated. A major problem in teaching the more subtle and complex concepts is in the analysis and definition of subject-matter properties. Such analysis becomes even more difficult when semantic confusion exists and where there is disagreement among experts. Assumptions regarding what elements

ERIC

of knowledge or what schedules of reinforcements are required to enable the learner to attain the desired terminal behaviors of the more complex concepts are apt to be faulty and must be tested. Performance measures designed to test assumptions upon which enabling objectives have been determined, improve the process of making the steps in the instructional system valid.

Stage (4): The fourth stage in the systems design flow chart is to specify instructional system designs. This is a complex process as Section II of this manual points out. However, it can be reduced to the major areas of:

1) specifying learner characteristics, i.e., identifying those aspects of the target audience which will require "tailor-made" instructional developments;

2) determining the learner responses called for during instruction. (Once the types of learning represented in each objective have been identified then strategies by which they can be made manifest can be systematically developed and tested. Such a function narrows down possible alternative strategies from which to choose in developing the product.);

3) describing the learning context, i.e., when types of learning have been determined, then it becomes necessary to identify the events that provide the conditions of learning to occur from each objective:

4) designing the mediational format, i.e., specify what form (verbal, visual, etc.) is to be used for the various types of learning identified to transmit the content to the learner. At present, this is perhaps our weakest link in translating specifications into products. Some research evidence is available to guide the selection of audio and visual elements in instruction but this evidence is still very thin. Educated guesses and

empirical evidence of pilot tryouts are still needed to determine the "best" form to use. Unfortunately, in the teaching-learning context, time and costs prohibit large scale controlled trials.

5) Defining the systematic routes by which the learner achieves feedback about his learning efforts.

All stages in the system design thus far discussed are essential to the actual building of the system. Methods for gathering inputs to develop the system and for determining required output of the system have been described in detail. Inputs are to be identified in terms of prerequisites, enabling objectives and conditions for learning; required outputs take the form of specified behavioral objectives. In addition, the means of identifying instructional strategies along with supporting media forms are to be described. The portions that now follow in Section III are addressed to the specific activities necessary to build and validate the prototype system. In order for this discussion to be meaningful, the reader must assume that the content for the instructional system has been identified and sequenced and that special equipment required in engaging the system is available. Criterion measures, both intermediate and end of course, are to be prepared. Now, the parts of the prototype system must be produced and systematically assembled, tried out with learners, appraised for effectiveness, analyzed to determine where to make necessary modifications and the whole process recycled until the levels of desired outcomes are achieved.

The development of a complex instructional system includes planning of the interactions among all elements which constitute the system. However, if such a complex instructional system is to be successfully developed close attention must be given to validating the system. In the past, failure to

test the validity of instructional planning has limited the effectiveness of instruction. Only as all aspects of an instructional system are systematically tested can feedback be obtained to direct attention to portions that are weak. But equally important to testing is an organized instructional system that can be replicated, thereby permitting the validation process to occur.

Stage (5): The fifth stage in developing the instructional system is to produce the instructional prototype. This becomes the first stage of actually creating the instructional system--systematically following whatever instructional specifications have been provided to give the system substance, form, and order. The process is one of translating written statements (specifications) into prototype materials, equipment designates, and instructional sequences. The process is analogous to a Boeing structural engineer creating the complex 727 system from a set of blueprints and specifications. In both cases the degree of system integrity determines the extent of functional success--obviously more noticeable if lacking in the 727 system but no less important in the instructional system.

In creating the prototype system it is essential to adhere very closely to the instructional specifications. Content must be written into statements or translated into visual and/or audio forms, then arranged into sequences that will accomplish desired behavioral changes. Media forms must be determined for each aspect of the content; i.e., whether to use printed forms, slides, overhead transparencies, audio tapes, 8mm loops, 16mm movies, various combinations of the above. Formats for each selected media form must be established; i.e., whether printed statements are to be hand lettered or typed and enlarged; whether to photograph real objects, caricatures, or abstract symbols; whether to use black and white or color. Strategies for sequencing

ERIC

the content and media and for involving the learner and the teacher must be decided upon which will be most effective in producing behavior changes and interesting to the learner. Decisions must be made detailing exactly what content elements go first and which should come next; transitions between elements must be specified; what actions and routines the learner should experience must be determined; how the teacher is to interact in the learning situation must be planned.

In addition to the above, the preparation of a teacher's manual or guide is important. If the instructional system is to be employed effectively, the teacher using it must be provided a detailed set of instructions explaining the exact purpose of the system and how it should be employed. Simple systems, such as a self-instructional program, usually have minimal teacher involvement during instruction; however, the more complex the system the more crucial the role of the teacher.

During prototype development it is essential to maintain close liaison between the content specialist, media specialist and behavioral scientist. To the extent each knows what the other is doing greater developmental continuity will result. For example, if visuals are specified for certain content portions, considerable headway can be made if a graphics artist, content specialist and learning specialist can meet together during which suggested sketches can be quickly prepared, jointly viewed, and interpreted.

Even with specifications given, numerous decisions must be made during the development to satisfy the stimulus dimension required to produce optimal learning. For example, although content may have been defined in the specifications, exactly what combination of words should be used, what language level represented, what format for displays, which sensory examples

to use, what exercises to include, etc., must be eventually resolved.

Regardless of the detail in specification preparation, it must be realized that still lacking is the sophisticated knowledge necessary to prescribe all the conditions, strategies, and forms required in the instructional system to produce optimum outcomes.

A useful categorization has been offered by Edling (1966) of classes of sensory cues into some meaningful and identifiable dimensions. The model presented in Figure 2 represents the entire spectrum of instructional stimuli arranged in a single continuum. Although there is conflicting evidence regarding the sensory mechanism and the nature of the learner interacting with more than a single class of stimuli in a particular situation, until better evidence and information are available this model is offered simply as a guide in deciding what stimulus elements are to be used in the instructional system prototype development. Admittedly, the subclassifications presented are highly debatable, but at this stage, the same could be said of any other categorizations. Each of the separate classification sections of the continuum will be briefly discussed.

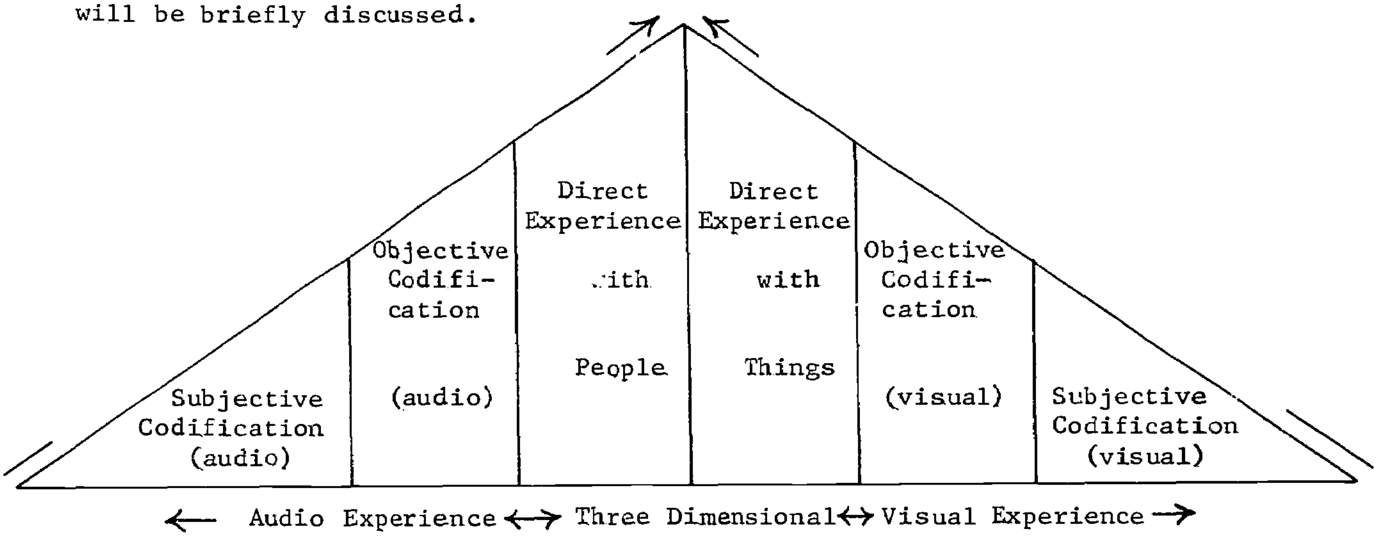


Figure 2. The continuum of instructional stimuli.

The classification suggested in Figure 3 represents three-dimensional direct experiences with people or people surrogates, who have the capacity for symbolic communication. The category "authentic situations" is used to identify life experiences that are not preplanned or structured in any artificial sense. This category is intended to represent the real life occurrences. It can be hypothesized that each succeeding category, moving from authentic situations, introduces increasing dimensions of artificiality while at the same time reduces in the potential number of cues.

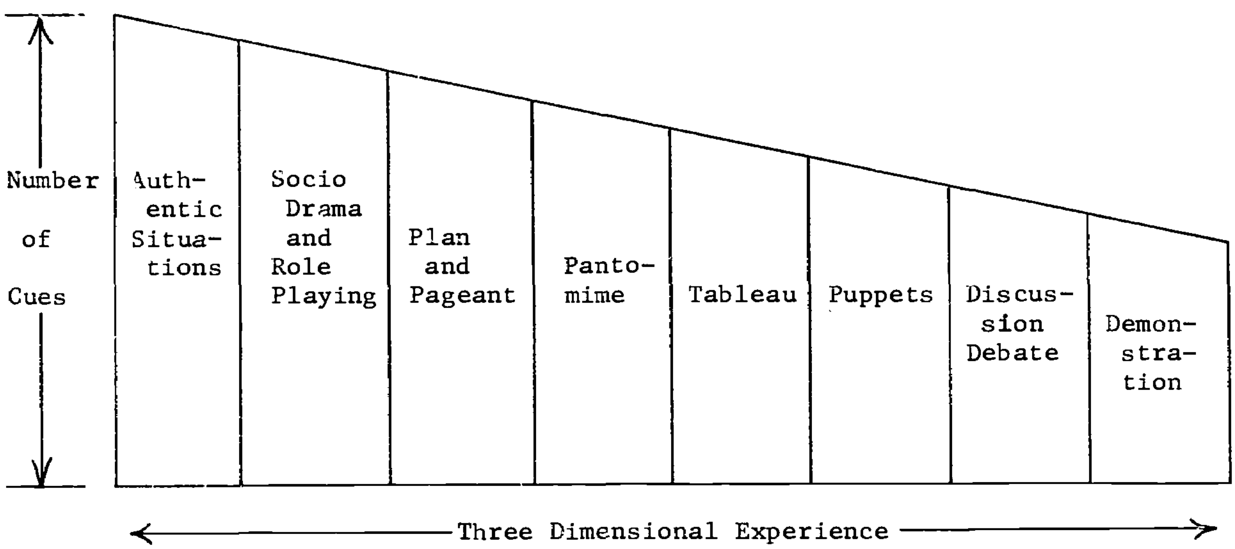


Figure 3. Direct experience with people

Direct experience of learners with three-dimensional objects other than people is represented in Figure 4. The actual object itself would appear to be the source of the largest number of cues. Other categories to the right of object itself tend to reduce in number of cues; however, they do not easily lend themselves to these classifications. The usefulness of

the classification of any object would, of course, be dependent on the learning objective.

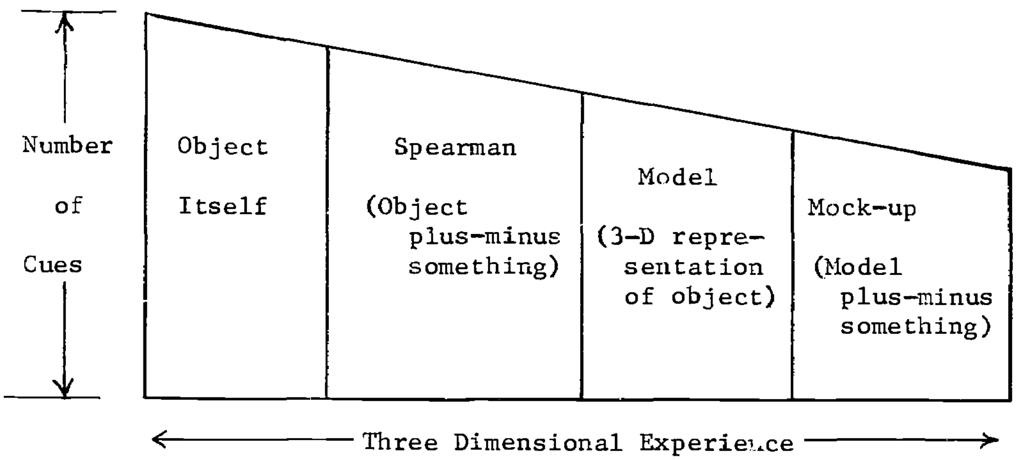


Figure 4. Direct experience with things.

The next two sections of the continuum of instructional stimuli concern <u>objective codification</u> of visual and audio experience. Objective codification is defined as: of, or having to do with a known or perceived object as distinguished from something existing only in the mind of the subject or person thinking, hence, being or regarded as being, independent of the mind; real; actual. Determined by and emphasizing the features and characteristics of the object or thing dealt with rather than the thoughts, feelings, etc., of the artist, writer, or speaker; as an objective description, painting, etc., hence, without basis of prejudice; detached; impersonal.

Objective codification with visual experiences moves from three-dimensional stimuli to the single sense of vision. Figure 5 lists the various sub-classes of elements represented in two-dimensional visual experiences. Motion

pictures in color which create the illusion of three dimension may be very close, perceptually, to three-dimensional experience. The number of potential cues decreases as one moves from this dimension to those of lesser realism.

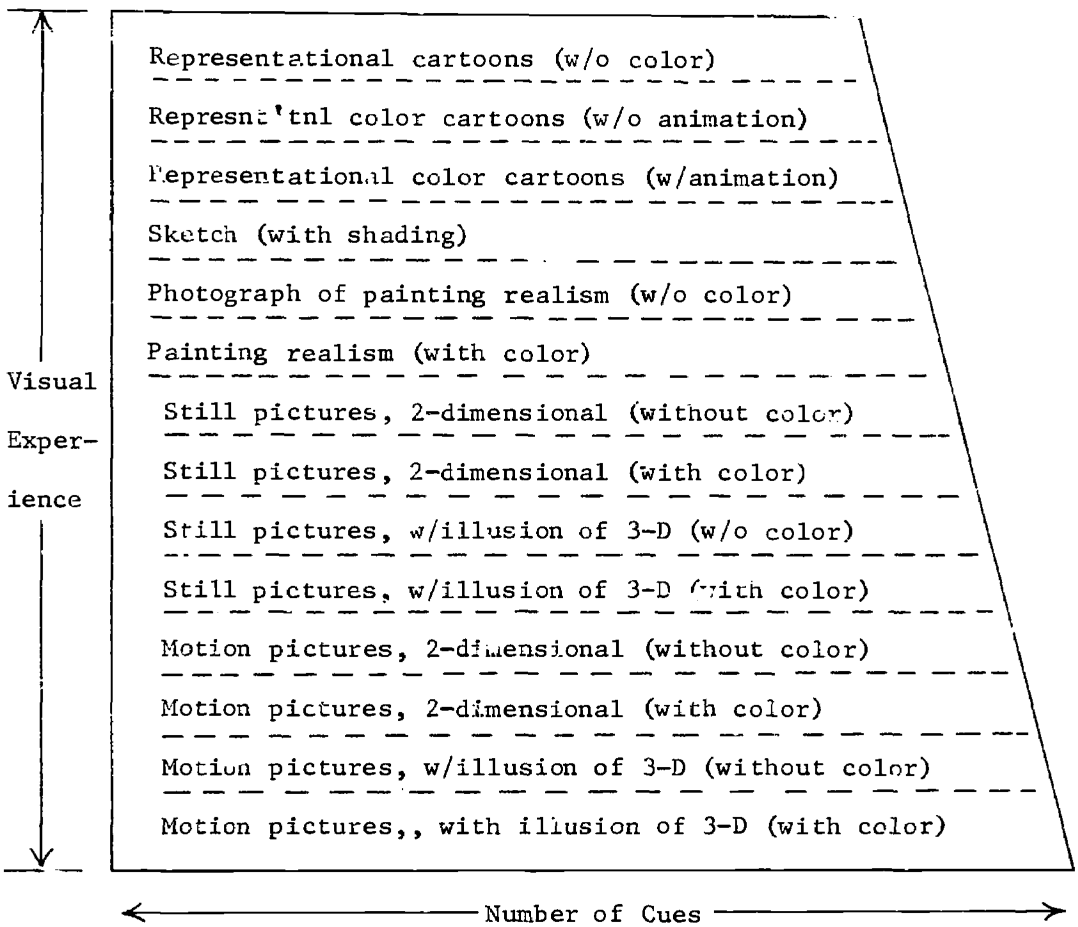


Figure 5. Objective codification with visual experinece

The continuum section representing <u>objective codification</u> employing <u>audio</u> experiences is presented in Figure 6. Sounds represent directional elements and audio frequencies present in reality. As the fidelity of the recording

decreases and when artificial representation of natural sounds occurs the number of cues present are reduced.

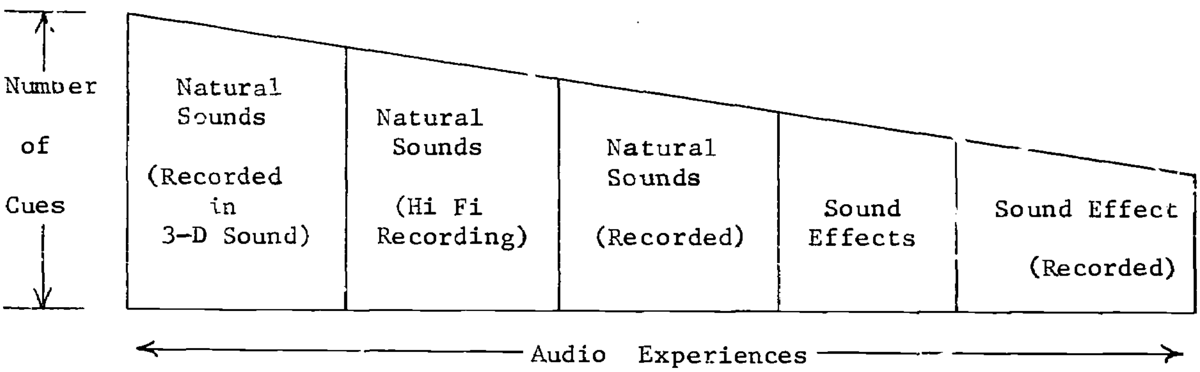


Figure 6. Objective codification with audio experience

The tail sections of the continuum of instructional stimuli concern <u>subjective codification</u> which is defined as: of, affected by, or produced by the mind, or a particular state of mind; or resulting from the feelings or temperament of the subject, or person thinking, rather than the attributes of the object thought of: as, a <u>subjective</u> judgment. In <u>Psychology</u>: existing or originating within the observer's mind and hence incapable of being checked externally or verified by other persons.

Figure 7 is a diagram of subjective codification with visual experiences. All of these subclasses require that the learner must have previously known the symbol (the code) in order to be able to associate it with the object it represents.

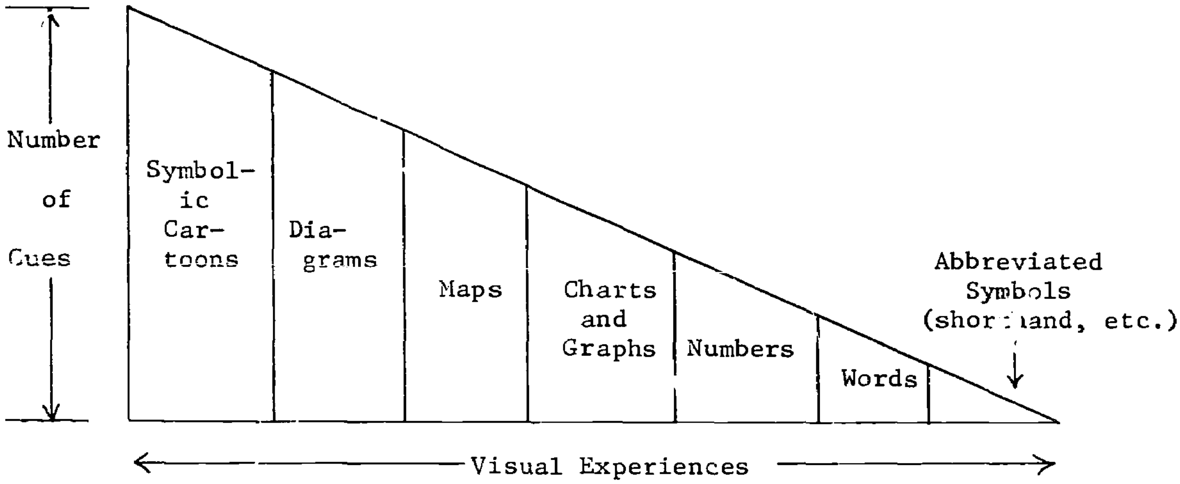


Figure 7. Subjective codification with visual experiences

The continuum section of <u>subjective codification</u> with <u>audio</u> experiences is presented in Figure 8. Here too, symbolism is involved which requires a knowledge of the code before interpretation and association with the object are possible. For example, musical notation may represent the pounding of the surf but many naive learners would not make that association without prior knowledge of the symbolic representation.

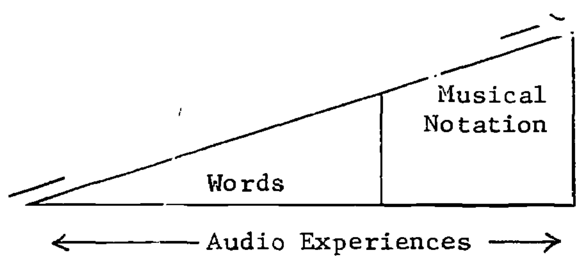


Figure 8. Subjective codification with audio experiences

The following are a few general guidelines which derive from the continuum of instructional stimuli model and are suggestive of the type that might be helpful in making decisions about the form of the stimuli in instructional system prototype.

1. If the object or event to be taught is available and cues can be perceived, then use direct experience. For example, if youngsters are being taught about the post office and one is available in the immediate area, depending on the objective it might be preferable to take the youngsters to the building rather than employ symbolic representation of a post office.

2. If the object or event to be taught is not available and cues are not readily perceived, use pictorial representation. For example, some underwater phenomena are unavailable for direct classroom instruction, therefore objectively codify them (present them by visual means).

3. If the object or event to be taught makes vibrations and these are significant cues and are available, use natural sounds. For example, the sharp rattling sounds emitted from the tail of a rattlesnake are more effectively taught by demonstrating the natural sound.

4. If the object or event to be taught makes vibrations and these are significant cues and are not available, use auditory representations. For example, heartbeat sounds symptomatic of various maladies are readily available through objective codification (tape recordings).

5. If the object or event to be taught is psychological in nature, that is, no iconic (image) representation is possible, use audio and/or visual symbols. For example, since no one has ever seen an atom, subjective codification must be employed in instruction.

6. If the object or event to be taught is confusing, complex, and/or hidden use symbolic representation. For example, subjective codification (visual and audio experiences) must be used to transmit musical ideas.

Stage (6): The next stage in the development is to conduct try-outs of the prototype. This is basically a quality control measure built into the developmental process. The concern of this stage is to subject the newly developed system to try-outs with appropriate target populations to determine whether the content, strategy, supporting facilities, equipment and materials of instruction actually do produce the desired changes in behavior. The primary purpose here is to obtain data from observations and other measures such that weaknesses in the system can be identified and changed.

Although this stage implies end-of-development try-out, actually try-out must be ongoing throughout the entire development. If the instructional

development process is not interrupted periodically to account for what has already been completed, faults in the specifications or the production already completed are apt to be overlooked thereby resulting in considerable waste of time and money. For example, as the content of the instructional system is developed into specific written statements, unless statements are checked for accuracy, tried out for language level, judged for interest level, etc., only limited confidence can be held that these elements will indeed be adequate at completion.

The term try-out tends to imply involvement only with appropriate type learners to determine if their behavior will be adequately changed. This interpretation is much too narrow and would be too slow and limited a process to produce all of the desired feedback. The term target population used above includes all possible sources to provide feedback of the system's effectiveness. For example, in addition to feedback from learners, written portions of the system should be submitted to an "unbiased" discipline expert to check accuracy of content. Unbiased is used here to mean some one has not previously been involved in the new system development. Feedback from a learning specialist should be sought to check integrity of examples to learning types and the efficacy of the events employed to produce intended types of learning. Media specialists should check fidelity and quality of audio and visual elements; a "typical" teacher should react to the instructional strategy to identify awkward and/or impractical routines. The above are representative check areas and others should be engaged as the system suggests. To the extent that feedback can be obtained from various specialists prior to involvement with actual learners, the chances of obtaining maximum beneficial feedback in the try-out with learners is increased.

ERIC

Extremely important to the try-out is selecting the individuals to cooperate. Care should be taken that they represent an appropriate level of competence or ability. Learners should be representative of the audience for whom the system is intended. However, the effort should be made to locate individuals willing to take the time and "go to the bother" of conscientiously trying out and reacting to the new prototype materials. Hopefully, they will be individuals who will give "honest" responses. Subject matter, learning, media and other specialists should be skilled and obviously competent to advise in their respective areas.

The learner must be thoroughly briefed on what behavior is expected of him during try-out. The emphasis should be placed on his seriously trying to learn from the materials. However, he should be prompted to verbalize about the places he finds confusing, too difficult, already known, etc. Further, he should be encouraged to comment if he finds dull and uninteresting areas.

Close observation must be maintained during the try-out with learners to produce the required feedback. Whereas curriculum studies and other materials developments, such as commercially prepared products, tend to take large segments of the completed materials into the classroom for testing, the process here advocated is to try-out and test small segments of the unfinished system with a single learner at a time. The learner is instructed to identify any elements of the materials that are confusing or difficult to understand, such as words, phrases, pictures, symbols, test items, etc. The try-out monitor makes careful record of the location of all such identifications along with any other significant occurrences during the try-out, i.e., puzzled expressions, evidence of being bored, undue time required to complete various sections, etc.

ERIC

If the monitor suspects the learner is having difficulties that he is not expressing he should interrupt the instruction and tactfully question the learner. Care must be taken in questioning not to create anxiety in the learner. If the subject becomes anxious the value of that try-out is greatly reduced and it would be better to discount the effort and reschedule another try-out with a different learner. It is important to conduct try-outs with as much instructional realism as possible, i.e., present the learner with intact instruction including the designed teaching routines.

As large segments of the new instructional system are developed, more emphasis is given to learner try-outs and less to specialist feedback. The emphasis shifts from individual to small group learner try-out. Less close observation to the individual is maintained and more reliance is placed on test results.

To the extent that tests have been developed for sub-portions of the new system and their validity and reliability are judged sufficiently high, such tests become diagnostic tools for determining weakness in the materials.

Certain details accompanying try-out activities are important not to overlook. Clearance for using try-out subjects with appropriate administrative level authorities prior to try-outs will prevent possible subsequent misunderstandings and will greatly improve the chances of obtaining additional future try-out subjects. In some instances providing transportation for subjects will avoid time lost and will preserve schedules. Monitors must be trained to try-outs in what to observe and how to behave during try-outs. In situations involving total classroom try-outs, detailed instructions to cooperating teachers specifying all required aspects of the try-out are essential.

ERIC

When all development of the prototype system is completed, field testing in context is the necessary final stage of try-out. By this time, segmented and small scale try-outs should be completed and necessary final revisions or modifications to the system processed. It is essential to try the total instructional system in the context of an actual classroom, with all its normal contingencies (expected or otherwise), to see if the expected changes in learner behaviors still occur. The system should be sufficiently complete in all parts--instructional materials, teacher guide, tests--that when handed intact to a teacher, classroom instruction can be readily initiated. If, in context, the system brings the expected proportion of learners to criterion within the allotted amount of time then development has been successfully completed.


3.1

Stage (7): Another quality control point in developing an instructional system is to analyze the try-out of the prototype. The purpose of this stage is to determine whether the instructional system attains its objectives. Two primary uses are made of this analysis: 1) to determine what portions of the new system are faulty or not adequate and require modification, and (2) to determine which specifications were unrealistically and/or improperly established and should be changed.

Analysis of the try out of a new instructional system is not nearly so precise as that resulting from the customary experimental research: the principle difference centers around controls. Where experimental studies place considerable emphasis on appropriate controls, developmental research of new instructional systems, on the other hand, seldom offers any basis for

controls. The new development is just that, new! No suitable comparable thing exists against which it can be compared. Similarly, no substantial basis exists against which to control.

Analysis must be maintained on an objective rather than comparative basis. That is, how far does the system go toward achieving the ends for which it was developed rather than how much better did the system perform compared to some other system. When a new system has been sufficiently developed and refined and in use over a period of time such that its accomplishments can be predicted, then comparative analyses with another established system are appropriate. When the system is still new and essentially untried the only basis for judgment lies in how well it attains the objectives for which it was created.

Other developmental research efforts also employ the objective oriented analysis for making judgments. For example, much of the performance of the Boeing 727 in wind tunnel try-outs was analyzed against new objective standards. Although certain factors were common to other older systems, the complex combination of old and new had to be very rigorously tested against newly established standards. The only basis for acceptance was successful attainment of the standards. If a portion of the 727 system failed in the test, careful analysis was made of what contributed to the failure so that modifications could be made. The analysis seldom could fix the exact cause of the fault; however, it could pin-point the area of weakness and thereby contribute significantly to planning for modifications and subsequent new trials.

How, then, does one proceed in objective analysis? It is tempting to say all is fair in love and new instructional system analysis and anything goes.

However something better must be said if the concepts of research and the "scientific method" are to be amplified. The difficulty is that the scientific method, in the form it is generally regarded, other than the ideal of rigorousness, has little guidance to offer this form of developmental research. We still know too little about how best to analyze the results of systems try-outs to prescribe any precise methodology. Literally, any and all means for analyzing results that meet the concept of rigorousness are currently being employed.

Generally speaking, analysis must closely parallel try-out activities. That is, as small segments of the instructional system are tried out with various specialists and individual learners, careful recording of the details of these tryouts must occur in order that necessary modifications to the segments and next steps in the system development can be made.

Feedback from specialists can be obtained in various forms depending upon the arrangements made. Discipline specialists will usually prefer to study the subject matter content of the materials privately, making editorial notes of both a technical and stylistic nature. The developer usually reviews these notes with the specialist for clarification before analyzing their implications. Media specialists will frequently prefer to judge the materials in concert with the developer to be sure of interpretations. Criticisms in these instances are noted in detail by the developer and through this interaction take the form of analysis. Try-out with a typical teacher is usually jointly managed with the developer present. As in the situation with the media specialist, the analysis results from the interaction of the two.

Analysis of small segment try-outs with learners depends upon the nature of the observations employed and the form in which they are recorded. In

many instances, the very fact that learners express having difficulty with a particular portion of the new system will cause the developer to view that portion in different perspective. Verbal flaws or illogical sequences previously overlooked sometimes almost "spring out" at the analyst as a result of such learner feedback. For example, analyze the Readers' Guide elements you hypothetically tried out in Workbook Exercise 3.1. It probably became immediately clear that the learner could not find the author's name in the visual and that the visual did not include an author, yet discussion about the author was called for in the worksheet question. This form of analysis points to obvious flaws in the development that were somehow overlooked and which lead directly to revisions. Although the example is a simple one the concept applies to all situations.

Other forms of analysis are often less obvious. For example, perhaps the only evidence of weakness in an instructional segment is that each try-out subject has not learned a particular bit of information coming from a specific part of the instruction. However, no obvious flaws, as in the above example, are evident in the materials. Careful study of the segment might reveal, for example, that double meanings are possible in the choice of words included and have led learners to an erroneous interpretation, or that particular sequential elements have combined to produce conditions of retroactive inhibition.

A more formal means of analysis than that described above occurs with criterion testing. Considerable attention to validity and reliability of the tests are necessary. Gagne, in describing research dealing with attainment of learning sets relevant to terminal objectives, demonstrates one example of how to analyze the try out of an instructional program using criterion measures (1961, p. 10). He first developed a hierarchy of learning sets by employing

the question, "What would an indivudal have to know how to do in order to achieve successful performance of this particular class of task, assuming he were given only instructions?" Five major levels of learning sets were thus formed.

In testing the measure of transfer among learning sets Gagne analyzed four possible empirical relationships for passing and failing relevant to higher-lower learning set combinations. These four were as follows, each combination of two plus- and/or -minus signs to be read from right to left rather than the conventional left to right pattern: (1) ++, indicating positive transfer from lower to adjacent higher learning sets; (2) --, indicating failure of lower level set followed by failure of adjacent higher level set; (3) +-, indicating failure of lower level set followed by passing of adjacent higher level set; and (4) -+, indicating passing of a lower level set followed by failure of adjacent higher level set.

The patterns of pass-fail relationships on a criterion test administered following the learning program are presented in Table 1 (Gagne, 1961, p. 10.) Although these results were primarily intended for other purposes they do demonstrate a good example of determining weak areas in a newly developed instructional system. Those interested in the interpretation of the total table are referred to Gagne's monograph.

ERIC

Table 1
Pass-Fail Relationships of Achievement Between Adjacent Lower and Higher Level Relevant Learning Sets (N=118)

Transfer to Learning Set	Frequency of pass-fail Pattern--Higher, Lower ++ (1)	-- (2)	+- (3)	-+ (4)
1. IV2 from IVA1	110	0	0	8
2. IV5 from IVA3	113	0	0	5
3. IIIA1 from IV2, IV3	85	0	7	26
4. IIIA2 from IV4, IV5, IV6	94	5	10	9
5. III1 from IV1	45	9	1	63
6. III2 from IV3, IIIA1	68	30	6	14
7. III3 from IVA2, IV3	75	25	7	11
8. III4 from IIIA2	62	40	4	12
9. II1 from IV2, III2, III3	34	70	3	11
10. II2 from IVA2, II3	41	60	2	15
11. II3 from III4	37	72	3	6
12. II4 from III4	9	85	0	24
13. I1 from II1, II2	25	78	2	13
14. I2 from II2, II3	28	80	3	7
15. I3 from II3, II4	6	104	0	8

In Table 1, the frequencies in column (4) -+ (passing lower level, failing higher level) indicate the places in the instructional program at which greater or lesser effectiveness was attained. Large frequencies are interpreted as points at which the program was relatively ineffective, small frequencies the points at which learning was effective. For example, as shown in Table 1, learning set number 5, progression from lower level learning set IV1 to adjacent higher level learning set III1, was failed by 63 of the 118 learners. Examination revealed that only one frame of the program was devoted to the content of set III1. Obviously insufficient instruction was provided these 63 learners

in that one frame for them to fully understand the new information. Similar results pertained to learning sets number 3 and 12 in which 26 and 24 failures respectively occurred at the higher level after success at the lower level set. Depending upon the percent of errors acceptable by the developer, other frames in the program in this example would be subject to similar scrutiny. This form of analysis obviously relies very heavily upon a powerful criterion test--one that tests all aspects of the instruction system.

Stage (8): As try-out and analysis of the new instructional system progresses, the purpose of all this is to provide a systematic basis for _instructional material modification_. During this stage, results of the analysis are translated into changes in the product. In some instances revisions can be made with much more certainty than in other situations. For example, analysis, such as the omission of the author's name in the Readers' Guide materials, should prompt a clear-cut addition to the visual; however, other revisions to the visual and worksheet based on the additional analysis cannot be made with such confidence. In all cases, any modifications made must be tested for effectiveness. This latter is a crucial point to the successful completion of an effective product.

Modifications should be planned with the same team of specialists who contributed to the development. For example, consider the _Readers' Guide_ elements introduced in the workbook exercise. Analysis revealed that the visual contained too much information at one time. Methods of reducing the amount of information should be determined in conjunction with the media specialist.

Stage (9): The final stage of developing the new instructional system concerns the _recycle_ of all product development stages, that is, returning to

ERIC

stage 5, (or earlier stages) producing the modifications to the instructional product (or stages), and going through the cycle again. The purpose of this stage is to point out the importance of continuing with the development, testing, analysis and revision process until the desired outcome learner behaviors are achieved.

Obviously this total process is not one of a series of four discrete stages of which the second cannot be started until the first is completed, etc.; rather, it is a complex process of overlapping and combining. It is not unusual for different stages to occur together; i.e., a trade back and forth between production, try-out, etc., involving various segments of the system. It might be more efficient, for example, to develop several similar visual elements at the same time even though they are located in different parts of the system. Feedback regarding their fidelity and readability can very often be obtained before the instructional sequence to which they belong is ready for trial.

To the extent the developer can persevere in the development cycle until desired behaviors are achieved his product will improve; assuming, of course, that the desired learner behaviors are indeed desirable.

Bibliography

De Cecco, John P., (Ed) _Educational technology_. New York: Holt, Rinehart, and Winston, 1964.

Edling, Jack V. Media technology and learning processes, in John W. Loughary (Ed) _Man-machine systems in education_. New York: Harper & Row, Publishers, 1966.

Glaser, Robert, Toward a behavioral science base for instructional design, in Robert Glaser (Ed) _Teaching machines and programmed learning, II_. NEA: Washington, D. C., 1965.

Glaser, Robert (Ed) _Training research and education_. New York: John Wiley & Sons, Inc., New York, 1962.

Gropper, George L. Learning from visuals: some behavioral considerations, _AV communications review_, 14, pp. 37-69, Spring, 1966.

Innovation in education, _Educational technology_. Washington, D. C.: Aerospace education foundation, Spring, 1967.

Lange, Phil C. (Ed) _Programed instruction_. NSSE, 1967, Part II.

Tracey, William, R., Flynn, Edward B., Jr., and Legere, C. L. John, Systems approach gets results, _Training in business and industry_, 4, pp. 17+, June, 1967.

Vandemeer, A. W., Systems analysis and media--a perspective, _AV communications review_, 12, pp. 292-301, Fall, 1964.

Section IV: Evaluation of Instructional Systems

Editorial Foreword

In this chapter, Casper Paulson continues the presentation of instructional systems development by describing several techniques which may be used in try-outs or prototype testing of instructional units.

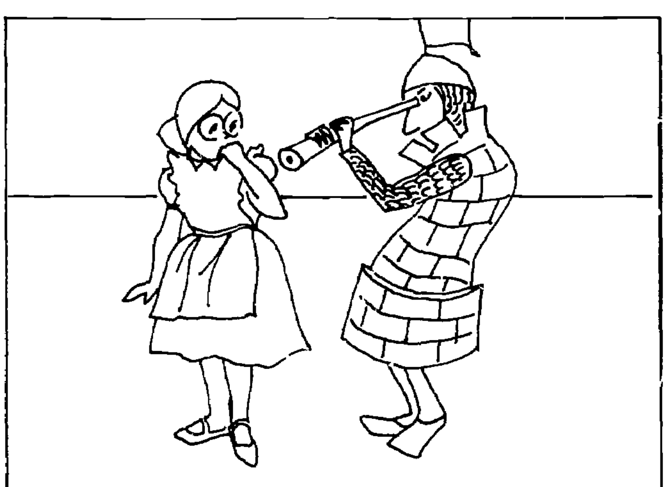
The reader should be forewarned:

1. That the chapter is not a review of the field of evaluation. It is focused on techniques to evaluate a developing instructional system.

2. That within the above arena, the author's primary interest is upon programmed instruction, or at least systems in which idiosyncracies of the individual instructor play a minor role.

IV

There are, however, several sections which are of broad relevance to project directors, project staff and evaluators. The initial section defining the area of evaluation presents a client-centered and practical, rather than scientific point of view, that is worth reading. The second section, on the value of evaluation presents a careful and balanced discussion of both sides of the question "Is evaluation worth the effort?"



"You're going the wrong way."

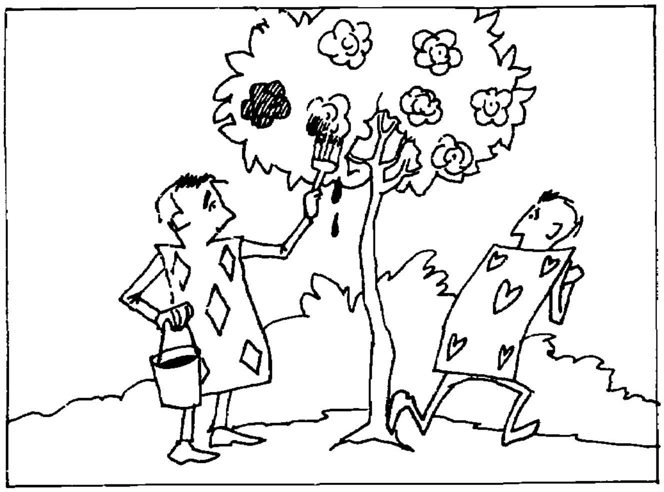
From this point the chapter then spirals into a complete emphasis on formative or developmental evaluation, targeting on techniques that have been widely used in the development of programmed instruction. The author, and the editor, agree that there are generalizable techniques that could apply across a wide range of instructional development. However, little attempt is made in the chapter itself to illustrate this generality for the reader.

Three major techniques discussed are:

1. the tutorial approach
2. large-group frameanalysis
3. the black-out technique.

The discussion of the use of pre-tests in the context of large-group frame analysis seems particularly sensible to the editor, and the issues discussed will apply to the use of pre-tests in other contexts.

IV



"Hurry, the Queen's coming on a site visit!"

Paulson identifies seven points in the learning process from point of origin, or initial pre-instructional behavior of the learner--through to the

ERIC

attainment of the ultimate objective of the instruction. These are points at which useful observations can be made to assist the improvement of the instructional unit. I think these seven stages for observation can be used to make productive revisions with almost any kind of instruction.

I regret that Bud Paulson did not include a more complete picture of evaluation, i.e., summative evaluation and various models of evaluation that are currently used across the country. However, references are given at the end of the chapter. He argues that only in highly organized instructional systems, particularly ones in which the influence of the instructor is minimal, can valid general information be tendered by an evaluation.

IV

Evaluation of Instructional Systems

Casper F. Paulson

The process of education involves a number of economic and interpersonal transactions in which certain valued resources, such as money and pupil and teacher time and effort, are expended to achieve valued effects, such as academic achievement, social behavior, personal satisfaction and well-being. Unfortunately, this system of transactions must presently be described as a barter economy, at best, because there is no agreed-upon conversion table by which all the relevant input and output values can be translated into a common currency. The valued elements themselves are frequently quite difficult to quantify. Effective operation of a small business, to say nothing of a large economy, would be most difficult under such circumstances. Prudent decisions require adequate information, and complex systems require an uninhibited flow of unambiguous information, if they are to operate effectively.

Evaluation may be described simply as the examination of certain objects and events in the light of certain value standards for the purpose of making adaptive decisions. Obviously, this is not the only definition of the term in current use. Conceivably, it is not even the best one. But it serves our purposes here. In order to explicate the definition, let us examine certain of the words.

"Examination" refers to a much broader range of activities than would be included in the term "measurement" used in the classical sense. For example, the term refers to efforts to identify value dimensions, as well as applying value standards. In many cases, it would either not be possible or not be

ERIC

desirable to reduce the object of inspection to measurement form. However, measurement will usually play an important part in the examination process.

The words "objects and events" are used primarily to achieve inclusiveness in this part of the definition. Thus they include any and all sharable sensory experiences. Excluded from our definition, but not from our concern, are completely subjective experiences and constructs not amenable to direct observation. While the evaluator may be very much concerned with the latter, he should be aware that the only way he can get a handle on them is to observe what are presumably their manifestations in real objects and events.

The words "value standards" represent a crucial part of the definition. If the evaluator hopes to serve well the information needs of his client, he must give serious attention to the values expressed by him. These values will often be expressed abstractly and ambiguously, and translating them into useful and operational form should be a cooperative effort. Ultimately, evaluation will be an exercise in futility if the evaluation information is concerned with dimensions radically different from those most crucial to the decision maker. Such information will simply not be used.

It is possible however, that the evaluator may be able to recommend instruments and techniques that get at some relevant aspect of the value dimension that had not occurred to the client. If such efforts are mutually satisfactory and provide information with decision value, they should be pursued. One should beware of the natural tendency to be concerned with that which is most readily measured, rather than focusing measurement skill on that which is of most concern. (It should be noted here that when we speak of "evaluators" and "clients" we are speaking of different roles, but not necessarily different people. Sometimes

ERIC

the need for technical expertise or objectivity dictate the use of a "third party" as evaluator. However, this is by no means always the case.)

The more extensive the behavioral science training of the evaluator, the more difficult it will be for him to accept his client's chief concerns at face value. He will be accustomed to exploring scientific constructs with scientific methodology. Evaluation, as we have defined it, involves using a rather different methodology to examine a rather different set of constructs. Failure to take a client's orientation seriously will lead to a situation where the evaluator is quite satisfied with the results, but the client finds them neither credible nor relevant to his needs.

It should be noted that we are concerned with both input and output values. These may include factors that are not readily translated into dollars, as for example, student time and effort. Some values are intrinsic and some derived. The happiness of a child may be an end in itself, requiring no outside justification. The noise of excitement that often accompanies creative activity may be valued only because it indicates that a desired process is indeed occurring. Some objects and events have both intrinsic and derived value. In addition to its intrinsic value, an engagement ring signifies both a past sacrifice and a future commitment.

The concept expressed here, that evaluation information should serve the needs of a decision-maker, is not universally accepted. The nuances of this debate need not concern us here. There are those who feel evaluation information should not be directed at any specific decision-maker target but should rather be an accumulation of information "for the record." This approach assumes that such information will be assimilated in a body of literature that will ultimate-

ly have some utility. Unfortunately, these reports are not very widely read, except by those directly concerned and their immediate families, nor do they have broad applicability.

On the other hand, there are those who believe that the evaluator should not stop with supplying information, but should himself proceed to render a judgment. This view assumes the existence of some ultimate set of values possessing broad, if not universal, validity and acceptability. Even if we felt so enlightened, we would hesitate to challenge the tradition of local autonomy so long cherished and such a vital element in American education.

In any event, the evaluator should know who and how he intends to serve. It is our choice to serve the educational decision-maker, by supplying him with the information he needs to make "adaptive" decisions. We cannot presume that these will be perfect decisions, only that they will be better than chance and better than he was able to make without evaluation information. It is worth commenting here that such decisions should be sufficiently better that they justify the cost of evaluation.

Implications

One rather obvious implication of the above discussion is that the decision oriented evaluator has a very restricted and readily identified audience for the information that he is to supply. Very early in the evaluation process, the evaluator should identify who it is that are making the decisions, what kind of information they want, what kind of information they need (these are not always the same), what kind of alternatives are considered in making the decisions, and to what extent can new alternatives be generated.

ERIC

Typically, educational decisions are made "on-line." Classes must be met, institutes must proceed on schedule. While instructional systems are typically prepared "off-line", with respect to the instructional programs of schools, they are usually very much "on-line" in terms of some developmental schedule. This presents certain reality constraints with respect to the information the evaluator is to supply, which include timeliness, manageability, and credibility.

If evaluation information is to have impact on the quality of a decision, it must be available at least at that point in time when the decision is made. In many cases it is simply not humanly possible to develop, refine, validate, administer, score, summarize, and report the results of sophisticated test instruments within the time constraints set by the needs of the given project decision-maker. Sophisticated measurement people often get queasy at the prospect of developing and using what they consider to be sloppy and unsophisticated techniques and instruments, and avoid their use altogether. While this is a laudable posture in the basic sciences, where faulty findings may have far-reaching and costly effects, it is unacceptable in the evaluation context. It is equivalent to saying that it is better to make decisions blind than to make them on the basis of somewhat tentative information. Our concern is not that all decisions be perfect, but simply that the net affect of a number of such decisions be better than it would be in the absence of information we can supply.

Decision-makers are busy people, with very specialized talents. They may not be familiar with the jargon of educational pyschology, and they will usually be unwilling to wade through a veritable Sears-Roebuck Catalog of evaluation information to find that particular bit of information that is relevant to their

immediate need. Thus the information presented must be readily understandable, with a minimum of distracting information. If one has to search for a "needle in a hay stack", it isn't likely that he will be able to make a "stitch in time".

It is not unusual, though it is somewhat dismaying, for readily available and understandable information to be completely disregarded in the decision process. Often this can be attributed to the fact that the information lacked some quality of persuasiveness, urgency, or believability. If we were of the "judgment" school of evaluators, then we would have to give some serious attention to identifying our value standards. When we leave this judgment to the educational decision-maker, it is crucial that we make a concerted effort to determine his value standards. If we have done this job well, he will be considerably more likely to make use of the information we supply.

What Is the Value of Evaluation?

If the purpose of evaluation is to determine and demonstrate that the observed achievements of an instructional effort are somehow commensurate with the costs, then it would seem appropriate to apply the same tests to the evaluation process itself. There are those who believe that evaluation activities are parasitic to instructional efforts, and should be performed only under coercion. Often experience has provided sound basis for such beliefs.

Others feel obligated to contribute to the scientific literature a descriptive account of the impact of their efforts, to be read and used at some future time by an audience whose size and needs cannot be presently determined. Since the ultimate value of such information is indeterminable, it is difficult to determine the appropriate amount of evaluation effort to be expended. Frequently the extent of the evaluation activity is determined by the amount of

uncommitted project funds available, or by application of a tithing principle that assumes that all righteousness will be fulfilled by the allocation of a certain percentage of project funds to evaluation.

An important implication of decision-oriented evaluation is that it usually cannot be justified on the basis of its assumed contribution to the scientific literature, since its specific purpose is to serve the information needs of a specific decision-maker at a specific point in time. Thus it must be justified virtually entirely by its contribution to the quality of a certain set of decisions. While this orientation provides no ready formula for determining whether and how much evaluation effort to apply, it does suggest crucial areas for consideration.

Both information and ignorance have their costs. The costs of evaluation information are much easier to determine than the costs of ignorance of such information. In the absence of evaluation, there is no way of determining the cost of inappropriate decisions. In all candor, however, we must concede that in some situations it will cost considerably more to evaluate than may be saved by using evaluation information.

Under what conditions, then, is the contribution of evaluation most likely to exceed its cost? Costs, in terms of "lost learning" and inefficient use of student and teacher time and effort (we don't mean to imply that these are the only costs) increase directly with the size of the audience affected. Take, for example, a situation where a choice is to be made between two spelling programs, one of which is 20% more effective than the other. The cost of an incorrect selection depends directly upon the number of students who will use the program. If one thousand students are involved, an incorrect decision

costs ten times as much as if only one hundred students were involved. The effect is the same as if ten times as many bad decisions had been made.

There is an assumption here, however, that the evaluation information has a consistent kind of relevance across all situations. In order for evaluation information to have this consistency, it is necessary for the thing being evaluated to be similarly consistent. In a number of education situations, this consistency is unattainable, because the unique influence of teachers and situations cause radical variations in the nature of the outcomes.

Instructional systems, however, are relatively less vulnerable to unique teacher and situational influence, and consequently they represent an ideal opportunity for rigorous and systematic evaluation. Instructional systems may be defined as replicable configurations of instructional stimuli and other prescribed instructional events, possessing described and predictable learning effects. Not only do they possess the necessary characteristic of stability, but to the extent that they are media-oriented, or otherwise replicable, they may be used appropriately with large audiences, thus reducing the per unit cost of evaluation activities. It is feasible to expend a considerable amount of effort in developmental evaluation to achieve optimum effectiveness of the instructional system, and if it is intended for a broad audience, a considerable responsibility is incurred to describe in detail the fully developed program to be distributed.

Evaluating Instructional Systems

It should be noted that this opportunity for, and emphasis upon, evaluation did not occur concurrent with the development of the new media, but rather with the development of programmed instruction and instructional systems. There is a

vast difference between mediating information and mediating instruction. Mediating information as is typically done in print, film, television, and tape, provides only the initial informational stimulus for a learning situation. The mediation of instruction utilizes these tools as part of an organized system constituting not merely instructional stimuli, but also prescribing student activities, providing a means for appraising these, and anticipating or reacting to student difficulties, with an extent of comprehensiveness that leaves little to chance. We can hardly find a better place to begin a discussion of the practical aspects of evaluation than in the context of the ideal case represented by instructional systems.

In 1963, Cronbach (Cronbach, 1963, pp. 672-683) advocated that evaluators focus their attention on means of identifying deficiencies and proposing improvements for existing curricula, rather than upon simply appraising the net affects of such curricula. Developing this concept, Scriven (1967) proposed that two evaluation roles be identified, which he identified as "formative" and "summative". Formative evaluation is the process by which information is used to develop a unit of instruction to the point where it is ready to be broadly applied. Summative evaluation is the process of describing the effects of such fully developed units of instruction. We shall now concern ourselves with the rationale and applications of certain approaches to formative evaluation.

Formative Evaluation

In the instance of an unevaluated unit of instruction, there is an assumption, however tentative, that a certain planned course of instruction

ERIC

will lead to some desired ultimate outcome. The latter may be beyond the scope of accessibility, in that it may refer to a remote condition such as economic productivity in adult life. Even if information on this kind of outcome could be obtained, the lag time between presentation of the instruction and observation of the consequences would represent an extremely long and inefficient feedback loop, virtually useless for developmental purposes. About the best feasible approximation to such ultimate outcomes is an end-of-unit measure that might shed some light on how the unit should be modified for its next application. While such measures represent only indicators of the likelihood of achievement of the ultimate aim, they are within the scope of the project and usually represent feasible attainments reasonably attributable to project activities. As such, they may legitimately qualify as "objectives".

Unless the unit of instruction is extremely small, the findings of such end-of-instruction instruments are extremely difficult to interpret, because of the multiplicity of factors and elements included in the instruction. This difficulty has lead some, particularly those in programmed instruction, to focus their attention on the immediate responses of students to small elements of instruction, such as program frames, which may be only a sentence or two in length. These approaches represent opposite extremes, and as such, they have certain deficiencies. It now appears that with the evolution of the systems approach to the development of instruction some of these deficiencies are being overcome.

The chain of events that leads from the introduction of an instructional stimulus to the ultimate attainment of an instructional objective is not a single contingency, or one "if ... then" statement, but really a long series of such contingencies. Many of these are not amenable to observation, and even

if they were, the task would soon prove unmanageable. There are, however, a number of points in this sequence where systematic observation can be quite meaningful ful and useful. One such sequence can be described as follows:

Point of origin

Instructional stimuli

Process indicators

Learning indicators

Diagnostic indicators

Criterion indicators

Ultimate values or goals

Point of origin. In many cases it is important to have a well-defined point of origin which describes those conditions that exist at the point of introducing the instructional system. Since the instructional systems are specific tools, rather than general, they are usually not intended for use with all students, but with a well-described population of students. Thus this population should be described with respect to all characteristics most relevant to the project, and may include, for example, I.Q., reading ability, mathematics aptitude, attitude toward science, etc. If, perchance, the group of students actually working through the program deviates to some extent from the intent of the design, at least the extent of this deviation should be known. If one wishes to infer the extent of change accomplished in students by using the program, then it is also important to get pre-test achievement data comparable to the post-test achievement instrument to be used. It should be noted at this point, however, that obtaining pre-test data is not always necessary and is sometimes undesirable. Some of the considerations in determining whether

ERIC Full Text Provided by ERIC

or not to pre-test will be discussed in a later portion of this paper.

Instructional stimuli. The next point of concern is the instructional stimuli. Various activities may be involved here. Proofreaders may check for spelling and punctuation errors. Content specialists may be concerned with questions of accuracy and adequacy of the subject matter. Educational psychologists may appraise the extent to which the presentation is consistent with learning principles. Presumably, each of these specialists has a sound rational basis for the judgments that he renders. The judgments lack a certain empirical or pragmatic quality, however, because nobody knows yet whether the materials will actually work with a given set of students.

Process indicators. The most important member of the developmental team is uniquely qualified, not by his expertise, but by his naivete. That is the student who does not know the material that one is trying to teach. Only he can tell whether the terminology used is within his vocabulary, whether the narrative is understandable to him, where he encounters unsurmountable difficulties, and the specific nature of those difficulties.

It is precisely here that programmed learning developers have made one of their most significant contributions. Careful, step-by-step observation of students working through programs has enabled the developers to isolate and remedy learning difficulties as they occur. The class of student behavior that is described as "process indicators" includes such things as student responses to individual frames of program instruction materials, note-taking behavior, or even visible sign of boredom or excitement. They do not necessarily indicate that learning is going on. What they do indicate is whether or not certain processes presumed to facilitate or cause learning, like active student involvement, exercise, or drill, are actually going on.

ERIC

Learning indicators. Next in the sequence is that class of events, called "learning indicators", that indicate actual learning achievements presumed to be landmarks on the way to the ultimate objective. The student who has demonstrated that he can spell "Argentina" presents some small glimmer of hope that he may ultimately be able to write an acceptable report on that country. These learning indicators may be obtained as the instructional process is occurring, and subsequent instruction modified as needed.

Diagnostic indicators. While the indicators described in the above paragraphs may be considered to be diagnostic in that they enable problem areas to be localized, there is also the familiar procedure of diagnostic testing which may occur during or following the administration of an instructional program. A single instrument or battery of instruments is administered in order to collect efficiently at a certain point in time the kind of information that would be necessary to localize problem areas. This approach is most useful when the procedures described above are not feasible.

Criterion indicators. The concern of criterion tests is focused upon the extent to which students can demonstrate achievement of the terminal objectives of the unit of instruction. These should be differentiated from diagnostic tests in that the items contained are derived directly from the terminal objectives, not from enabling objectives. The purpose is to describe simply the probability of students being able to demonstrate a criterion level of performance after instruction, not to describe the condition or locate the difficulty of students not achieving criterion.

Ultimate values or goals. "Ultimate values and goals", as noted before, usually cannot be observed directly. They serve here chiefly to identify the

source from which all the other indicators are derived.

We have described a variety of classes of observations, and their sequential relationship, to prepare a background for discussion of formative evaluation practices used currently or recently.

The Tutorial Approach.

An approach used frequently by developers of self-instructional materials is simply for the developer or evaluator to sit down with the student as he works through the material and note carefully the student's reaction to each step or frame in the program. (For those readers not familiar with Programmed Instruction jargon, a "frame" is a small "bit" of instruction, frequently only a sentence or two in length, followed by a question which requires some response from the student.) Whenever a student encounters difficulty in working through a frame or making a response to it, he describes his problem to the "tutor" who, in most cases, will immediately revise the frame. While this procedure may seem to be simplicity itself, its success depends upon subtleties difficult to translate into verbalized principles. Susan Markle (1967, pp. 122-123) describes current practice in this way:

> Procedures for eliciting these data vary. Some testers prefer to talk to the student throughout the process, a procedure which, of course, renders the student's final performance suspect, if not invalid. Others prefer to query the student who hesitates or errs, leaving him to his own devices when no danger signals are apparent. The data which may be missed under this condition are exemplified by statements which some of us have heard often: "I know what you want here, but..." and "I see your point, but it seems to me...". There are at present no firm rules. Each programmer has his own.

Unfortunately, though experience may be the best teacher, the tuition is frightfully high. Nor will recounting a diverse set of principles proposed

by various programmers obviate the need for such experience. The present state of the art does not permit this.

In order to provide a concrete referent for the general concept of "tutorial approach" and in order to provide the reader with some understanding of at least one methodology within this classification, we shall describe one of the better defined methodologies, without implying that this is the best or only one.

Robert E. Horn (1966) has developed a self-instructional program entitled Developmental Testing for the purpose of training evaluators or programmers in the tutorial approach to formative evaluation. For this he is to be commended in that he has used the programming format to instruct about programming technique, an undertaking surprisingly rare in the field of programmed instruction. Embarrassingly, our discussion of the merits of this approach are based solely upon our experience and examination; we have no empirical data, nor were we supplied any by the author. At any rate, proceeding upon what we consider to be the valid assumption that it was insufficient to simply discuss elements of technique, Horn has prepared a program in which students face realistic simulated problems in program development, and receive feedback on the appropriateness of their solutions to these problems. It is the opinion of this writer that in doing so, Horn has proceeded one necessary step beyond informational description. Thus, while we cannot presume to achieve through description that which this program achieves through instruction, it may be helpful to inform the reader of some of the principles and guidelines contained in Horn's instruction. These are displayed in Figures 1 and 2.

1. The programmer should first explain to the tryout student that the materials he is to be given are intended to help him learn subject matter designated in the title.

2. The programmer should emphasize that the role of the student is to help the programmer evaluate some new education materials. Comments and suggestions that the student makes will help the programmer make revisions.

3. The programmer should then explain that he has to know how much the student already knows about the subject matter and whether or not the student has all of the prerequisites to learn from the materials. He should then give the student the pre-test (always) and the prerequisites test (if required) timing the student on both. Both of these may be done when the test subjects are being selected.

4. When the tests have been completed, the programmer should show the student the program and explain again that it is the material, not the student, that is to be tested from now on. This is an especially important point about which the student should have no question.

5. The student should be given a ball point pen with which to write his answers. (This will prevent him from erasing potentially valuable information for revising the program). He should be provided with answer sheets, if any.

6. Tell the student to put an "X" next to the items he thinks he got wrong after he has checked his answer. If the program contains open-ended questions, tell the student about this.

7. Explain to the student that if he doesn't know an answer, he should take a guess and write "guess" on the answer sheet. If he simply can't think of an answer, he should leave the answer blank and place an "X" next to the item on the answer sheet.

8. Tell the student the time limits placed on the tryout session and that he can take a break whenever he feels like stopping.

9. Re-emphasize that any comments he wants to write or express to the programmer will be useful and welcomed.

10. Then ask the student to commence with the materials. (If the student asks what he should do or asks if he's doing it right, the programmer should gently insist that all the directions necessary are given in the materials. It is important to try out the directions, too.)

11. The programmer should note carefully the time at the beginning and end of each tryout session and keep track of "break time".

Figure 1. Checklist for the First Tryout Sessions (Horn, 1966, p. 6)

Principles

I. If the student can continue through the program even though he has difficulty with an item, it is best to let him continue. Ask him about the difficulty at the end of the tryout session. Watch him very carefully for three or four frames. If he's consistently in trouble, it may be well to interrupt.

II. If the student has so much difficulty with an item that he cannot proceed with the rest of the program, the programmer should intervene. His first step should be to try to revise the program on the spot, presenting a revised or new item to the student. This may be done orally or the programmer may make written changes in the program. He should do this revision with a minimum of explanation to the student.

III. If these on-the-spot revisions do not work or if the programmer can't figure out the difficulty, he may then query the student directly with such open-ended questions as: "Will you tell me about the difficulty?" or "What seemed to be the trouble with this item?"

Figure 2. How to Intervene in the Tryout Process (Horn, 1966, p. 12)

We believe that it would be a mistake for the reader to conclude that simply reading these guidelines and principles constitutes an adequate summary and a viable substitute for use of the program. The program itself is rather inexpensive and, we believe, effective, and those presented with the task of actually undertaking a formative evaluation using this approach could be well advised to acquire and use this program.

General comment on the tutorial approach. Markle (1967, p. 123) discusses three kinds of problems commonly associated with the kinds of evaluation discussed above:

1) It is a costly and time-consuming procedure.

2) The subtlety of the "tryout editor's" task, and the difficulty in describing it, and performing it.

3) Representativeness of the students to be observed.

Like Horn, Markle uses the term "developmental testing" rather than "formative evaluation", but the terms are roughly comparable and her comments are particularly applicable to the tutorial approach presently being discussed. With respect to the first problem, Horn believes that his approach can be used effectively with as many as four students at one time. Doing so would tend to alleviate the cost and time requirements. Some of the procedures to be discussed later will go even further in this direction. Horn's procedure also speaks to the second problem by providing specific, if arbitrary, procedures for the "tryout editor" to follow. While the effective implementation of these procedures will still require considerable good judgment and intelligence, the process will be a good deal less chancy than it otherwise might be. With respect to the third problem, Horn claims that this difficulty can be alleviated considerably by proper choice of subjects. He advocates the use of one student of relatively low ability, one student of average ability, and one student of high ability.

In general, the tutorial approach has certain clear-cut advantages. It is very sensitive to specific learning problems as they appear. Like some of the procedures to be discussed later, it contributes to effectiveness largely by identifying the more serious inhibitors or impediments to learning, such as mistaken assumptions about the learner's vocabulary, and ambiguity or lack of clarity in the narrative. It is diagnostic only in the sense that it identifies discrete process difficulties, not more complex learning difficulties. It assumes that, in general, the pathway to the terminal objective is correct, and functions simply to remove rocks and debris from the trail.

ERIC Full Text Provided by ERIC

Unfortunately, this approach also has certain deficiencies. As discussed above, it is relatively more vulnerable than other approaches to the possibility that the students employed are unique, rather than typical. Consequently, developmental changes may to some extent be cyclical, rather than incremental. If information from a second trial tends to contradict information from a first trial, it is difficult to feel that progress is being made.

The more serious difficulty with the tutorial approach is that, while it is quite sensitive to communications problems, it is frequently blind to learning problems. Unless used in conjunction with other techniques, the information obtained indicates only that a certain desired interaction of the student with the materials is going on. Other than this it tells little or nothing of actual progress toward the terminal objective, in terms of what has been and what will be learned. The approach also tends to be insensitive to oversimplified, and hence inefficient, instruction. While students may occasionally comment that frames are too easy, in the typical case they will work through them without comment. This problem may occur when the frame size is too large or step size is too small. Frame size is determined simply by the number of words in a frame. Step size is determined by difficulty. Thus, a small frame may constitute a large step simply because the student has to make a response on the basis of a limited amount of information. It would, however, be an oversimplification to say that the one is simply the inverse of the other. Large frames may be loaded with relevant content, or extraneous details. Small frames may demand considerable thought on the part of the student in order to make a correct response, or they may be so easy that the response is considered trivial.

ERIC

Since students are less likely to comment to their tutor either about the triviality the responses that they are making or the possible extravagance of extraneous material, correction by amplification of the content, or "stepping out" the instruction is more frequent than correction by simplification or deletion. Consequently, there is a tendency for revisions to be longer than the initial draft. Since there is no balancing tendency toward shortening the program, revisions may tend to compromise learning efficiency as they enhance learning effectiveness.

Large-Group Frame Analysis

An alternative or supplement to the previously described tutorial frame analysis and revision is to try out complete sequences of programmed material with a large group, say 20 or more students, with provision for recording and analyzing their responses to individual frames or instructional elements.

While this approach may be adapted for use in a variety of situations involving systematic instructional efforts, it is most easily described and most typically employed in programmed learning development, since such situations are structured to elicit student responses systematically and continuously.

It is often just as easy to get complete classes to work through experimental trials of programs as it is to get individuals, and it often involves even less direct cost to the evaluator. If the materials are relevant to public school curricula, they may be presented as an integral part of the instruction. Those who object to the use of such untried or unvalidated materials should recall that typical classroom instruction is subject to the same criticism.

The advantage of this type of evaluation is that the instructional activity involved is more similar to actual conditions in which the material will ulti-

mately be used than is the case with some of the other evaluation procedures discussed. Students are not continually interrupted by the necessity to comment on their progress, or their learning difficulty, nor is their attention focused on the "trouble-shooting" nature of their participation, or the tentative, developmental nature of the instructional materials. The data obtained by this procedure are less vulnerable to unique personal characteristics, since the responses of a number of students to each frame or learning element are considered simultaneously. Thus it is possible to get more accurate data on the probability of correct learner responses to individual frames. Since no revisions are attempted until all the responses and other data have been collected, it is possible to get a much clearer profile of program effects, and thus to marshal a considerable amount of information in making decisions about program revision.

Procedures for collecting response data. Since this procedure makes extensive use of student responses and reactions as they work through the program, we shall first discuss appropriate means for obtaining such data. While it may be the ultimate intent to have students write their response in a program book, or only to make covert responses, where they simply think but do not write their answers, it is desirable at this point to construct separate response booklets or sheets on which students will write responses to each frame or element in their learning program. The assembly and format of the response booklet should be such that it is easy for the students to use, with a minimum of distraction or manipulation while they proceed through the program. There should be similar concern for ease of scoring and analysis by the evaluator. This is the chief reason for having responses recorded in a separate response booklet. An example of such a response sheet is presented in Figure 3. (Paulson, 1963, p. 41).

Name ____________ SET 2: TIME: Begin ______
Stop ______

MOVEMENTS OF THE EARTH: ROTATION & REVOLUTION

2-1 ____________________
2-2 ____________________
2-3 r o _ _ t i o n
2-4 r _ _ _ _ e s
2-5 c __________________
2-6 a __________________
2-7 __________________ s
2-8 a _ _ _
2-9 On back of page
2-10 ____________________
2-11 ____________________
2-12 ____________________
2-13 ____________________
2-14 4 ____________ tion
2-15 ____________________
2-16 ____________________
2-17 ____________________
2-18 e __________________
2-19 ____________________
2-20 ____________________
2-21 ____________________
2-22 _ _ _ _ l u _ _ _ _
2-23 ____________________

2-24 and 2-25 On back
2-26 ____________________
2-27 ____________________
2-28 __________ P ________
2-29 ____________________
2-30 S _ _ _ _ _
2-31 ____________________
2-32 ____________________
its ______________
2-33 ____________________
2-34 ____________________
__________ & _______
2-35 ____________________
2-36 ____________________
2-37 r __________________
r __________________
2-38 ____________________
2-39 r __________________
2-40 r __________________

(front)

SET 2

Please do each drawing in the box provided

2-9

2-24

2-25

(back)

Figure 3.

Obviously, each response blank should have a number corresponding to the item number in the program, but in addition it is desirable to include in the response sheet all cues presented to the student in the program. Thus if an initial letter cue is given in the program, that cue should also appear in the response item. The number of blanks in the response item should correspond with the number of words required in the response portion of the program. This will be helpful to both the student and to the person scoring the response sheet. Also, if commas, conjunctions, or other short contextual word cues are employed in the program, they should be included in the response sheet. In the case of branching programs, where students select from alternatives the answer they consider most appropriate and turn to the page that accompanies that answer, the alternative page numbers should be listed in the response booklet, with students directed to circle the page number they chose.

It is often convenient, and thus tempting, to collect data on the time required by each student to complete each segment of the program. However, in the formative stages of the program development this information has little utility in making revisions. It is useful in evaluating finished programs, but should then be collected during administration of the final draft of the program.

In addition to collecting the actual responses to each frame, it is very useful to have students record their comments with respect to specific difficulties that they encounter in working through a frame, and also to record general comments about the program. It is desirable to collect the latter at the end of each work session, while impressions are still in the minds of the students.

ERIC

Accompanying the response booklet should be a specific set of directions. If students are unfamiliar with the programmed learning format, this should be explained. For such students, and for all students in the lower grades, it is desirable to have sufficient practice exercises that the examiner can be assured that students understand and follow the instructions. It is often useful to explain to students that you are testing the program, and not them. It is important that students record all their answers, and if they choose to change an answer, they should mark out but not obliterate their initial answer and write their next answer beside it. Recall that this injunction can be facilitated by supplying the students with ball point pens or pencils with no erasers.

Once the students have started working through the program, it is desirable to monitor the procedure closely, in order to assure that they are following directions closely and recording all their answers. A common difficulty in program tryouts is that some students will copy the feedback answers into the response blanks rather than think out their own responses. While it might appear that this would be detrimental to learning, research evidence, surprisingly, does not support this as long as students at least read the frames. Unfortunately, however, such a procedure renders response data useless and for this reason it should be discouraged.

Occasionally students will come to the tryout monitor with questions. When this happens one should try to determine the exact nature of the difficulty and subsequently record it. However, answers to such questions should usually be restricted to the minimum amount required to permit the students to continue working through the program. The trial seeks to determine whether the program

can stand by itself. If it can't, revision, not augmenting instruction, is needed. While it may take a little hardness of heart to carry out this procedure, it is more likely to reveal needs for revision that will ultimately benefit many students.

It is quite likely that the students' comments and reactions will focus upon conspicuous defects and "nit-picking" errors in the program, such as incorrect spelling or punctuation, or typographical errors. If their frame of reference in playing the role of constructive critic emphasizes this copy-reading function, which could better be left to professionals, they will devote little or no effort to describing the more subtle learning difficulties they encounter. Thus, unless the program is exceptionally "clean", it may be advisable to precede the main trial of the program with a pilot effort on a portion of the group, or a different group. The objective of such a pilot would be to render "first-aid" to the program, in which conspicuous defects and readily correctable errors may be dealt with quickly before the principle trial is commenced. The "first-aid" trial should not be considered or treated as a major revision effort.

Design and implementation concerns. Since major revision activities do not commence until all the response data have been collected, it is feasible to marshall a considerable amount of supplementary information which may also be employed in the revision process. Such information may include post-tests to indicate whether the behavioral objectives of the instruction have been attained, the same or similar instruments administered as pre-tests to indicate whether the objectives had been attained prior to instruction, tests of general abilities such as intelligence and reading relevant to achievement

ERIC

and measures of prerequisite skills necessary to benefit from using the program. In determining the nature and extent of such data collection, one should bear in mind that its sole purpose is to facilitate the revision process. Formative evaluation efforts seldom make a direct contribution to the scientific literature. Thus, the normal zeal for collecting information should be tempered with prudent judgment regarding its cost and utility for the revision process.

Dick (1968) has proposed and tried a method for integrating and using data from varied sources in the revision process. He describes the method as follows:

1. Study the item analysis of the end-of-lesson test to determine those concepts which were most often missed by the students.

2. Study the incorrect responses to these particular test items to determine if there was a straightforward misunderstanding of notation, a complete lack of comprehension of the concept, or a variety of errors.

3. Use the guide to determine those frames in the program which dealt most directly with the concept(s) missed on the test.

4. Study the student error rates for these frames. If the program frames are quite similar to the test item, and the error is quite low, more practice frames should be provided. If the error rate is quite high, these frames need revision.

5. Study the sample of incorrect student responses to this segment of the program. These responses should suggest the nature of the learning difficulty and the type of revision needed.

6. Study the comments of both the students and the program reviewers for further suggestions concerning the problems encountered with these particular frames.

7. If no frames in the program correspond to a test item missed by a large percentage of the students, consider the addition of frames that will "bridge the gap" between the present learning materials and what would be considered a transfer type item.

The actual trial of this method was revealing, and is particularly relevant to our discussion of the large-group and tutorial methodologies. Dick summarizes the results in this way:

> The general consensus among the writers was that the frame error rates and reviewer comments were the most informative data. If the error rate became excessive (which depended upon the writer's own point of view), the incorrect responses to the frames and the student comments were studied. Few of the writers studied the results of the end-of-lesson tests or related test item performance with particuar frames in the program.
>
> It was clear that none of the writers had followed the suggested sequence through the data. The writers reported that the end-of-lesson tests, which they had not constructed, did not completely represent the objectives they themselves would have tested. The writers did indicate an interest in the students' overall impression of each section of the test, e.g., impressions of continuity, readability, etc. They were also interested in knowing more about the ability level of the students who had made specific comments about segments of the program.
>
> It was of interest to the author to note that when the writers were given a hypothetical alternative of gaining information about the program by going through it personally with three or four students vs. gaining statistical data from 40 or 50 students, the latter procedure was much preferred. There seemed to be a greater acceptance of a more limited type of information from a greater number of students (which appears to provide greater generalizability), and an acknowledgement of difficulty of obtaining suitable guinea pig students.

These findings highlight both the strengths and the potential deficiencies of the large-group method. Apparently, specific methods of translating and transmitting data that are presently little used must be developed, or there is little point in collecting the data in the first place. Careful consideration must be given to what kind of data to collect, and how that data should be presented. Dick's paper is recommended reading, both because of its candor and conciseness, and for helpful suggestions he offers for future efforts.

Post-tests. While it would seem that even the most minimal effort to collect supplementary information would include post-test data regarding the

achievement of the behavioral objective, even this assumption is open to some question. Though it is almost self-evident that such information would be useful, the ultimate question is whether it is actually used. As noted previously, Dick (1968, pp. 99) found that in the revision of a calculus program, test item data were seldom used. It would be comforting to believe that this finding was unique and exceptional, but more probably it is not. First, the typical philosophical and theoretical orientation of programmers implies that if each step on the way to criterion performance is achieved, criterion must also inevitably be achieved. Thus, attention is focused primarily upon the steps, not the goal.

It may also be that something about the operating procedure in revising programs makes the use of post-test information unfeasible or at least difficult in the revision process. Typically, program revision involves retracing the program frame by frame, modifying it where the need is apparent. Response data regarding errors on each frame are readily available. On the other hand, it is hardly feasible to reexamine the final examination in the process of revising each frame.

If the factor inhibiting the use of post-test data is a relatively unchangeable characteristic of programmers, then there is little hope of remedy. On the other hand, if post-test information can be adapted and supplied to programmers in a form that meets their procedural requirements, there is indeed some hope that post-test information will enhance the quality of the revision process. Procedures for doing this will be discussed in a later section.

Pre-tests. Almost as broadly accepted as the need for post-testing is the desirability of pre-testing the same students with the same instrument, though

ERIC

there are a few dissenters. If the pre-test is identical, or at least equivalent, to the post-test, and if it is valid to assume that the post-test measures something that students are able to do after instruction that they were not able to do before instruction, then such a pre-test inevitably represents a failure experience for students. As such, it represents a rather inauspicious beginning to their learning experience with the program. Fear that this would impair program performance is not altogether unreasonable. Further, discouragement at attempting an unreasonably difficult task, supplemented by awareness that they are not expected to do too well, may lead to apathy during the pre-tests and a consequent underestimate of their true ability to perform.

One way to minimize the undesirable affects of pre-test difficulty is to arrange test items in order of difficulty, if this is possible, and to precede the items to be scored with a number of "lead-up" items that are easy enough for almost all students to answer correctly. This gives students a "running start" in the test, and may lead them to making a more serious effort throughout the test. It is also important to urge the students, both before and during the pre-test, to make their best effort. These easy "lead-up" items should not be scored or analyzed, since they will tend to obscure more than clarify interpretation of the data.

Another argument against using pre-tests is that they focus learning in an undesirable way and develop a set toward learning certain specific program content. If experience from a pre-test tells the student precisely what he is to learn from the program, this experience may reasonably be expected to affect his post-test performance. One way to minimize this problem is to generate a large pool of post-test items and then divide them

ERIC

into two forms of the test. This division process may be random or systematic, depending upon the way the data will be analyzed and used. Half of the pre-test group can be given Form A, the other half Form B. It would be desirable to constitute these two groups randomly. At the post-test, the students would be administered the test opposite that which they had for a pre-test. While this procedure is useful for making group decisions, it has certain obvious deficiencies in measuring individual pre to post gains. Often, however, such gain scores for individual students would have little utility in the revision process.

In some cases, the nature and content of the instruction is so unique that it would only be reasonable to assume pre-test scores of zero or not significantly higher. If one can afford not to test that assumption, then pre-testing in such situations makes little sense.

It is frequently the case in public school instruction, however, that students vary considerably in relevant knowledge that they bring to a program. In such cases it is highly desirable to know how much of their post-test performance is attributable to the program, and how much to prior knowledge. Use of the pre-test in such situations may make much clearer to the programmer the actual entry level of students working through the program, and thus facilitate revisions early in the program. Also, if one is concerned about incremental gains toward the stated objectives, as well as the probability of achievement or nonachievement of those objectives, then pre-test information will be necessary. In general, the simplest way to describe the usefulness of the pre-test is to say that it makes post-test findings much easier to interpret.

Entry tests. Occasionally tests that do not meet the foregoing description of "pre-tests", but that serve other useful purposes, are administered before students work through a program. While these also are sometimes referred to as "pre-tests", they serve quite different purposes. In some cases they determine the presence or absence of certain essential prerequisite abilities, or the extent of certain skills considered highly important for program achievement. In some cases they are used to screen out those for whom the program was not designed, for example, those who have inadequate reading ability. In other cases, they are used to determine differential effects of the program on students who vary considerably with respect to some relevant characteristic.

Use of the pre-test, as it was earlier described, that is, one similar to the post-test, as a screening device is usually a questionable procedure. It is legitimate to use such a test to screen out those students who really do not need the program because they already have the skills represented by the objectives. Their use is also appropriate to determine where in the program students should begin. However, selecting high scorers on such a test as the only ones to take the program amounts to an admission that the only students who can profit from the program are those who already know much of what the program has to offer. Such a practice also has effect of minimizing pre-test to post-test gains. Thus it may tend to make the program appear to be less effective than it really is, or could be.

Record data. In the ideal case, the programmer and evaluator have a very clear concept of what characteristics the entering students should possess. When this is the case, one can proceed forthwith to develop an entry test

ERIC

to determine whether candidates for the program actually have the prerequisite skills. Unfortunately however, prerequisites are elusive and typically ill-defined. A more realistic approach may be an attempt to determine pragmatically and after the fact, which students with what characteristics are unable to benefit from the program. Thus, until the concept of prerequisite requirements has been clarified, perhaps the most effective procedure is to utilize the information obtained from conventionally administered standardized tests that appear to have some relevance to program achievement. Typical candidates for such use will be intelligence tests, tests of other general abilities like reading, and standardized tests in the subject area. The data obtained from such measures can be related to program data to determine the differential impact of the program on various levels of students with respect to these other relevant characteristics.

A further advantage of use of data regarding general characteristics of students is that these are usually available from student records and do not require any special testing immediately before administration of the program. Not only is such special testing taxing for the student, but it also calls his attention to the experimental nature of the program administration, thus increasing the so-called "Hawthorne Effect". A rule of thumb for the evaluator should be to keep his data collection procedures as unobtrusive as possible, and to interfere as little as possible with the ordinary train of events.

The data collection procedures described above are by no means exhaustive. Little has been said, for example, about the use of subject matter specialists to review the program, or the use of the programming or writing specialists. If it is economically feasible to bring this kind of talent to bear in an evaluation effort, it certainly merits consideration. What has been attempted

ERIC

is a bare outline of common practices in this type of formative evaluation.

Summarizing and Presenting Data

If we bear in mind what has been said about the role of the evaluator in supplying information for decision making and the fact that in the context of our immediate example the decision-maker is a program writer, we will be able to derive some useful clues as to how the data should be summarized to best serve his needs. First, as we have mentioned before, program revision usually involves moving through the program in the same order in which it was written and rewriting frames as appears necessary. Thus, it would appear desirable to provide a similar frame-by-frame focus of the evaluation information.

This microscopic inspection should be supplemented, however, by information that provides a somewhat broader view of program effects. The large-group procedure presently under consideration enables one to view a profile of the effects of each and all of the frames contained in the program supplemented by general comments from students and other sources. In some cases, information will be directly relevant to the revision of particular frames. In other cases, it will be more relevant to determining a general strategy. For example, if the instructional program appears to be tedious and monotonous, perhaps a systematic variation of program style rather than rewording of frames is indicated. Thus, general and program profile information can be reviewed at the beginning of the revision process, followed by a systematic frame-by-frame reworking of the program.

The response data obtained from students working through the program lends itself well to the microscopic analysis of frames. The first step in

summarizing this data is simply to tally the number of anticipated and unanticipated ("correct" and "incorrect") responses to each frame by students.

"Error rate". In order to understand the significance of this data, and the rationale of certain procedures employed in handling it, a brief discussion of the term "error rate" is in order. The reading of the self-instructional frame and the active response resulting therefrom are both parts of the brief learning experience. When a student makes an incorrect response, two things are indicated: 1) the information did not prepare him adequately, and 2) he has performed or practiced an inappropriate response. In a linear program, the only remedial action taken is to provide feedback to the student to indicate what his response should have been. Frequent errors cause inefficient learning, so it is desirable to keep error rates at a minimum. The generally accepted error rate per frame is 5-10%. The term "error rate" has received considerable usage in programmed instruction over the past several years, though it has some deceptive implications. Rather than coin a new term we will simply voice certain reservations about the old one.

It frequently happens that a student will respond to a frame in a way that is unanticipated, but is still not wrong. That is to say, if the student were interacting with a teacher and made that response, the teacher would sense that his response was equivalent or synomonous with the anticipated response and tell him so. Thus, in correcting a program, there is a temptation to regard equivalent responses as correct, in order to be fair to the student. It should be recalled, however, that this data is not being employed to grade students, but to assess the effects of the program. While unanticipated responses may be innocuous in a number of cases, they

have at least two undesirable characteristics. First, the programmer may have intended to exercise the student in the use of a certain term in order to lead up to a subsequent learning experience. Use of an equivalent term, while technically correct, may not provide that exercise. Further, if a student's response deviates from the correct answer feedback given to him, he may either be uncertain as to the correctness of his response or wrongly conclude that he responded incorrectly.

Thus, "error rate" is more appropriately thought of as an index of program performance rather than student performance. It simply indicates the relative frequency of responses other than those anticipated by the programmer.

Tabulation procedures. The tabulation of unanticipated responses will usually be done by hand for a variety of reasons. In the case of constructed response programs, where students write out their answers, human judgment is presently required to determine the correctness of the response. Under development are devices that will automatically evaluate written responses that meet certain format constraints, but it may be some time before such devices are widely available. In the case of selected response programs, wherein students select from a number of alternatives the answer that they consider to be correct, machine scoring is feasible. For the present evaluation puposes, however, the machine scoring is desirable only if information on each item is tabulated. Further, it would be higbly desirable to know the frequency of each of the several possible incorrect responses, not merely whether the response was correct or not.

While tabulating errors is not a formidable task in itself, care should be exercised in delegating this work to clerical personnel. Those who do so

ERIC

will soon recognize the need to develop very explicit correction rules and procedures. In some frames, correct spelling is a crucial quality in the response, while in others the spelling will not matter. The sequencing of words in the response may or may not be crucial. It is virtually impossible to anticipate all of the questions that are likely to arise in the process of scoring student responses, but a little foresight can markedly reduce the frequency of scoring problems.

One way to simplify the monitoring of the tabulation task is to have the tabulators copy down all incorrect responses to frames. If these are observed to include answers that should have been considered correct, appropriate changes can be made. Recording all incorrect responses provides useful information for revision, too.

If a number of people are involved in the scoring process, it is better to have each of the individuals responsible for correcting all of a certain portion of the program, rather than having a number of total programs tabulated by each individual. First, the scoring procedure proceeds more rapidly, because individuals become familiar with the appearance of correct responses on the response booklet page, and can tell virtually at a glance which responses are incorrect. Second, deviations from desired scoring procedures are more easily localized and remedied. The alternative procedure, when complete programs are tabulated by individuals, is to have the tabulator initial those programs he worked with Then if any of the tabulation has to be redone, it won't be necessary to work through the whole batch.

When the tabulation task has been delegated to clerical personnel, it is desirable to spot check all individuals and all parts of the program.

Rechecking a handful of programs may contribute considerably to the quality of the data obtained. It is worth noting, parenthetically, that individuals vary considerably in the speed and accuracy with which they perform this kind of task. If extensive tabulation is required, it would be desirable to either administer a work sample or observe the relative efficiency of various persons in early phases of the task, and select for the major portion of the effort those that demonstrate the greatest proficiency.

The procedure for tabulating responses in branched programs, where students select from a number of alternatives the answer they consider to be correct and turn to the page accompanying that answer, is somewhat different than that described above. Students working through this type of program should be directed to circle on the answer sheet the page number corresponding to their choice. Incorrect choices will lead students through a sequence of remedial frames. In this type of program, not all students work through all frames. The programmer will want to know how many students choose incorrectly at choice points, the frequency with which the various remedial sequences are used, and the error rates for the frames in those remedial sequences. It will be necessary to tabulate both correct and incorrect responses in remedial sequence frames in order to properly calculate error rates in those frames.

Perhaps the easiest way to tabulate errors is simply to use a copy of the response booklet and place a tally mark in the appropriate blank for each incorrect response. On completion of the tabulation process, these tallies can be translated into percents representing the error rates. In order to get a profile of the function of the total program, these percents

ERIC

can be presented in table form as illustrated in Figure 4 (Paulson, 1963, pp. 25-26).

Providing a larger view. In performing the revision task, the programmer will be interested both in the detail information about each frame and in the general profile that can be discerned by examining the table. It will be seen that the usefulness of the profile information presented in Figure 4 has been enhanced considerably by examining the differential effectiveness of the program frame on low, average, and high ability students in fifth, sixth, and seventh grades. An indication of the usefulness of such profile information can be seen by careful examination of the data. Error rates of virtually all students increased sharply on the twenty-sixth frame of Set 5, and remained high until the end of that set. This indicates not merely localized frame difficulty, but a general deterioration of program effectiveness at that point, and need for strategic remediation, not merely of frames, but of a whole section of the program. It should be noted, however, that if the program were intended for use by average or better seventh grade students, a wholesale revision of this section of the program might not be necessary.

Another useful technique to gain an overview of the functioning of relatively long sequences of program frames is flow charting. While this is most useful, and virtually mandatory on branching programs, it is also useful for simple linear programs. An example of a flow chart is presented in Figure 5 (Lang and Kersh, 1967).

Usefulness of the flow chart is enhanced considerably by labeling the content of each frame. Obviously, a flow chart representing a large number

ERIC

	5 Low N=18	5 Ave. N=16	5 High N=18	5 Total N=52	6 Low N=9	6 Ave. N=19	6 High N=20	6 Total N=48	7 Low N=12	7 Ave. N=14	7 High N=5	7 Total N=31
5-1	0	0	0	0	0	0	0	0	0	0	0	0
*5-2	33	31	17	27	44	21	10	21	33	21	0	23
5-3	17	13	0	10	22	16	0	10	0	0	0	0
5-4	0	0	11	4	11	5	0	4	0	0	0	0
5-5	11	13	0	8	33	21	0	15	0	7	0	3
5-6	11	6	0	6	0	11	5	6	0	0	0	0
5-7	17	6	0	8	22	16	5	12	8	0	0	3
5-8	17	0	6	8	22	26	10	19	8	0	0	3
5-9	28	0	0	10	33	11	5	12	17	0	0	6
5-10	17	13	0	10	22	0	0	4	0	7	0	3
5-11	22	0	0	8	11	5	0	4	8	7	0	6
5-12	6	6	0	4	11	0	0	2	0	0	0	0
5-13	28	13	0	13	22	11	5	10	0	0	0	0
5-14	6	0	17	8	22	5	5	8	8	0	0	[illegible]
5-15	17	13	0	10	44	16	5	17	0	0	0	0
5-16	17	0	0	6	11	16	5	10	8	0	0	3
5-17	33	6	6	15	33	21	5	17	0	14	0	6
5-18	22	13	0	12	22	5	5	8	0	0	0	0
5-19	6	6	0	4	0	0	0	0	0	0	0	0
5-20	6	0	0	2	0	0	0	0	0	0	0	0
5-21	17	0	6	8	11	5	5	6	8	0	0	3
5-22	6	0	0	2	11	0	0	2	0	0	0	0
5-23	17	0	0	6	11	0	0	2	0	0	0	0
5-24	6	0	0	2	0	0	0	0	0	0	0	0
5-25	17	6	0	8	22	5	10	10	8	0	0	3
*5-26	28	19	33	27	78	84	50	69	25	43	40	35
*5-27	33	13	11	19	44	37	25	33	0	7	0	3
5-28	39	31	6	25	11	11	0	6	25	7	0	13
*5-29	39	25	11	25	56	26	5	23	8	7	0	6
5-30	33	6	0	13	22	11	20	17	8	0	0	3

Figure 4.

(continued on next page)

	5 Low N=18	5 Ave. N=16	5 High N=18	5 Total N=52	6 Low N=9	6 Ave. N=19	6 High N=20	6 Total N=48	7 Low N=12	7 Ave. N=14	7 High N=5	7 Total N=31
*5-31	39	19	6	21	56	16	15	23	33	0	0	13
5-32	22	13	6	13	33	21	0	15	0	0	0	0
*5-33	61	38	44	48	67	79	45	62	25	43	0	29
5-34	22	25	0	15	33	0	5	8	0	0	0	0
*5-35	44	31	11	29	33	37	20	29	17	7	0	10
5-36	33	25	11	23	44	11	5	15	8	7	0	6
*5-37	50	44	28	40	56	42	20	35	33	14	0	19
*5-38	50	38	11	33	56	21	15	25	8	14	0	10
*5-39	28	31	17	25	44	32	20	29	0	7	0	3
5-40	33	38	0	23	22	5	5	8	0	7	0	3
*5-41	44	31	6	27	22	21	20	21	0	0	20	3
*5-42	50	25	22	33	22	11	10	12	17	0	0	6
5-43	44	31	6	27	11	16	0	8	8	0	0	3
*5-44	33	25	22	27	33	21	5	17	25	0	0	10
5-45	22	19	6	15	33	11	0	10	17	0	0	6
5-46	33	19	6	19	22	5	0	6	8	0	0	3
*5-47	44	13	11	23	33	21	10	19	0	0	0	0
*5-48	50	25	22	33	67	32	15	31	17	0	0	0
*5-49	44	31	22	33	67	26	5	25	0	0	0	0
6-1	6	6	0	4	0	5	0	2	0	0	0	0
6-2	6	0	0	2	11	16	0	8	0	0	0	0
6-3	0	0	0	0	0	0	0	0	0	0	0	0
6-4	17	0	0	6	11	5	5	6	8	0	0	3
6-5	11	19	0	10	0	0	0	0	0	7	0	0

Figure 4.

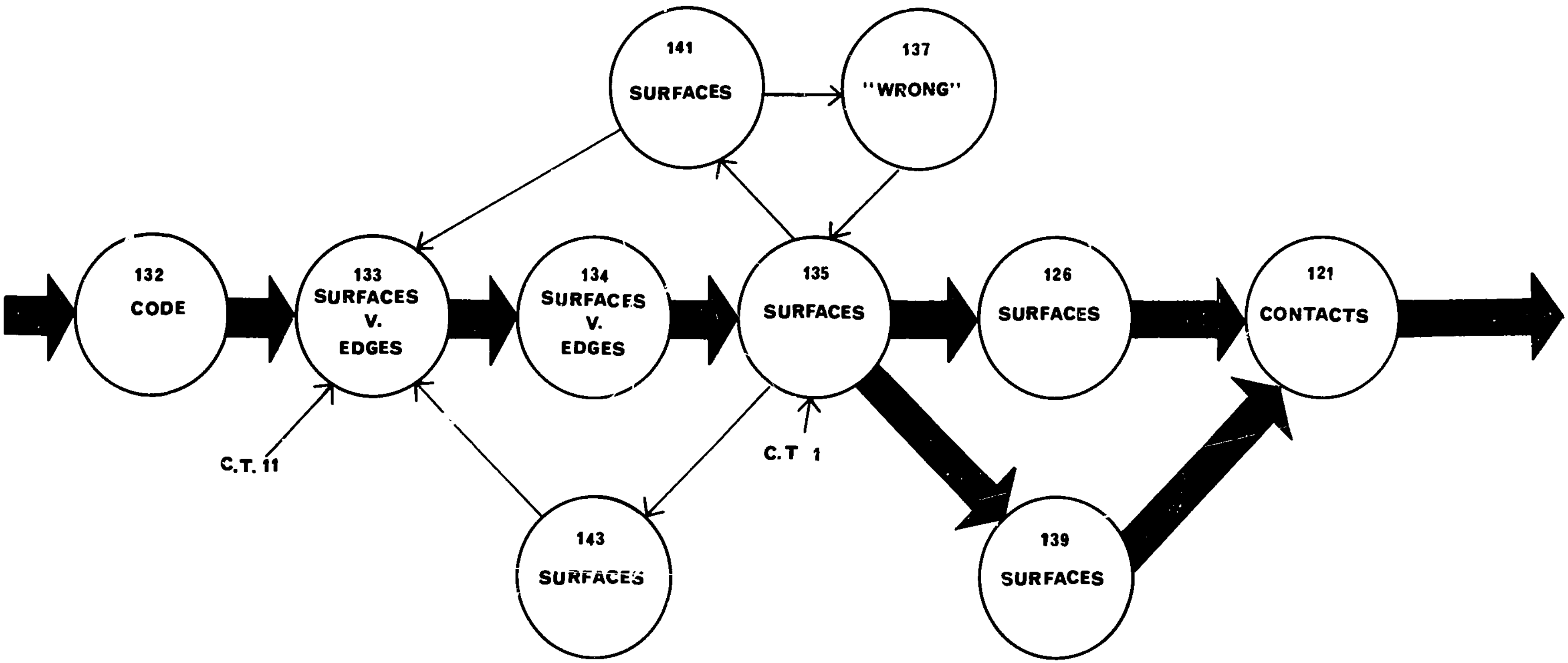

132
CODE
133
SURFACES
V.
EDGES
134
SURFACES
V.
EDGES
135
SURFACES
126
SURFACES
121
CONTACTS
141
SURFACES
137
"WRONG"
143
SURFACES
139
SURFACES
C.T. 11
C.T 1

Figure 5

of frames, presented on paper of conventional size, will not permit a very detailed description of the content of each frame. For reasons that may be deduced from earlier discussion, the words best representing frame content will probably be the words representing the anticipated response of students.

An alternative procedure is to cut up one copy of the program and attach each frame to a small note card, and then to place all the cards in order on a bulletin board (a large sheet of Fir-Tex will do nicely). In addition to the frame itself, error and use rates and other useful information can be indicated on each card.

This procedure will enable the programmer to examine troublesome frames in terms of the concepts contained, and to track back through the program the instruction aimed at teaching those concepts. Thus the programmer need not assume that a higher error rate on a certain frame indicates a deficiency specific to that frame, but may indicate a deficiency in certain previous frames that may not themselves have had a high error rate.

Terminal frames as learning indicators. Programmers frequently employ "terminal frames" which may be distinguished from instructional frames by the fact that they are not intended to teach anything, only to test the adequacy of recall of what has been learned from previous frames. Terminal frames contain no new instruction, they simply present the student with a problem with a minimum of cues regarding the correct solution. When this programming technique has been employed, terminal frames should be specially designated. Responses to these frames represent "learning indicators", as described earlier in this paper, while the more typical instructional frame

responses represent "process indicators". A high error rate on a terminal frame casts doubt on the adequacy of all the frames immediately preceding it which were presumably intended to prepare the student for the terminal frame. This is true even if the error rates on the preparatory frames themselves appear to be acceptable.

Making information useable. While general and profile information will be useful to the programmer as he prepares for the actual revision process, he will actually undertake the revision process frame by frame. Thus he will need to have information about each frame focused in such a manner that it meets his procedural requirements. One way to accomplish this was alluded to above in describing the flow charting process. A copy of each frame can be attached to a note card. The tables used to summarize frame data may be cut apart so that the data on a row representing a given frame may likewise be attached to the card. In addition, space may be provided for other information, such as student comments and results on relevant post-test items.

In view of Dick's findings as stated earlier in this paper, it is probably not realistic to expect that programmers will inspect post-test results and implement them in any meaningful way in the revision process. Assuming that the stated behavioral objectives are taken seriously, the function of the post-test is simply to indicate when the revision cycle can be stopped because an acceptable level of performance has been achieved. Summary data about post-test results do not lend themselves to articulation in the revision process. However, it is possible to determine individual item results, and then to correlate these with frames relevant to that specific achievement. If this is done, frames related to deficient post-test item

performance may be identified. In some cases, these frames may already have been identified as d-ficient due to a high error rate, but in other cases they may not. In any event, this procedure is one way of sensitizing the programmer to problems that might otherwise have been overlooked.

If pre-tests were employed, a similar procedure may be followed, only this time paying attention to item success rather than item failure. Items typically passed on the pre-test may indicate unnecessary frames in the program.

There will also be a considerable amount of narrative information that will require organizing and presenting. For the most part, this information will be in the form of comments by reviewers and students, some of it general and some of it relevant to specific frames. Both general and specific comments should be coded according to their origin. This is important because it enables the programmer to determine the authority and expertise of the person making the comment. Hopefully, those reviewing the program (except for students working through the program) will have exercised sufficient restraint that their comments will be within their own area of expertness. Subject matter specialists should not comment on programming style nor should writing specialists comment on content. The programmer should take quite seriously those comments that are within the reviewer's area of expertise, but should use his own judgment as to how seriously to consider comments that overreach that area.

Student comments should be coded in the same way. Interestingly, they are perhaps the best qualified of all reviewers, because they know, in a way that is not possible for the others, what is understandable and what is ambiguous to students, where and for what reason the students are likely to encounter

difficulties. Thus, except where technical accuracy or correctness is concerned, the comments weighed most seriously should be those of students. Mager (1962, p. 5) makes this point persuasively, in providing a list of tips to programmers:

1. Whenever you are stuck, and for whatever reason, you can get unstuck by asking a student to help you. Ask his reaction to the explanation, or to the item you are having trouble with, or ask him to suggest an analogy or example that might help.

2. Your colleague will probably be more trouble than help. While he can be of great help in reviewing the technical accuracy of your material, he will find it impossible to keep from commenting on your choice of subject matter ("They won't really understand it unless you include this and that, etc."), upon the philosophy of programming ("That's not the way I teach it... All you're doing is spoon feeding them... If it ain't hard it ain't learnin'."), or about your style ("You've taken a whole page to say what you could have said in just one sentence."). You will find that your program draft behaves much like an ink blot and when you ask your colleague to comment on it his comments are likely to tell you more about himself than about your program.

3. To repeat, whenever you are not sure what to do next, ask a student.

As suggested previously, comments regarding specific frames should be reported on individual frame revision cards along with the frame itself and other relevant data. This will facilitate considerably the use of such information by the programmer in the revision process.

General comments by the students and reviewers will be somewhat more difficult to handle. They should be identified or coded as to source, and classified according to their nature. For example, comments about subjective feelings or attitudes regarding the mode of instruction might constitute one grouping, and comments regarding writing style another grouping. If these various classes contain a large number of comments, the main thrust of their meaning may be lost in verbiage. In order to synthesize this information in a form most useable to the person revising the program, redundant comments

may be summarized in a single statement, and the frequency of that type of statement, in terms of the number of separate statements from which the summary statement was derived, can be noted parenthetically. In order to further insure that the more frequent comments are given appropriate consideration, summary comments within each classification may be listed in order of their frequency.

Another helpful step that may increase the likelihood that general comments will actually influence specific program changes is for either the evaluator or the programmer to examine the program frames in view of the general comments and flag each of those frames in which a general comment appears to have relevance.

Summary. It appears that, in general, the large-group approach to formative evaluation enables the marshalling of a large and diverse pool of evaluative data, but that careful consideration must be given to the manner of organizing and presenting this data if it is to be useful in a formative revision process. Assuming that this is done, the approach enables the evaluator to go a long way in meeting earlier described constraints of timeliness, manageability, and credibility. Both general profile and specific information can be supplied to developers at the time they need it, in a form that they can use it, and of such a nature that it is considered a valid input to the revision process.

The comprehensiveness of this evaluation procedure can be seen by recalling the sequence of contingency classes described earlier in this paper and then noting the specific types of indicators provided for, with respect to each of these contingency classes. These are indicated as follows:

Point of Origin	Pre-Tests General Abilities Prerequisite Behavior
↓	
Instructional Stimuli	Student Comments Various Reviewer Comments
↓	
Process Indicators	Monitor Comments Frame Responses Analysis
↓	
Learning Indicators	Terminal Frame Analysis
↓	
Diagnostic Indicators	Synthesis of Other Data
↓	
Criterion Indicators	Post-Test

The large-group procedure is not as likely as the tutorial procedure to result in lengthening the program, but neither is it likely to shorten it appreciably. This approach is still largely insensitive to inefficient programming, where steps are too small, or frames too wordy. Unfortunately, these defects may also becloud the interpretation of frame data which can and should shed light on the relationship between the stimulus and process-indicating response contingency relationship. The procedure to be discussed next offers some promise of coping with these deficiencies. It is perhaps best used as a supplement to rather than a substitute for the procedures we have discussed.

4.1

<u>The Blackout Technique</u>.

This technique was developed by Holland and Kemp (1965) in an attempt to

provide a measure of the degree to which the material is programmed. An interesting by-product of the approach is that it may itself provide a means for revising materials. Of particular interest here is that its use may tend to minimize some of the deficiencies of the procedures previously described. Holland (1967, p. 89) describes the rationale for the procedure thus:

> Although most programmers would agree that not all sets of short completion items are programs, there are as yet, no unequivocal criteria for identifying the point when material can be termed "programed". If the programmer had at his disposal clear-cut measures of the degree to which material is programmed, he would have powerful help in designing material. Also, program editors could easily identify well-done programs, and could easily communicate their evaluations to authors. Experimenters with questions as to comparisons of programs and conventional techniques, or those with questions as to best presentation modes for programs or best response modes for programs would be able to identify first whether or not they were using real programs. Strangely, many experimenters have rushed to make such comparisons without first being able to demonstrate that they were in fact dealing with a program.

Figure 6 shows how two items from a program were changed by application of this technique.

The blackout procedure involves simply obliterating with black crayon phrases considered not absolutely necessary to elicit a correct response from students. This gradual, iterative process is continued until an increase in error rate of students working through the frame is noted. When the blackout process is completed, only the information necessary to maintain the original error rate will be retained. In poorly programmed materials, this process may result in blacking out as much as three-fourths of the original material. The blackout ratio is determined by the percentage of words in the original text that can be blacked out without increasing error rate. Poor programs, or

Normal	Blacked-out
22. A B--- of Exchange (Draft) is convenient for the payment of debts.	22. A B--- of Exchange (Draft) [redacted]
23. The seller of merchandise by sending a Bill of E------- drawn on the buyer and attaching the shipping documents to a bank for collection can be assured that the merchandise will not be delivered to the buyer until the buyer pays for it.	23. [redacted] a Bill of E------- [redacted]

Figure 6. Samples of Normal and Blacked-out Items from the Monetary Program. (Holland, 1967, p. 89)

relatively unprogrammed material, will have a relatively high blackout ratio. Highly programmed material will have a low ratio.

While this technique may be useful for eliminating redundant or superfluous material, it should be obvious that it will not transform poor programs into good ones. Holland emphasizes that "the blackout measure is not a direct measure of whether the material teaches well or poorly, nor is the blacked-out version considered to be an improvement of the original version. The blackout ratio is only intended as a measure of the extent to which the original material is programmed." Holland quotes Skinner (Skinner, 1963, pp. 168-77) in defining programming as "the construction of carefully arranged sequences of contingencies leading to the terminal performances which are the object of education."

The contingencies referred to in this definition are concerned almost exclusively with the second and third level of the sequence of contingency classes described earlier in this paper. It is in the examination of the relationship between these two classes that the blackout technique is most useful. In the ideal case, the program will be designed and executed in such a manner that the learner is lead through a carefully planned sequence of steps to successful achievement of the learning objective. Each step involves the presentation of instructional stimuli, the elicitation of a meaningful response from the student, and the presentation of feedback regarding the adequacy of that response.

In the ideal situation, each frame will contain the necessary and sufficient conditions to enable a correct response. Appropriate responses indicate that the intended process is occurring. The sufficiency of

the instructional stimuli is tested by the error rate of students working through the program. Given that the error rate is acceptable, the necessity of the various elements of the instructional stimuli may be tested by eliminating words until decrements in performance occur. Ideally, through this procedure everything necessary but nothing unnecessary would be contained in the program.

In summary, the blackout technique provides a means of assessing the extent to which a certain desired rythmic stimulus-response instructional process is occurring. In this sense, it serves as an index of the extent to which the instructional process is controlled by the programmer, or conversely, the extent to which learning is dependent upon uncontrolled contingencies. Holland (1967) argues persuasively that application of the technique will enable much more fruitful examination of programming variables. He attributes the equivocal findings of much programmed learning research to the fact that many of the programs used were not, in his terms, highly programmed. It seems quite possible that the approach that he advocates, the careful examination of stimulus-response contingencies, may enable the discovery of many principles for effective programming, in which case the additional rigor would certainly be justified.

While this is a cogent argument for a science building, it still leaves some questions as to the relevance of the technique for formative evaluation purposes. Aside from scientific concerns, is the blackout technique of practical relevance to the program developer?

One can develop an argument on a rational basis that it is, particularly in the case of longer programs or situations where programs are used extensively. If one accepts the theoretical basis for the use of reinforcing feedback

in facilitating learning, it can be argued that over a period of time the student is being conditioned by selective reinforcement to focus his learning activity on just those stimulus elements that are required to enable his correct response. If this is true, as time goes by he will be less and less likely to attend to those stimulus elements that are not relevant to his immediate response. More and more he will learn only that which he is responding to. The writer's experience indicates that this is indeed the case. Very likely, this phenomenon would not occur in the shorter programs typically used in research efforts.

Conversely, if students perceive the need to read carefully each frame regardless of what portion of it is responded to and reinforced, then in effect the significance of feedback is being systematically reduced, with a consequent loss of the presumed benefits of reinforcement. Thus the high blackout ratio, while possibly acceptable in a short program, would appear to be a danger sign in a longer one.

It should be emphasized that the blackout technique illuminates best the learning process, not its product. Dashboard instruments give some indication of a car's proper functioning, but they do not reveal its destination. Likewise, the blackout technique reveals little about the achievement of an ultimate objective more than the extent to which the means are viable.

4.2

ERIC

References

Cronbach, L. J. Course improvement through evaluation. _Teachers College Record_, 1963, 64.

Dick, Walter. The methodology for the formative evaluation of instructional materials. _Journal of Educational Measurement_, Summer 1968, 5 (2).

Holland, James G. A quantitative measure for programmed instruction. _American Educational Research Journal_, March 1967, 4 (2).

Holland, James G., Kemp, Frederick D. A measure of programming in teaching-machine material. _Journal of Educational Psychology_, Dec. 1965, 56.

Horn, Robert E. _Development testing_. Ann Arbor, Mich.: Center for Programmed Learning for Business, 1966.

Lang, Robert L., and Kersh, Bert Y. _Dental anatomy_, unit II. Research completed for the University of Oregon Medical School. Eugene, Oregon: University of Oregon Medical School Press, 1967.

Mager, R. F. _Getting started with programmed instruction_, Varian Associates, 1962.

Markle, Susan. Empirical testing of programs (Ch. V). In Lange, Phil C. (Ed.) _Programmed instruction_, the sixty-sixth yearbook of the National Society for the Study of Education, part II. Chicago: University of Chicago Press, 1967.

Paulson, C. F. The effectiveness of the self-instructional program, longitude and latitude. Monmouth, Oregon: Teaching Research, 1963.

Scriven, Michael, The methodology of evaluation. In Tyler, R. W., Gagne, R. M., and Scriven, M. (Eds.) _Perspectives of curriculum evaluation_, AERA monograph series on curriculum evaluation, Vol. 1. Chicago: Rand McNally, 1967.

Skinner, B. F. Reflections of a decade of teaching machines. _Teachers College Record_, 1963, 65.

Section V: Measurement

Editorial Foreword

The editor has always considered this chapter on measurement by Del Schalock, one of the better professional contributions from Del and from our division. The chapter presents a broad perspective on the topic of measurement, and some unique ideas on the criteria for judging the adequacy of measures. These ideas which diverge from the customary notions of reliability and validity, are presented primarily in Part Four of the chapter entitled, "Evidence needed in the support of the trustworthiness of a measure."

The chapter does not require prerequisite psychometric training. However, the reader would gain in depth of outlook if he were able to contrast Schalock's point of view with that typically found in introductory books on tests and measurement.

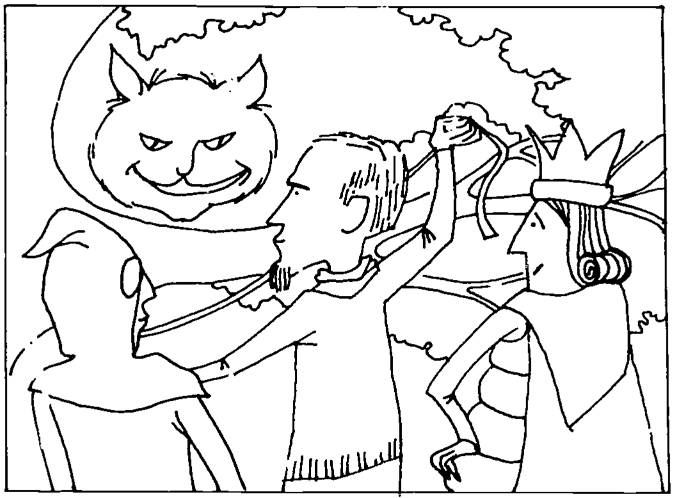
The chapter is well organized, although profuse in parts. In Part One of the chapter, Dr. Schalock amply presents a perspective of measurement as it relates to the other topics in this manual. Naturally, he finds measurement both prerequisite and ubiquitous.

Some readers, including the editor, have difficulty with the lofty organization and abstract style of the author. I can testify that however formidable, the chapter is worth wading through.

All the chapters in this manual either directly assert or imply that discovering and clutching sound evidence is essential to their area. And, although both the chapter on evaluation and the one on design present overall plans for organizing this collection of evidence, the details of collecting

V

useful information are dealt with only in this chapter on measurement. In this sense, the chapter represents the infra-structure of the entire volume.



King. "If it's got a head,
it can be beheaded."

Schalock: "If it's a construct, we
can measure it somehow."

V

SECTION V

Measurement

H. Del Schalock

Table of Contents

INTRODUCTION AND STATEMENT OF OBJECTIVES

The present chapter occupies a somewhat unusual position in the manual since it stands as a bridge between chapters dealing with instructional systems development and evaluation and those dealing with instructional research. Placing the chapter in this position was done purposely since measurement is common to both evaluation and research activities, and the theory and methodology underlying measurement does not vary by activity. It is obvious, of course, that specific measures will vary according to the nature of the activity being pursued, and that somewhat different criteria operate in the selection and use of instruments in the pursuit of research and evaluation functions, but the underlying principles of measurement and the broad general classes of measures available for use in the behavioral sciences are applicable across both research and evaluation efforts.

Because of the frequent and often not too critical use of such terms as research, evaluation, development, measurement, design and analysis a person reading in these areas is often confused as to the distinction and interrelationships that exist amongst these terms. In an effort to place the process of measurement in perspective, and thereby enhance the information that can be taken from the chapter, the discussion of measurement is preceded by some defining, differentiating and interrelating statements.

1. <u>The relationship between development, evaluation and research</u>. In general terms instructional systems development refers to the systematic design, building and testing of a set of instructional experiences to bring about a given outcome for a given learner or set of learners

in a given setting; instructional evaluation refers to the examination of products or events in light of specified standards for the purpose of making adaptive decisions; and instructional research refers to the set of activities that lead to the identification of empirically verifiable generalizations about the relationships between instructional acts, learner outcomes, learner characteristics, and instructional settings. Within this context the process of evaluation supports instructional systems development efforts by providing the kind of information the systems developer needs in order to make adaptive decisions relative to the developmental process. As Dr. Hamreus has pointed out, three kinds of information are continuously needed by the developer:

a) information relative to the degree to which learners achieve the level of performance that is expected of them as they move through the instructional systems;
b) information relative to the effectiveness of each of the elements in the instructional system in bringing about the outcome intended for it; and
c) information relative to the appropriateness of the conceptual framework that underlies the instructional system.

Instructional research contributes to instructional systems development efforts by identifying "principles of instruction" that can be used in the design of instructional systems.

2. <u>The role of measurement in evaluation</u>. As pointed out by Bud Paulson instructional evaluation can assume many forms and serve many functions:

it can aid in the identification of the need for new instructional efforts; it can help in the formulation or design of such instructional efforts; it can assess their effectiveness and/or impact; or it can assess their relevance or appropriateness. In any of its uses, however, evaluation is dependent upon measurement as a basis for information gathering.

Because of the demands upon evaluation to move information at a rate and in a form that permits decision makers to make the on-line decisions that must be made, much of the measurement in instructional evaluation is informal and unsystematic; any procedure is acceptable for use which provides the information needed to make a critical decision at a given point in time. Thus, casual on-the-spot observations and interviews, questionnaires, records, and passing comments often constitute the "stuff of measurement" that is brought to bear in instructional evaluation. This is not to imply, however, that all evaluation efforts involve such informal measurement approaches. The basic rule underlying any evaluation effort is to obtain measures that are as reliable as possible within the time and resources available, but the commitment to provide the information that is needed to make on-line decisions sets such severe constraints upon time for data gathering that more formal measures often cannot be used in evaluative efforts.

3. The role of measurement in instructional research. Because instructional research is concerned with the development of "generalizable instructional principles" the demands upon quality measurement are more

stringent than they are in instructional evaluation. Moreover the demands for the inclusion of specific classes of measures in instructional research efforts are also more stringent than in evaluation efforts. In the search for instructional principles it is desirable to include at least five classes of measures in each research study: a) measures of the content carried by an instructional message; b) measures of the strategy used in an instructional act; c) measures of the characteristics of learners which interact with instructional content and strategy to determine that which is learned; d) measures of the characteristics of the instructional setting which interact with content strategies and learner characteristics; and e) measures of the outcome that is expected to derive from the learning. To the extent that one or more of these classes of measures are left unattended "uncontrolled variance" enters the research results, i.e., the less the explanatory power of that study. Obviously not all dimensions of learner characteristics, setting characteristics or teacher behavior can be measured or controlled in every study, but the educator who is serious about engaging in instructional research must attend to as many of them as possible. Historically, one of the ills that has characterized instructional research has been the failure to attend to as many of these variables as should have been attended to, or to attend to them so grossly as to make their measurement relatively inconsequential.

4. The relationship of measurement to design and analysis. Instructional evaluation and instructional research always in. ~~ve~~ issues of measure-

ERIC

ment, design and analysis. In instructional evaluation all three activities may be considerably less formal on occasion than they are in instructional research, but each evaluation activity must, by its nature, involve elements of all three. Instructional research of course always requires careful attention to all three. Operationally, measurement, design and analysis are linked together in the following way: measurement involves a set of operations that provide observations of the object or event that is the focus of concern; design pertains to the conditions established under which measurements are taken, where the aim of those conditions is the reduction of unwanted sources of variance that influence the measures taken; and analysis involves a set of operations that orders the observations taken through measurement in such a way that conclusions appropriate to the limitations of a given design may be drawn. As such the activities in measurement, design and analysis are inseparably linked, and decisions in one area both influence and are influenced by decisions in another. In order to carry out either instructional research or evaluation it is paramount that the principles of measurement, design and analysis, and the interactions between them be understood.

While the distinctions and interrelationships between research, development and evaluation and between measurement, design and analysis are critical to the full understanding of any of these concepts individually the present chapter does not eleborate the relationships spelled out above. Its focus instead is upon the theory of measurement generally, the classes of measures available to the behavioral scientist, and the kinds of information needed about a given

measure in order to have confidence in the data that derive from its use. The position from which the chapter is written is simply that instructional systems development, evaluation and research are completely dependent upon measurement and that the quality of the measurement incorporated within these activities is, along with the quality of design and analysis, responsible for the quality of the product that evolves from these activities.

Broadly speaking the purpose of the chapter is to provide a framework which will permit the users of measures in instructional research, development and evaluation to build into their activities the highest quality measurement possible. Specifically the objectives of the chapter are five:

1) to develop in the reader a sensitivity to the role of measurement in instructional systems development, evaluation and research;
2) to develop in the reader a point of view toward the use of measurement in education;
3) to develop in the reader an understanding of the concept of measurement;
4) to develop in the reader a familiarity with the major classes of measures available to the behavioral scientist, and the particular strengths and weaknesses of each; and
5) to develop in the reader a sensitivity to the kinds of information that is needed in order to insure that any measure used is trustworthy and that it is applied in a trustworthy way.

The chapter does not have as an objective the development of the ability to build new measures. This requires knowledge and skill that is relatively specific to each class of measure available to the behavioral scientist and would involve a degree of detail that cannot be accommodated within the space limitations of the chapter.

Hopefully, the chapter will serve an additional purpose, namely, an extension or reformulation of measurement theory in the behavioral sciences. By-and-large, measurement in the behavioral sciences is poorly understood. While it is not hard to understand or accept the principles involved in some of the more common measurements used in the natural sciences, for example, length, weight, and volume, it is hard sometimes to understand and accept the fact that measurement of such characteristics of individuals as achievement, intelligence, aggressiveness, and anxiety involves basically and essentially the same thinking and the same general procedures. It is also hard to accept the fact that if the rules of measurement can be set up on some rational or empirical basis measurement of anything is <u>theoretically</u> possible. Above and beyond the difficulties most people have in understanding the "philosophy" of behavioral science measurement is the difficulty they have in making reasoned sense of the concepts used to describe the adequacy or trustworthiness of behavioral science measures. The concepts of "reliability" and "validity" have had many changing referents, and, in the opinion of the writer, not entirely adequate ones. As a consequence, a major aim of the chapter is to develop in the reader an awareness of an alternative set of constructs which define the adequacy or trustworthiness of a measure. While the provision of such a framework moves considerably beyond that which might be thought of as content appropriate to a chapter in the application of measurement in education, it is anticipated that the framework will have utility for the reader as well as for the field generally.

It is also a prerequisite to an understanding of the chapter as a whole since it provides the integrating framework for the thinking that is outlined in it.

The chapter is divided into five sections: (1) a point of view toward the use of measurement in education; (2) the concept of measurement; (3) classes of measures used in the behavioral sciences; (4) evidence needed in support of the competency with which a measure is applied. While much of the content of the chapter may appear on first encounter to be unduly technical or complex, experience has shown that persons with little or no formal training in measurement can understand it.

5.1

PART I: A POINT OF VIEW TOWARD THE USE OF MEASUREMENT IN EDUCATION

The content of the present chapter derives from a point of view about the use of measurement in education, and the reader should be aware of it. The point of view is reflected in three interrelated assumptions:

1. An educator or educational researcher must understand the fundamental nature of measurement, particularly as it applies in the behavioral sciences, if he is to pursue instructional development, evaluation and research activities judiciously. Simplified guides as to when to use what kinds a measures upon whom are not enough. If education is to be effective, and if a "science" of education is to emerge, the nature of educational measurement must be understood by those who use it. Lindquist (1950, p. 158) states the view well when he says "If measurement is to continue to play an increasingly important role in education, measurement workers must be much more than technicians . . . unless they show much more concern with what they measure as well as how they measure it, much of their work will prove futile and ineffective."

Toward this end, the present chapter introduces the reader to content that is designed to make him "more than a technician".

2. An educator or educational researcher must be willing to undertake the measurement of classes of educational objectives that heretofore have gone largely unmeasured. Thus far in education both standardized achievement tests and informal teacher examinations have focused primarily upon the measurement of content objectives that derive from established school subjects. Objectives concerned with such things as sensitivity or consideration toward others, artistic abilities, artistic or aesthetic preferences, moral values, attitudes toward social institutions and practices, habits relating to personal hygiene and physical fitness, managerial or leadership ability, etc., have been seriously neglected--even though most educators hold them to be of importance. The point of view adopted here is that if these are important classes of objectives they should be so specified, and measures developed to assess whether the school experience brings them about.

   Toward this end the present chapter aims to sensitize the reader to a wide range of educational objectives and to suggest alternative strategies that could be used in their measurement.

3. If the educator or educational researcher does undertake the measurement of the full range of educational objectives to which he is committed, he will have to rely upon more than "paper and pencil" measures. This point of view rests upon the assumption that the aim of any educational achievement measure is to obtain a measure of behavior or a product of

behavior that is acceptable as evidence of the realization of the objective desired. How else can one have confidence that an educational objective has in fact been realized? Given such a point of view, paper and pencil tests are not at all appropriate to the measurement of the full range of educational objectives desired. When the focus of measurement is upon "knowledge", e.g., concepts, facts and principles, or upon the application of knowledge to a set of tasks that require only the manipulation of symbols, e.g., the solution of a mathematics problem, outlining the steps involved in building a house, or writing a theme on the expression of consideration in one's relations with others, paper and pencil measures are perfectly appropriate. However, when the focus of measurement is upon <u>building</u> a house, or upon <u>relating</u> considerately to others, a paper and pencil measure won't do--unless one is willing to make the assumption that being able to outline <u>how to build a house or how to behave considerately is in fact</u> related to the ability to build a house or to act considerately. While most people would probably accept the idea that the knowledge factor is <u>related</u> to the concrete performance factor, few would accept the idea that the relationship is perfect. If this is true, then measures other than or in addition to those which require only the manipulation of symbols are needed in order to assess some of the objectives of education.

The chapter rests on the assumption that the educator or educational researcher must be familiar with the full range of measures available to the behavioral scientist. These include such "obtrusive" measures as interviews, standardized objective tests, teacher-made

ERIC

tests, systematic observation (face-to-face observation, tape recordings, video-tape recordings) and standardized projective tests, and such "unobtrusive" measures as physical traces through erosion and accretion, e.g., wear on library books or the accumulation of used paint tubes, documents and products, simple observations, and contrived observations through hidden hardware. Each of these are specialized measurement methodologies, some of which are appropriate to one kind of measurement problem and some appropriate to others, but in order to handle the full range of measurement tasks that he faces an educator or educational researcher needs to be competent in them all.

Toward this end the strengths and weaknesses of the major classes of measurement methodologies used in the behavioral sciences are summarized.

Hopefully such a point of view about measurement in education will be widely shared for until it is, and the measurement capability which it assumes is available, both the science and practice of education will be handicapped in their progress toward the ends desired for them.

5.2

PART II: THE THEORY OF MEASUREMENT

Lorge (1950, p. 533) has reported that in a sample of 2 1/2 million words the term "measure" occurred more than 400 times and was used in 40 different ways. It was used to indicate the act of weighing, the instrument used in weighing, and the numeral that expressed the result. The term also referred to less exact instruments, processes, and units. In fact, any instrument that is used as a basis for comparison, even when that comparison involves the pro-

cess of estimation or judgment, generally carries the term "measure". Thus, as used popularly, measure not only refers to procedures that have precision, but also to acts of objective estimation, such as the estimation of beauty or the estimation of an individual's intelligence.

In its broadest technical sense, measurement is the assignment of numerals to objects or events according to rules (Stevens, 1951). As measurement is conceived within the physical sciences this is rather a straightforward definition: by using a device such as a ruler or a scale in accordance with relatively simple principles one can assign inches to a table top or ounces to a cup of flour without difficulty and with considerable (though never exact) accuracy. It can be a different matter in the behavioral sciences.

> "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." (Kerlinger, 1965, p. 41)

The author goes on to point out that while this example may seem ridiculous as an illustration of measurement, it does in fact meet the definition of measurement. 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 contained in the instructions to the judges. What makes the

ERIC

example seem far fetched as an illustration of measurement is the fact that <u>the properties or the characteristics being judged were not at all specified</u> i.e., what is meant by "nice", "strong", "intelligent", etc., <u>the rules for assigning the numerals to the properties were not clear</u>, i.e., what properties had a value of 1, 2, 3, etc., and <u>the conditions under which the observations were to be made were not spelled out</u>. Without these minimal considerations, "reproducability of observations" is not likely to be reached. This of course is one of the first criteria of measurement that scientists and users of measures demand, for without reproducability a measure is of little value. <u>To become reproducable, the properties of that which is to be observed, the rules by which numerals are to be assigned to those properties and the conditions under which the observations are to occur must be made explicit and public</u>. Unfortunately, a definition of measurement contains no statement about the quality of the procedures involved in it.

<u>Constructs and Measurement</u>[1]

Perhaps the most basic of all concepts in measurement is the simple notion that in order to measure something one must know what it is he wishes to measure. Unfortunately, like the definition of measurement, this also is a deceivingly simple statement. Two requirements are inherent in it: 1) some <u>construct</u> of that which is to be measured exists and 2) some notion of the <u>measurable properties</u> of the construct exist. Put in another way, that which is to be

---

1 Constructs are concepts which have added meaning of being deliberately or consciously created for a special scientific purpose. Intelligence is a case in point.

ERIC Full Text Provided by ERIC

observed is dependent upon man's ability to conceive of it and, then, of his ability to observe it (Lorge, 1950, p. 536).

Specifying the constructs that are to be observed. An essential feature of any science is the development, extension and refinement of its constructs, for it is through its constructs that it gains its power. Conceptualization (the building of constructs) is a constant, cyclical data-dependent process wherein new data (observations) give rise to new constructs and new constructs give rise to new data. The cycle doesn't begin at a particular place, nor end; it is ever present and so long as man inquires it will forever be present.

The point that is critical here is that constructs don't "just happen" or don't "first exist". They are man-made and they are constantly evolving. For ages, man disregarded certain objects or their effects because he did not know of them or their behavior. He did not notice ultraviolet radiation, nor the fact that quartz reacted differently from glass to ultraviolet light. Nor did he notice electrical currents in the brain, or "unconscious" motivation, or that some people can taste certain things and others can't. We now have such constructs, and they have opened an appreciable store of knowledge. We don't have them in the same form as they were conceived initially however. Atomic structure, circa 1968, is not atomic structure circa 1938. Neither is "intelligence" nor "motivation" nor "learning". Constructs come and go, may be powerful or weak, but always they change. Moreover, whatever their form, they dictate that which is to be observed and measured.

Specifying the properties of constructs that are to be observed. Having invented a construct, or having revised one, is not enough in itself to permit its measurement. The properties or characteristics of the construct must also be specified, for it is these that in fact permit measurement, i.e., permit

the assignment of numerals to indicate quantity. Thus, it is not sufficient simply to invent a construct such as intelligence, or achievement, or anxiety, or poverty, or cultural deprivation. One must also specify what it is that constitutes such constructs.

This constitutes one of the most difficult problems in measurement, for as yet there are no rules governing how it is to be done and no way of knowing whether it ever gets done. The procedure followed in specifying the properties of a concept has been labeled generally as "constitutive definition" (Margenau, 1950; Torgerson, 1958) and refers to a procedure whereby a construct is simply defined with progressively lower-order constructs until one either encounters a set of constructs whose properties have been defined in other terms or reaches a point beyond which defining constructs can't be found. Thus, one may begin with a construct such as teacher behavior, break it down into its major components, e.g., caretaking, teaching and routine-administration behavior, then break each of these down into their major components, etc. Ultimately a point is reached beyond which lower-order constructs are not applicable, and at that point the properties of a construct will have been specified so far as available knowledge permits. This whole process will be recognized of course as a procedure comparable to the hierarchical analysis of educational objectives that Dr. Twelker has reviewed in Chapter II.

Specifying indicators of the properties of the constructs that are to be observed. In the paragraph above attention was given the fact that in any measurement the act of measuring is directed to the properties of objects or events (the properties of a construct) rather than the object or event itself. While this is true, it does not mean that the properties themselves are actually

measured, for the properties of an object or event are themselves constructs. What actually gets measured in the act of measuring are indicants of the properties of objects or events. Thus distance is indicated by such standardized measures as an inch, a mile or a light year or by such non-standardized measures as three days of travel on a good horse or as far as the eye can see on the open prairie and anxiety is indicated by response to such standardized paper-and-pencil items as are contained in the Taylor Manifest Anxiety Scale or such non-standardized indicants as the GSR, flushing, trembling, perspiration, etc. When thus conceived the "reality" of measurement is much different than that anticipated by many people when they think about measurement. Conceiving of measurement as the assignment of numerals to indicators of properties of objects or events somehow makes the process seem less dependable or exact than the label historically has implied. Fortunately or unfortunately, the process is the same in all measurement, whether in the physical or the behavioral sciences.

Some "problems" people have in relation to constructs and measurement.

1. THE PROBLEM OF INFERENCE. Objects, events, processes, etc. or constructs of objects, events or processes, can be thought of as falling along a continuum of concreteness-abstractness. If one conceives of that which is to be measured in this way it becomes clear that some of the properties of objects, e.g., width, height, weight and hardness, or some of the outcomes of education, e.g., reading ability, knowledge of mathematics or skill in athletics, fall on the concrete end of the continuum and are therefore amenable to rather "direct" or "low inference" observation. This is not the case for such constructs as intelligence, morale, anxiety, hostility, or creativity. These are constructs that exist only by inference; they are assumed or inferred to exist because

of certain observed regularities in behavior. As one moves away from the measurement of indicators of concrete properties of objects, that is, properties that are available to direct observation, and into the measurement of the indicants of "inferred" objects or constructs, reliance upon inference becomes heavier and heavier. As this happens uneasiness about measurement increases. As soon as relatively simple physical properties are left behind for more complex and elusive properties, direct observation of properties is impossible. Hostility cannot be observed directly; nor can morale, anxiety, intelligence, creativeness, talent, etc. Such constructs require indirect, "high inference" measurement in that their properties or characteristics must always be inferred from observation of presumed indicants of those properties.

When using indirect or high inference measures indicants are taken as reflections of an underlying property or characteristic. They are seen as "something" which points to something else. If a boy continually strikes other boys, we may say that his behavior is an indicant of underlying hostility. If someone's hands sweat excessively, we may say that he is anxious. If a child answers a certain number of items in an intelligence test correctly, we say he has a certain level of intelligence. In each of these cases, some identifiable behavior is assumed to be an indicant of any underlying property of a given construct that is to be measured.

It is understandable that people view measurement as being rather shaky when it involves making inferences from observed behavior rather than directly observing properties like skin color, size or sex. To measure a child's cooperativeness, dependency or imaginativeness is very different from measuring his height, weight, or wristbone development. The fundamental process of measure-

ment is the same but the rules are much more difficult to prescribe. Inference dominates, and this creates one of the more vexing problems of psychological and educational measurement.

The distinction between direct and indirect measurement, or between low and high inference measurement, has major implications when considering the evidence one needs in judging the trustworthiness of a measure (see Part IV in the present chapter), and will therefore be returned to in later pages.

2. THE PROBLEM OF RESTRICTIVENESS. One aspect of measurement that is bothersome to people is highlighted in the previous discussion, namely, the fact that any single measure can attend to only one or a few of many properties of a complex construct. No single measure of a teacher's behavior can ever measure all of the properties or dimensions of her behavior; nor can a single measure of intelligence ever measure all of the rich and diverse properties of human intelligence, or a single measure of creativity ever measure all of the characteristics of human creativity. Even more bothersome is that in science generally an effort is made to measure a property or characteristic of an object with little or no regard for any of its other properties or characteristics. The length of a table may be estimated without reference to its color, width, height, wood, style, or shape. Psychometrists try to make estimates of the "intelligence" of an adult with little or no consideration of his race, personality or economic circumstance. Given such constraints, a single measure can at best give only an approximation to a construct, and if the construct is one which is complex, i.e., involves many properties, it will require many measures to yield an adequate approximation of it.

3. THE PROBLEM OF MEANING. When measurement is recognized as no more and no less than the set of operations which assigns numerals to indicators of properties of objects, events, or processes the meaning that can be given to that which is measured sometimes suffers. Anxiety, intelligence, creativity, beauty, speed, density and all of the other constructs that are so rich in meaning to so many people pale when the operations that constitute their measurement are applied to them. As carried by most of us such constructs are rich in surplus meanings, i.e., meanings which are not ordinarily reflected in the operations which measure them, and as such they have broad utility, high affective loading, etc. Unfortunately they are not very exact, and therein lies the problem from the point of view of objectivity, empiricism or science. In science the meaning of a thing is limited to the operations that are used in its measurement! In essence, an operational definition is a definition that assigns meaning to a construct in terms of the operations that are used in measuring the indicants of that construct. Intelligence, for example, if only one measure of a construct is used, may be defined as that which is measured by text X, or anxiety defined as the number of facial tics emitted over a given period of time and under specified conditions. By taking such an approach to the problem one danger is reduced, namely the danger that one will impute a reality or an existence to a construct which it does not really have (the process of reification). While this is a major gain it does not do away with the fact that one is still dealing with an abstraction rather than a concrete reality, or the fact that a single measure rarely taps all of the properties of an object or a construct. As such the acceptance of operational definitions represents a constructive step forward in measurement theory but not a panacea as once thought.

While operational definitions are indispensible ingredients in all scientific measurement they are dependent for their worth or power upon the conceptual activity that breaks out the properties of the constructs that are to be measured and the indicants of those properties. In combination, these two activities constitute the essence of the scientific enterprise, that is, the shuttling back and forth between the level of construct-hypothesis-theory building and systematic observation.

5.3

Operations and Measurement

Assuming that the properties or characteristics of a concept have been identified (defined, conceptualized), there is still the task of assigning numerals to these characteristics in such a way as to indicate their quantity. This procedure must be governed by "rules" or "operations" which, like the properties of the concepts being measured, are made explicit and public. Only in this way can a measure be "reproducable", and thus admissible as objective information.

Specifying the rules by which numerals are to be assigned to that which is to be measured. In thinking about the assignment of numerals to properties it needs to be recognized that a *numeral* is a symbol of the form: 1, 2, 3..., or I, II, III.... It has no quantitative meaning unless it is given such a meaning. It is simply a symbol of a special kind. The term numeral is used because measurement ordinarily uses numerals which, after being assigned quantitative meaning, become *numbers*. A *number* is a numeral that has been assigned quantitative meaning.

As used here, the term "assigned" refers to the mapping of the objects of

one set (numerals) to the objects of another set. In behavioral science research the members of one set are usually individuals, or indicants of properties of constructs which relate to individuals, and the members of the other set are usually numerals.

Rules govern the assignment of the objects of one set (numerals) to the objects of another set (persons or properties). In the behavioral sciences a "simple minded" rule might say: "assign the numerals 1 through 5 to individuals according to how nice they are. If an individual is not at all nice, let the number 1 be assigned. If he is very, very nice, let the number 5 be assigned. Assign to individuals between these limits numbers 2, 3, and 4." Another rule might be: "If an individual is male, assign him 1. If an individual is female assign her 0." Both rules of course assume previous definition of constructs and isomorphism with reality (Kerlinger, 1965, p. 413-417).

As with anything else the rules of measurement may be "good" or "bad". To the extent that they are good or bad, measurement is likely to follow suit. Many things are relatively easy to measure because the rules are easy to draw up and follow. To measure sex, for example, is easy since several simple and fairly clear criteria can be used to determine sex and to tell the investigator when to assign 1 and when to assign 0. It is also easy to measure certain other human characteristics: hair color, eye color, height, weight. Unfortunately, most human characteristics are much more difficult to measure, mainly because it is difficult to specify clearly the properties of the characteristics to be measured and to devise clear rules that govern the assignment of numerals to them. Nevertheless, rules of assignment must always govern the measurement process, and more and more attention must be directed to the specification of such if the behavioral sciences are to advance.

ERIC

Categorization and scale placement. Ordinarily when one uses the term measurement, he has as a referent something as manageable and stable as inches or yards or pounds or cents; something that can be added, subtracted, divided and multiplied. Unfortunately, measures in the behavioral sciences do not contain the characteristics which permit all of these operation, and because they don't there is serious debate about what one can in fact do (statistically and otherwise) with them. The purpose of the present discussion is to review briefly the concept of levels of measurement, and then speak briefly to the implications of this for statistical analyses. The discussion which follows draws heavily from the work of Coombs (1953, pp. 472-485) and Kerlinger (1966, pp. 419-428).

Most authors classify measurement activities into one of four levels: nominal, ordinal, interval or ratio, though Coombs (1953) identified five. For purposes of the present statement the traditional classification is accepted.

1. MEASUREMENT IN TERMS OF A NOMINAL SCALE. Measurement in its simplest form consists of substituting symbols or names for real objects. When measurement consists only in this mapping of objects into symbols, the symbols constitute a nominal scale. Thus a system which classifies occupation into families or the symptoms of patients into psychiatric classifications represents a nominal scale. Nominal measurement depends upon the most elementary postulate of measurement that exists, namely, (a = b) or (a ≠ b), but not both. Translated this postulate states that "(a) is either equal to (b) or not equal to (b), but not both." For purposes of classification, one must be able to assert either that one object is the same in a characteristic as another or that it is not the same. In measurement "the same" does not necessarily mean complete identity:

it can mean "sufficiently the same to be classed as members of the same set." Saying that two boys are the "same" in maleness is in one sense accurate (they both are males) but it is likely that one boy may actually be more masculine than the other. This is a criterion matter. To be able to say that "the two are the same", one must only meet the criterion or the set of criteria that have been established as a measure of sameness. All criteria have one requirement, however: they must be sufficiently unambiguous to make classification possible, that is, to satisfy the condition the postulate states. In addition, the relation of equality must be symmetric and transitive. By symmetry is meant that if the relation holds between (a) and (b), it also holds between (b) and (a); symbolically, if $a = b$, then $b = a$. By transitivity is meant that if $a = b$ and $b = c$, then $a = c$.

This level of measurement is so primitive that it is not always recognized as measurement, but it is a necessary condition for all higher levels of measurement.

2. MEASUREMENT IN TERMS OF AN ORDINAL SCALE. Sometimes the objects in one class of a nominal scale are more than just different from the members of another class--they may bear some kind of a relationship to them. One such relationship is that the members of the one class are more of something than the members of the other class, and it is meaningful to say that the relation "greater than" (>) holds between the members one class and the members of the other in relation to some property. When this relationship holds for all members of the two classes the result is an ordinal scale.

Ordinal level measurement depends upon the so called "transivity postulate", that is, that "if $(a > b)$ and $(b > c)$ then $(a > c)$". Translated the postulate

ERIC

states that "If (a) is greater than (b), and (b) is greater than (c), then (a) is greater than (c)." Other symbols or words can be substituted for "greater than" (" > ") and "less than" (" < ")., e.g., "is at a greater distance than," "is stronger than," "precedes," "dominates", and so on. Most measurement in psychology and education depends on this postulate, for a goal of most measurement is to be able to assert ordinal or rank-order statements like "(a) has more of a property than (b); (b) has more of the property than (c); therefore (a) has more of the property than (c)."

The preceding statements may seem obvious, and in physical measurements the postulate is often satisfied: if stick (a) is longer than stick (b), and stick (b) is longer than stick (c), then stick (a) must be longer than stick (c). If student (a) has more items right on a test than student (b), and student (b) has more right than student (c), student (a) must have more right than student (c). But take the relation dominance: (a) may dominate (b) and (b) may dominate (c), but it is possible that (a) does not dominate (c). A wife may dominate her husband, and the husband may dominate their child, but the child may dominate his mother. If an investigator is studying dominance relations among children, he cannot simply assume that the postulate is correct. He must demonstrate that it is correct.

3. MEASUREMENT IN TERMS OF INTERVAL AND RATIO SCALES. In the two scales discussed heretofore--nominal and ordinal--the elements of the system were classes of objects, and the relationships were relationships of equality and greater than. Nothing was said about a concept of distance between classes. Thus, although (a) may have been observed to be greater than (b), and (b) greater than (c), nothing was said about (a) being greater than (b) by a larger

amount than (b) was greater than (c). In interval and ratio scales the concept of distance between classes enters the picture, and with it a significant increase in the power of measurement. The increase in power derives from the fact that once information is available on how large the intervals are in the property of the object being measured it then becomes possible to apply arithmetic operations to them.

When it is possible to specify only that intervals or distances of a given amount appear within categories one has interval level measurement. When measures reach this level of sophistication it is possible to apply the arithmatic functions of addition and subtraction to them (though it should be noted that with this level of data it is not quantities or amounts that are added or subtracted, but only intervals or distances). It is only when one is able to use ratio scales, however, that all of the arithmetic functions can be applied, that is, multiplication and division as well as addition and subtraction. This is made possible by the fact that a ratio scale, in addition to possessing the characteristics of nominal, ordinal, and interval scales, has an absolute zero that has empirical meaning. If a measurement is zero on a ratio scale, there is a basis for saying that some measured object has none of the property being measured. Numbers on the scale indicate the actual amounts of the property being measured; if a ratio scale of achievement existed, for example, it would be possible to say that a pupil with a scale score of 8 had an achievement twice as great as a pupil with a scale score of 4. For this reason ratio measurement is the ideal of all scientists.

Unfortunately, as indicated previously, measures in the behavioral sciences are far from reaching the ideal of ratio measurement, and there is no reason to

believe that they ever will. Most measurement is of the nominal and ordinal variety, though many psychological and educational measures are treated as if they approximated interval measurement. The implications of this are twofold: (1) behavioral science measures of necessity lack in precision, and (2) the statistical analyses that can be applied to most behavioral science data are limited. Dr. Beaird speaks to this point at length in the chapter on Data Analysis.

Error and Measurement

As Lorge has pointed out (1950, p. 538-539), empirical measurement ultimately depends upon the occurrence of a sense datum and its interpretation by an observer. Weight, even though measured by machine, is ultimately related to the kinesthetic sensation of heavier and lighter. The machine--that is, scales--merely allows for a simple and relatively objective perception of the effects of weight. The geiger counter is a machine that enables the observer to perceive a specified class of effects by extending the range of human sensation. A test of intelligence serves the same purpose by standardizing a set of stimuli and a set of rules for applying the stimuli to the assessment of given properties of an individual. Other machines facilitate measurement by the control of systematic, chance, or erratic conditions, or by magnifying effects. In this sense machines and tests simply allow more precise determinations of a particular class of effects. In the absence of instruments for extension of the senses, or for the control of conditions, human observations are liable to error. Instruments are a means for approximating more closely the property under observation.

ERIC

Unfortunately, there is a limit to the closeness of approximation in the measurement of a property by an observer or by an instrument. In calorimetry it is well known that the temperature of the measuring device affects the temperature of the material under observation, and in this respect observation must be corrected by calculation for the influence of the instrument. In intelligence testing an examiner can affect the score of the candidate by providing encouragement or its lack. In the physical sciences a great deal of attention is devoted to the reduction of the interaction of the instrument with the characteristic under observation but in the behavioral sciences this is rarely considered. Ultimately, however, the behavioral sciences also must be aware that it exists and specify carefully the conditions under which measurement occurs so as to be able to replicate these conditions when measurement is replicated.

Specifying the conditions under which measurement is to occur. In all scientific observations, whether direct or indirect, the conditions for observation must be carefully specified in terms of time, place, and circumstance. In physics and chemistry, observations at sea level at 25° centigrade may differ markedly from observations of the same thing at 0° centigrade and in an airplane 35,000 feet above sea level. In psychology, the behavior of an individual at 2 A.M. at his desk in his own home may differ markedly from his behavior at 10 A.M. at his desk in his office, or the behavior of a teacher may differ markedly in one subject area as compared to another or when working in a reading group with "good" readers as compared to "poor". To control for this source of variance all measures necessarily must specify the conditions under which observations are made.

ERIC

The problem of error variance. Even though a measuring instrument reflects a careful conceptual effort, sensitive operational definitions, and a detailed statement of the rules for assigning numerals to properties, error will stil enter the measures taken with it. Two people using the same yard-stick to measure the same table will come up with different results. A chemist uses chemical balances so sensitive that they must be kept in another room to guard against the influence of body heat or of air currents set in motion by the chemist's movements; nevertheless, he will weigh material several times and still settle for an average as the "true" weight. The control and/or elimination of error in measurement is one of the most critical tasks of the scientist.

With physical measurements subject to error, it should not be surprising that behavioral science measurements are still more so. For example, attempting to measure intelligence with paper-and-pencil tests would seem to invite all kinds of error. And it does, but measures of intelligence also reflect to some degree the trait being measured; they are not totally inaccurate. If a person scores high on an arithmetic test, for example, it is reasonable to assume, despite the likelihood of error, that he is pretty good at arithmetic.

Essentially, this is the basic assumption of measurement in the behavioral sciences; any measure contains an element of error and an element of truth. Mathematically the assumption is that any obtained measure X is the algebraic sum of a true measure (t) and a measurement error (e), or,

$$X = t + e$$

Unfortunately, the statement is not entirely satisfactory. Logically, a "true" score is one that is not contaminated by any kind of error, but two kinds of error always exist in every measure: constant error and random error. By

ERIC

definition, a constant error is one that appears consistently in repeated measurement; random error is error that influences different measurements to different degrees. With a distinction between systematic, repeatable errors and those that vary randomly, it is possible to rephrase the basic equation of testing as

$$X = s + e$$

with s (systematic measure) representing a composite of a true measure and any constant error. In this revision of the equation, (e) represents only the residual error which is random and unpredictable. (Guion, 1965, p. 29).

While both constant and random error need to be eliminated in measurement to as great an extent as possible, the elimination of constant error is by far the most critical. Constant error occurs with each measure taken and so influences the mean scores of groups being studied (the basic measure used in most statistical analyses) as much as it influences each individual measure. For this reason the presence of constant error in the measures that comprise a set of data distorts the data irreparably, for no amount of statistical manipulation can reduce the distortion or be free of it. Random error on the other hand, while it distorts an individual measure, does not distort sets of measures. This is because the same kind or source of error, if it is truely random, is not likely to appear twice in the same manner. Thus, if only random error were involved in a set of measures the mean of the measures would tend to give a good approximation of a "true" score (because the error scores would be essentially uncorrelated) and statistical manipulations on the data would be free of the damaging kind of bias that is introduced through constant error in

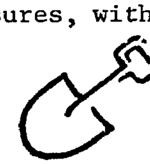
ERIC

measurement. As indicated previously, however, both constant and random error are always present in every measure, and so the above dicussion is rather academic.

Generally speaking, constant error results from either the inadequacy of a measuring instrument (see Part IV of the present chapter) or the inadequate or inappropriate application of adequate instruments (see Part V of the present chapter). Random error generally results from either an "accidentally" inappropriate application of a measure or from uncontrollable events within the measurement situation, e.g., respondent fatigue, noise, unanticipated interruptions, the respondents idiosyncratic reaction to an assessment situation. By-and-large constant error can be reduced, random error cannot. Sources of constant error in the development and application of a measure, and procedures for their reduction are detailed in the sections IV and V.

A general strategy for combatting error variance that derives from measurement: apply two or more measures to any construct being measured. There are two reasons which underlie such a recommendation: a) any single measure can hope to assess only a few of the many indicators that one would be willing to accept as evidence of the construct being measured, and b) because all measures, whether in the physical or behavioral sciences, involve a degree of error. This is especially the case in education, however, where many of the measures must of necessity be high inference measures, i.e., they are indicants of or supposedly related in some way to a given educational outcome. Given these two conditions the educator and educational researcher need to employ what Campbell and Fiske (1959) have called a "multiple operations" approach to measurement. In brief, it calls for multiple measures to be used in assess-

ing each factor to be measured in every study, and assumes that in combination the measures will "share in the relevant components (of that which is to be measured) but have different patterns of irrelevant components" (Webb et. al., 1966, p. 3). The basic assumption underlying the procedure is simply that once a proposition has been confirmed by two or more independent measurement processes, the uncertainty of its interpretation is greatly reduced. Campbell and Fiske argue that the most persuasive evidence for the existence of that which is being measured comes through a triangulation of measurement processes. The basic assumption underlying the approach is that if a proposition can be demonstrated when using a series of imperfect measures, with all their irrelevant error, confidence can be placed in it.



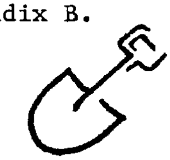
5.4

PART III. CLASSES OF MEASURES IN THE BEHAVIORAL SCIENCES

As indicated previously, the position adopted in the present paper is that to be maximally effective the educator and/or educational researcher needs to be familiar with the full range of measures available to the behavioral scientist. This includes, to use Webb, et. al. terminology, such "obtrusive" measures as interviews, standardized objective tests, teacher-made tests, systematic observation (face-to-face observation, tape recordings, video-tape recordings) and standardized projective tests, and such "unobtrusive" measures as physical traces through erosion and accretion, e.g., wear on library books or the accumulation of used paint tubes, documents and products, simple observations, and contrived observations through hidden hardware. Each of these are specialized measurement methodologies, some of which are appropriate to one kind of measurement problem and some appropriate to others, but in order to

handle the full range of measurement tasks that he faces an educator or educational researcher needs to be competent in them all.

Limitations of space and time make it impractical to enter into a disucssion of these various methodologies in the present paper. They are complex and each has its own unique set of problems. Nor is it practical to attempt to relate the kinds of evidence needed to insure the trustworthiness of a measure or the adequacy with which a measure is applied to the various classes of measures. To provide the reader with some idea as to the nature of these measures, however, the kind of data that they provide, and some of the particular strengths and weaknesses of each, the major classes of measurement methodologies in the behavioral sciences have been summarized in Table 1. While this brief summary will in no way prepare the reader to use the various classes of measures listed, it is hoped that it will sensitize him to the possibility of their use. Excellent discussions of the operations involved in the various measures appear in Festinger and Katz (1953), Lindzey (1954), Mussen (1960), Gage (1963), Kerlinger (1965), and Webb, et. al. (1966). A sample linkage of measurement methodologies to the evaluation of classes of education outcomes appears in Appendix A. A sample classification of measures that are reviewed by Buros in the 1959 Fifth Mental Measurements Yearbook appears as Appendix B.



5.5

Table 1. Classes of Measures Used in the Behavioral Sciences, some Comments as to the Advantages and Disadvantages of Each, and the Nature of the Data that Derive from Them.

Classes of OBTRUSIVE MEASURES		Data Form	Data Level: Nominal	Data Level: Ordinal	Data Level: Better
Interviews					
Structured: Fixed Alternative Items (agree-disagree, forced choice, rank order)		Category freq. counts (agree, don't know)	X		
Particular Strengths: Constant stimulus conditions; opportunity to clarify misunderstanting; ease of data preparation	Particular Weaknesses: Lacks the flexibility that is possible in an unstructured interview; data costly to obtain	Total scores (e.g., number right, wrong)		X	
		Rankings (e.g., preferences)	X	X	
Structured: Open-ended Items		Category freq. counts (content analysis)	X		
Particular Strengths: Constant Stimulus conditions; opportunity to clarify misunderstanding	Particular Weaknesses: Data costly to obtain, prepare for analysis; requires extensive coder training	Ratings (through content analysis)	X	X	
Unstructured					
Particular Strengths: Freedom to pursue a topic as the situation dictates	Particular Weaknesses: Stimulus only generally known, thus comparability of data questionable; data costly to obtain, prepare; extensive coder training required	Category freq. counts (content analysis)	X		
		Ratings (through content analysis)	X	X	

Systematic Observations ("naturalistic" or "experimental")

	Particular Strengths:	Particular Weaknesses:	Data Form	Data Level: Nom-inal	Ord-inal	Bet-ter
Diary Records	None except insofar as it is better than no recording of an observation	Subject to the error of re-call; subject to limitations of longhand recording; obser-vations usually not focused	Category freq. counts (content analysis)	X		
			Ratings (through content analysis)	X	X	
Check Lists	Provides focus and/or order to one's observations	Cumbersome for recording more than a few items; capable of handling limited range of data at one time	Category frequency counts	X		
Rating Scales	Provides focus and/or order to one's observations; per-mits the summarization of large amounts of information in one score	Meaning of rating uncertain, i.e., cues/criteria it is based upon, thus comparabil-ity questionable; generally unreliable, lacking in evi-dence of validity	Ratings	X	X	
Running Records (preconceived category sets)	Provides focus to observa-tions; handles large amts of info at one time; permits sequential ordering	Requires extensive observer training; data costly to obtain	Category freq. counts (sequential order)	X in sequence		

Standardized Objective Measures

Measure	Particular Strengths	Particular Weaknesses	Data Form	Data Level: Nominal	Data Level: Ordinal	Data Level: Better
Intelligence and Aptitude	Constant stimulus conditions; normative data available for reference; reliability and validity evidence available	Culturally and often symbolically dependent; limited sampling of behavior requiring use of "intelligence"	Part test and/or total test scores		X	X
Achievement	Constant stimulus conditions; normative data available for reference; reliability and validity evidence available	Culturally and often symbolically dependent; limited sampling of behavior requiring use of "intelligence" in reflecting level of achievement	Part test and/or total test scores		X	X
Personality, Attitude, Value, and Interest	Constant stimulus conditions; normative data available for reference; reliability and validity evidence available	Dependent upon the assumption that what one says is related to what one does, is, or believes; validity data limited	Part test and/or total test scores		X	

Standard Projective Measures

Measure	Particular Strengths	Particular Weaknesses	Data Form	Data Level: Nominal	Data Level: Ordinal	Data Level: Better
Association -- Completion	Standardized stimuli that invite free response; some use "realistic stimuli usually inviting strong involvement	Analysis highly dependent upon idiosyncratic interpretations; little evidence of reliability or validity	Association: Part/Total test scores		X	
			Completion: Category freq.;Ratings	X	X	

Teacher Made Tests

Test	Particular Strengths	Particular Weaknesses	Data Form	Nominal	Ordinal	Better
Short Answer (agreement-disagreement, forced choice, or rank order)	Constant stimulus conditions; easy to administer; easy to score; opportunity to standardize	Difficult to develop; response mode limited; test level limited to knowledge or described behavior in a situation response test	Part test and/or total test scores	☐	☒	☐
Essay Tests and Written Documents	Offers opportunity to assess writing skills and integration; relatively easy to develop	Limited sampling of situations relating to property being tested; scoring, standardization difficult	Ratings	☒	☒	☐
Products (articles, e.g., furniture, clothes, models, experiments)	Offers opportunity to assess application skills; easy to develop	Limited sampling of situations relating to property being tested; scoring difficult; little opportunity to standardize	Ratings	☒	☒	☐

ERIC Full Text Provided by ERIC

Classes of NONOBTRUSIVE MEASURES

Class	Particular Strengths	Particular Weaknesses	Data Form	Data Level: Nominal	Data Level: Ordinal	Data Level: Better
Physical Traces	In conspicuous; uncontaminated by the measurement	Of limited utility in the behavioral sciences; gross; validity and reliability hard to establish	Any form of evidence as to erosion, accretion			
Documents and Products	Rich sources of information; relatively easy to obtain; uncontaminated by the measurement process	Limited sampling of situations relating to property being tested; scoring difficult; little to standardize	Category freq. counts (content analysis)	X	X	
			Ratings	X	X	
Simple Observation	Some record of some things, which probably is better than no record at all	Lack of structure and clear rules for assigning numerals to observations to permit it to be called measurememt	Anecdotal Records	X		
			Ratings	X	X	
Contrived Observation (Hidden Hardward)	Permanent record of behaviors permitting leisurely, multiple analysis; uncontaminated by the measurement process	Violation of privacy; unethical	Category freq. counts (content analysis)	X		
			Ratings (through content analysis)	X	X	

PART IV: EVIDENCE NEEDED IN SUPPORT OF THE TRUSTWORTHINESS OF A MEASURE

A basic principle underlying all measurement is that before a measure can be used with any degree of confidence, or before the results of measurement can be viewed by others with any degree of confidence, evidence as to both the adequacy of the measure and the accuracy with which it is applied must be available. This is the case whether one is taking measurements in the physical, the social or the behavioral sciences. Technically, confidence in measurement requires two kinds of evidence: (1) that the measure being used is trustworthy, e.g., that a 12 inch ruler is, in fact, 12 inches long or that a set of categories designed to describe instructional behavior actually describes instructional behavior, and (2) that a trustworthy measure is used in a trustworthy way, e.g., that an accurate ruler is used accurately and it is used to measure distance rather than weight or that a set of categories that describes instructional behavior is applied accurately and it is used to describe instructional behavior rather than classroom management behavior.

Historically two criteria have been used in evaluating the adequacy of measures in the behavioral sciences, those of reliability and validity. Generally speaking, validity refers to the extent to which a measure measures that which it is intended to measure, and reliability refers to the extent to which it does so consistently. Within these broad meanings, however, a number of more specific meanings exist, for example, test-retest and split-half reliability, face validity, construct validity, concurrent validity, predictive validity. Unfortunately, while these have come to be relatively

common terms in educational parlance, the most recent document relating to the matter (Standards for Educational and Psychological Tests and Manuals, 1966) indicates that there is far from unanimous agreement as to what it is that they refer to or the conditions under which they have relevance.

In the opinion of the writer much of the confusion that has and still does exist around these issues stems from the tendency to (1) force all classes of evidence needed in order to judge a measure to be sound or trustworthy into the 'reliability or validity mold,' i.e., the constructs have been overworked, (2) to treat the various kinds of evidence needed as to the adequacy of a measure as if they were independent of the nature of the measurement being made, and (3) to fail to distinguish sharply between evidence needed in support of the trustworthiness of a measure and evidence needed in support of the adequacy with which a trustworthy measure is applied. In contrast to this procedure the point of view being taken here is that (a) the criteria by which one judges the adequacy of a measure differ depending upon the nature of the measure being taken, i.e., depending upon whether it is a low inference or a high inference measure, or whether either of these is being used as a predictor, (b) a sharp distinction has to be made between the evidence needed as to the adequacy of a measure and the adequacy with which a measure is applied, and (c) a new set of constructs can describe these criteria better than the old constructs of reliability and validity. For purposes of clarity the discussion of evidence needed in support of the trustworthiness of a measure is organized around the types of measures being taken.

Support of Low Inference Measures

Five criteria are deemed essential in assessing the adequacy of a low inference measure in the behavioral sciences: (1) its relevance, (2) its representativeness, (3) its fidelity, (4) its consistency, and (5) its accuracy. Some attention also needs to be given to the practical qualities of a measure, but this is not the same order of consideration as 1 through 5.

Relevance. As used here the relevance of a measure refers to the extent to which it is a logically appropriate measure for that which is to be measured. Technically, the relevance of a measure is determined by two considerations, the selection of an appropriate measurement methodology and the selection of appropriate item types to be used within the methodology. If either selection stretches credulity, then the relevance of the test is suspect. As used here, the concept of relevance is intended to replace the earlier concepts of face or content validity (see Standards, 1966). In terms of the concept of error variance, the greater the relevance the less the likelihood of constant error.

Representativeness. The representativeness of a measure refers to the extent to which it samples the situations in which the behavior under consideration is reflected. Generally speaking, the greater the number of situations sampled the more confidence one can have in judging whether the behavior observed is or is not a functional part of the behavioral repertoire of the individual. As in the case of relevance, the primary means for determining the representativeness of a measure is through logical analysis: if the situations sampled are not representative of the full range of situations available for sampling then the measure is suspect. A formal means for estimating representativeness also exists; namely, the comparison of scores obtained on two forms of the measure. If scores are similar on both forms then it may be assumed

(granting the acceptability of the measure by logical analysis) that the sampling of situations is adequate. The two forms of the measure may be developed separately or they may be formed simply by randomly dividing the situations sampled into two equal halves. This procedure will be recognized as "alternate form" and "split-half" reliability measurement, but as used here the concept of representativeness is intended to replace these earlier concepts. In terms of the concept of error variance representativeness is also a means of reducing constant error.

Fidelity. As used in the present paper the fidelity of a measure refers to the extent to which a measure draws upon the performance of the concrete behavior that is the target of the measure; does it require behavior that is isomorphic to the objective of measurement (identical to it) or does it call for behavior that is only in some way related to it. Ideally, all measures of educational or other performance objectives should be isomorphic, but since the realities of educational or other settings sometimes make this impractical the general rule to be followed is to make one's measures as high in fidelity as is practically possible.

Many years ago Lindquist (1950, p. 146) proposed four alternative approaches to the measurement of educational objectives: (1) give the examinee occasion to do some of the things that are specified by the objective (an isomorphic, situational response test); (2) give the examinee occasion to do things similar to some of those specified by the objective, (a "related behavior" situational response test); (3) describe a situation in which the examinee would have occasion to do what the objective specifies, and then ask him to tell what he would do in this situation or how he would do it (a "verbalized behavior" situational

response test); and (4) discover whether or not the examinee knows the facts, rules, principles, etc., that are presumably essential or conducive to the desired behavior, (a "knowledge" test). At the time Lindquist wrote his paper educators were measuring essentially at the knowledge level, and his plea was to get them to move to a higher level of fidelity. Today they are still measuring essentially at the knowledge level, and the plea is still relevant. This is not meant to imply that measurement should avoid focusing at the knowledge level; indeed, many educational objectives are focused entirely at that level. Also, there is obvious truth in the argument that knowledge is essential or conducive to the overt behavior with which an ultimate educational objective is concerned, i.e., there is a relationship between what and how much an individual knows and how he will behave in certain situations. The point of the discussion here, however, is that while the measurement of knowledge is in many instances a worthwhile goal, it cannot substitute for the situational response type measures as measures of educational objectives.

To date, the concept of the fidelity of a measure has not been emphasized in the literature on test theory, so the concept has had little empirical testing and carries no analogues. On the surface, however, it appears to be a useful concept and so it has been included here. In terms of the concept of error variance it too contributes to the reduction of constant error.

Consistency. Consistency is that attribute of a measure which speaks to the reliability with which it measures that which it purports to measure. As discussed earlier all measurement, whether in the physical on the behavioral sciences, contains a certain amount of chance or random error. Two sets of measurements of the same characteristics or properties of the same individuals

ERIC Full Text Provided by ERIC

will never exactly duplicate each other. This is termed the unreliability of measurement. At the same time, however, repeated measurements of a property within an individual will (if the measure is at all appropriate), show some consistency. For example, if a boy was the best reader in the room the first time that a class was tested it is highly probable that he will be among the best readers on another testing, even though he may not be the best. This tendency toward consistency for a repeated set of measures is what has been termed historically as "test-retest" reliability. Technically, consistency may be defined as the extent to which a set of measurements is free from variance due to random error, but in fact it tends to be more a measure of the consistency of the individual that is being measured than it is a measure of the test per se.

A number of factors may contribute to a lowering of the consistency of an instrument: (a) response variation by the subject due to fatigue or illness, an "incorrect response set", etc., (b) variation in administration or the administrator of the test, (c) variations in scoring. As long as the sources of error remain unsystematic, however, i.e., not constant, their threat to measurement is not great.

Within this framework there is essentially only one procedure for obtaining an estimate of the consistency of a measure, namely, through repeated measurement of an individual or set of individuals with the same measure. While there are a number of limitations inherent in such a procedure, for example, if one waits very long between measures individuals will change so much that the repeated measure will of necessity by quite different from the first, or if one retests too quickly the second score will be subject to recall error from the

first, it is still the one measure that provides data relevant to the concept of consistency or reliability. Consistency estimates are needed for all measures, whether direct or indirect, for random error always enters measurement.

A special problem is encountered in consistency estimation when the measurement methodology being used does not call for the rigid control of test stimulus materials, e.g., when the measure is the description (categorization) of the free play of children, teacher-child or parent-child interaction, group problem solving behavior, etc. Here the stimulus situation is never the same twice (unless one is working in a controlled, experimental situation) and consequently there is no reason to assume that the behavior in question will ever be the same twice. Nevertheless, it is still possible to obtain an estimate of the consistency of these kinds of measures if one is interested in doing so. The only requirement is that the <u>situation that is being observed be as similar as possible on repeated measurement occasions</u>, e.g., observing the same children in the same play area at the same time of the day under as many of the same conditions as possible.

<u>Accuracy</u>. The accuracy of a measure has often been confused with reliability, but they are in fact two quite different concepts. A test may be quite reliable (a person may consistently receive a similar score upon repeated measures) but it may not be accurate, that is, the test could show him with x amount of a characteristic when in fact he has an x + 1 amount. Nor is accuracy comparable to relevance, representativeness, fidelity, or validity. A test may be relevant, etc., that is, it tests x when it is supposed to test x, but it may be so gross as not to be able to distinguish between x and x + 1.

From a technical point of view, the concept of accuracy has no legitmate meaning in the behavioral sciences. This is because these sciences have access only to nominal and ordinal level data (see above); one must have an absolute measure against which to judge, such as those which exist in the Bureau of Standards, in order to have a measure of accuracy. This first of all requires measurement at the ratio level. Nevertheless, the concept of accuracy is a comforting one and is retained in the present list for whatever value it may have.

Practicality. While practical considerations can never justify the use of a test which gives worthless information, a relevant and technically sound test cannot be used where it is impractical. Thus users of tests need always to seek a viable "trade-off" between the demands of good measurement and the demands of reality. The major factors which need to be considered generally in relation to practical matters are cost, time required for administration and scoring, ease of administration and scoring, the availability of comparable test forms, user acceptability and potential usefulness of results. Nearly all textbooks in educational and psychological measurement deal with these topics, so they will not be pursued here.

Support of High Inference Measures

It will be recalled that high inference measures differ from low inference measures primarily in terms of the concreteness of the properties of the concept that are being measured. In contrast to low inference measures, where the properties being measured are directly observable and/or manipulatable, high inference measures center in "indicants" of that which is being measured. Put in other terms, high inference measurement centers upon the measurement of qualities that are only inferred (constructs). Intelligence, learning style, anxiety,

interest, and aptitude are concepts of this kind. One cannot see, hear, touch, feel, smell or taste these qualities; they can only be inferred from that which is observed.

With such a basic difference existing between direct or low inference and indirect or high inference measures one would expect the measurement processes involved to be quite different. This is not the case. The steps or operations involved in direct and indirect measurement are exactly the same. Also, the issues of relevance, representativeness, fidelity, consistency, accuracy and practicality are as crucial in high inference measurement as they are in low inference measurement. The one difference that exists in these two classes of measures is the requirement that high inference measures have some empirical evidence that they are in fact measuring that which they are supposed to be measuring. This calls for an attribute above and beyond those of relevance, representativeness, fidelity, consistency, etc., namely, evidence of construct validity.

As with the concepts of relevance, representativeness and fidelity, construct validity also relates to the issue of whether a measure is measuring that which it is supposed to measure. It differs from these other criteria, however, in the kind of evidence that is permissible in its support: the attributes of relevance, representativeness and fidelity involve essentially nonempirical, analytic, judgmental evidence (the notions of "face" or "content" validity) whereas the attribute of construct validity requires experimentally obtained empirical evidence. This requirement makes validity evidence costly and difficult to obtain, for it requires a full-scale research program to do so, and as a consequence only the better "standardized" measures are likely

to have it. Rarely will teachers be able to obtain evidence of construct validity for the tests which they develop within a class.

Two kinds of empirical evidence are desirable in support of the construct validity of a measure: (1) correlation with another measure that is known to be a valid measure of the property under consideration (concurrent validity), and (2) the experimental verification of hypotheses which involve the concept being measured and which have used the test being considered in the research that has provided that verification (theoretical validity). Either or both kinds of evidence provides confidence that the measure "is in fact measuring that which it is supposed to be measuring."

One further comment about construct validity: just as the evaluation of an instructional system permits the clarification and testing of the conceptual framework which underlies it (the hierarchy of enabling objectives), the pursuit of construct validity permits the clarification and testing of the constructs used in a discipline. In this sense, obtaining evidence of construct validity is as much a conceptual or theory developing activity as it is a measurement activity. While most practicing educators will not be involved in the development of instruments to test concepts or theory, they need to be familiar with the idea of construct validity, for whenever they use indirect measures in their research they will have to support them with evidence of this kind.

The evidence needed in support of the trustworthiness of both low and high inference measures is summarized in Table 2.

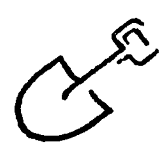
ERIC

Table 2. Evidence needed in support of the trustworthiness of low and high inference measures.

LOW INFERENCE MEASURES	HIGH INFERENCE MEASURES
Relevance	Relevance
Representativeness	Representativeness
Fidelity	Fidelity
Consistency	Consistency
Accuracy	Accuracy
	Construct Validity

Low and High Inference Measures Used as Predictors

Much of science is involved with prediction, for once prediction is possible it enables control. Educators are also concerned with prediction: who will pass and who will fail? Who is a good college risk and who isn't? Who is likely to benefit from curriculum A and who from curriculum B? To permit prediction there must be measurement, for one predicts whatever he is predicting from some set of measures. These may be as general as age or sex or socio-economic status, or as specific as number of questions asked or number of comments made in a class. Of significance to the present discussion, however, is the fact that whenever any measure is included in a generally used prediction scheme, whether it is a low or high inference measure, there must be evidence of its predictive validity. Predictive validity, like construct validity, requires empirical evidence, i.e., that the measure does in fact predict that which it is supposed to predict, and this requires a full scale research effort to demonstrate. As a consequence, it is likely that teacher-made tests used as predictors will rarely have demonstrated predictor validity.



5.6

PART V: EVIDENCE NEEDED IN SUPPORT OF THE COMPETENCY WITH WHICH A MEASURE IS APPLIED

Evidence as to the adequacy or trustworthiness of a given measure is obviously that which is looked for first by a potential user of a measure or by persons reading the reports of users of measures, i.e., reports of research, development, or evaluation efforts. For readers of such reports, however, and consequently for the user of a given measure, there is another kind of evidence needed, namely, evidence as to the competency or accuracy with which a measure is applied. This is evidence of a markedly different kind than that reflected in judgments of relevance, representativeness and fidelity or empirical estimates of accuracy, consistency or construct validity, but it is equally critical. An entirely adequate or trustworthy measure, for example, the Binet Intelligence Scale or the Flander's Interaction Analysis System or a carefully tested interview schedule can be totally useless or full of error unless it is administered with care and competence. Unfortunately little attention has been given this dimension of measurement in discussions of measurement theory, or even in reporting the results of research, development or evaluation efforts, and as a consequence the users of measures are relatively insensitive to it.

The kind of evidence needed in support of the competency with which a measure is applied is closely linked to the concept of instrumentation. In the physical sciences most measures are taken with the aid of some kind of measuring device. These range from the relatively simple ruler and scale through the amp meter, thermometer and speedometer, to the highly sophisticated gadgetry required to monitor the functioning of a rocket or the path of particles that derive from

the bombardment of atoms. By and large these devices are characterized by some kind of "hardware" which permits "readings" to be taken of that which is to be measured. They are also characterized by accuracy and a long history of careful development. In the behavioral sciences a great many measuring devices also exist, for example, the galvanometer, paper and pencil tests of intelligence and achievement, questionnaires and opinion surveys, and they are intended as close parallels to the measuring devices used by the physical scientist. Put in another way, they represent the counterpart of the physical scientist's "hardware," and the translation of responses to them represent the counterpart of the physical scientist's "readings" of that which is being measured. Generally speaking, such readings are relatively easily taken and require little formal evidence of ability of an observer to take them accurately. This is especially the case if these kinds of measuring devices have had a great deal of careful work in their development.

A great many measuring devices exist in the behavioral sciences, however, that do not at all parallel such an approach to measurement, _even though at first glance they may appear to do so_. These include individual measures of intelligence, such as the Binet or Wechsler-Bellevue, interview schedules where respondent's answers are coded or categorized, projective measures of personality and observational measures such as the Flander's system or Bale's system of interaction analysis. In one important sense these measures are similar to those reviewed above, namely, they are characterized by careful and extensive work in their development. They are critically different, however, _in that the responses to the measures are not translated into "readings" that can be taken easily from the measure_. Without exception these measures depend upon an

ERIC

elaborate, detailed, highly complex "coding" process for their data, and a great deal of training has to be given the person using these kinds of measures to guarantee that the coding process be done in a trustworthy manner. It is also critical that evidence as to the accuracy with which this coding can be done be accumulated before formal measurement is undertaken and be presented in detail for the consumers of the resarch development or evaluation effort. Generally speaking, data of this kind is referred to as observer or coder or tester reliability.

The general lack of concern on the part of measurement theorists or the users of measures in the behavioral sciences for evidence of the competency with which measures are applied probably stems from the traditional emphasis in measurement upon instrumentation. In measurement that involves tests or some other form of instrumentation, minimal involvement on the part of a person is required to administer them. For measures such as these one often finds that no evidence is provided as to the accuracy with which the measure has been applied. The apparent justification for this procedure is that the administration of measures such as these is so straightforward, and their dependency upon an administrator's involvement so minimal, that it is simply assumed that anyone at all trained in the administration of such measures can and will do the job accurately.

Much the same point of view holds with respect to measures which require high involvement on the part of a person to administer but minimal dependency upon interpretation or on-the-spot scoring or coding, for example, an interview which requires "yes-no" or "agree-disagree" answers or one which requires only that the respondent's answers are recorded on audio and video tape. As in low

involvement measures it is simply assumed that with some training the administrator of the instrument can and will do so accurately. To a large degree this philosophy holds even for some measures that require high administrator involvement and heavy dependency upon interpretation or scoring, for example, the Binet or Wechsler-Bellevue intelligence scales; evidence is rarely provided as to the accuracy with which these measures have been applied because it is assumed that the administrators of them have had extensive training in doing so.

A different orientation has to be taken to the accuracy of use issue, however, when a measure is of a kind that the administrator of the measure is required to make complex judgmentsor categorizations or codings of . ' which is to be measured. This is the case whether the measure requires on-the-spot coding such as that required in the content analysis of audio or video tapes or the protocols of interviews. Under either circumstance detailed evidence must be available as to the accuracy with which these measures are being applied. Generally speaking, this requires for its demonstration the comparison of records made by two or more persons who either observed an event or segment of behavior simultaneously but independently or who coded the same record of an event or segment of behavior simultaneously but independently or who coded the same record of an event or segment of behavior independently. As indicated earlier, an additional constraint on the user of coding systems is that evidence of this kind must be available prior to the use of the measurement system.

On the basis of the rather sharp distinction that can be drawn between measures that require only recording when they are used, and measures which require some kind of scoring or coding or classification before they can be recorded, it is possible to identify three kinds of evidence needed

in support of the accuracy or competency with which measures are applied: 1) evidence as to the reliability of the administration of the measure, 2) evidence as to the reliability of the recording of the data that derive from the measure, and 3) evidence as to the reliability of the scoring or coding or classification procedure that precedes recording. Since some measures require only adminstration and recording it follows that for these measures only the first and second types of evidence are required. For measures that require scoring or coding prior to recording all three kinds of evidence are required. These distinctions are summarized in Table 3.

Table 3. Evidence needed in support of the competency with which measures are applied.

Measures Which Require Only Recording	Measures Which Require Scoring/Coding/Classification Before Recording
Administrative Reliability	Administrative Reliability
Recording Reliability	Recording Reliability
	Observer/Coder/Scorer Reliability

By combining Tables 2 and 3 a summary classification of the kinds of measures used in the behavioral sciences, the kinds of evidence required as to their adequacy, and the kinds of evidence required as to the accuracy with which they are applied can be compiled. This is presented in Table 4.

Table 4. A Summary Classification of Measures Used in the Behavioral Sciences, the Kinds of Evidence Required as to their Adequacy, and the Kinds of Evidence Required as to the Accuracy with which they are Applied.

Kinds of Evidence Needed in Support of...	Low Inference Measures		High Inference Measures	
	Those Which Require Only Recording	Those Which Require Scoring, Coding, or Classification Before Recording	Those Which Require Only Recording	Those Which Require Scoring, Coding, or Classification Before Recording
the Adequacy of the Measure	Relevance Representativeness Fidelity Consistency Accuracy —	Relevance Representativeness Fidelity Consistency Accuracy —	Relevance Representativeness Fidelity Consistency Accuracy Construct Validity	Relevance Representativeness Fidelity Consistency Accuracy Construct Validity
the Accuracy with which the Measure is Applied	Administrative Reliability Recording Reliability	Administrative Reliability Recording Reliability Observer/Coder/Scorer Reliability	Administrative Reliability Recording Reliability	Administrative Reliability Recording Reliability Observer/Coder/Scorer Reliability

Administrative Reliability

In applying any measure in the behavioral sciences that involves an intrusion into people's lives administrative issues become critical. One has to obtain the permission of those being measured to be measured, provide the desired structure or evoke the desired set within which to administer the measure, be certain the person being measured understands his role or responsibility or task in the measurement process, reduce all possible sources of contamination or error, record identifying or other relevant administrative data, etc. Only when there is reason for believing that all of these matters have been properly cared for can one accept the data coming from the application of a measure as trustworthy.

In the application of most educational evaluation or research measures administrative procedures are relatively well known, relatively straightforward and rather easily followed: those being measured are usually in a setting where measurement or "testing" is common and those doing the measuring are usually aware of the basic rules of test administration and in a position where the gaining of cooperation, the control of irrelevant sources of variance, etc. is relatively manageable. This is not always the case however, and as the measurement situation departs from the traditional or the accepted or the controllable increased care has to be taken with respect to both administering measures and describing the procedures followed in their administration. Examples of measurement situations in which educators might find themselves which would fit this description include the assessment of parents' attitudes toward a school program, the study of the relationship between teacher behavior and pupil outcomes and the study of the relationship between teacher morale and

administrative practices. In all of these studies the administration of measuring instruments varies considerable from the familiar classroom testing situation; the respondents are not accumstomed to being measured, and the one doing the measuring probably has little control over those he is measuring. To illustrate the care that needs to be taken in administering measures under these circumstances, and to illustrate the care that needs to be taken in describing these procedures to those who will be reading the research or evaluation report, the administrative procedures followed in a study that compared the interaction patterns of mothers and their preschool children in a free play laboratory and in the home (Schalock, 1956) will be described.

> After securing the cooperation of the mother for the study, an appointment was made for an interview with both the mother and child in the home. The purpose of the interview was primarily that of establishing a "working" relationship with the mother, and thereby making her feel relatively comfortable during the actual observations. It was assumed that the more relaxed and comfortable the mother felt in the observer's presence, the more nearly "customary" her behavior would be.
>
> In the course of the interview the observer explained the purpose of the observations, the observer's role in the observations, how the observations were to be made, and what was expected of the parent and child during the course of the observations. The observer pointed out the necessity of his not establishing a close relationship with the child. The desirability of having the mother engaged in an activity that was routine during the period of observation was also stressed. The entire interview was focused on trying to convey to the mother the desirability of customary routine behavior during the course of the observations. At the close of the interview appointments for the observations were made. . .
>
> Having become acquainted with the parent and child during the course of the interview, and having explained what was to be expected during the observation, little difficulty was encountered in actually beginning the observations. After an initial greeting, the observer simply indicated to the parent that he was ready to begin the observation. At this cue the mother returned to her regular activities and the observation proceeded.

Observations were not limited to behaviors ocurring inside the home. Frequently the child played out of doors during the entire observation. Whenever the child was outside for any length of time the mother would also go outside. This procedure was adopted in order that as much interaction take place as possible in a thirty-minute interval. It did not mean, however, that the mother necessarily took part in the activity in which the child was engaged.

The observer's relationship to the mother and child while in the home was carefully defined. He was to stay close enough to the interaction to enable him to see and hear what was taking place, but was to remain as much apart from the interaction as possible. If the child made friendly or attention getting overtures toward the observer, he was recognized but not interacted with in an overt, verbal sense. As a rule the children did not need more than this brief recognition to enable them to return to or begin an activity. If the child persisted in his overtures, observation was stopped and a word of explanation was given. It was pointed out to the child that the observer was working and could not talk with him at this time, but would as soon as he was through. The mother was also generally helpful in this respect. In no instances were overtures continued by a child throughout an observation period.

The position taken by the observer in relation to the occurring interaction was dependent primarily upon the scene of the interaction. In any setting, however, the observer strove to maintain as unobtrusive a position as possible. If the mother and child were in a relatively small room, the observer usually stationed himself in or near the doorway to the room. In this way the observer was quite well apart from the immediate scene of interaction. He was able to move out of the way of either the mother or child if they left the room, or was in a good position to follow the scene of interaction if it changed. In the case of the mother and child being in a larger room, the observer could again operate from the doorway if he could assume an unobtrusive position on either side or end of the room. When the interaction occurred outside, there was no one particular place for the observer to stand. The criterion here again was to remain as unobtrusive as possible, but to be within seeing and hearing distance of the mother and child.

Whenever the mother and child were not together, that is, whenever one could not be seen by the other, the observer remained with the mother. The rationale underlying this procedure was that the observer's presence with the mother might act as a stimulus to the child's attending the mother or at least operating within the same room as the mother, and, contrariwise, being alone with the child might tend toward observer-child interaction. (pp. 18-21)

Recording Reliability

Ordinarily, formal or systematic data on the reliability with which the user of a measure records the data that derive from the measure is not required. Rarely, for example, in the usual pursuit of educational research, does one expect or require that data be presented to support the accuracy with which a researcher records height or weight measures, achievement test scores, days absent, etc. Nor does one expect this kind of data in support of the recording of the category or coding judgments that are made in conjunction with measures that are dependent upon such judgments: once a behavior is labeled as "evidence of hostility" or "indicative of a conservative attitude" the recording of that level is not questioned.

This is not to imply, however, that such data are irrelevant or in all cases unnecessary. Recording error is always possible, even when the most sensitive or common of instruments are used. For example, it has been shown that repeated recordings of the weight or height of a child within a given hour by a single researcher almost invariably produce a range of scores. As a consequence, depending upon the degree of accuracy demanded by a study, the user of a measure must always make a judgment as to whether he should or should not obtain evidence as to the accuracy with which the recording of data occurs. If the demand for accuracy is high, or if there is some reason to believe that recording accuracy is low, then evidence should be presented which indicates the reliability of recording; if the demand of a study for accuracy is not extremely high, and there is no reason to believe that recording accuracy is low, then such data need not be presented. If data on the reliability of recording are required the methodology used in obtaining it is the same as that used in obtaining data on the reliability of coders (see below).

ERIC

Coder Reliability

For measures that require classification or coding as a part of the measurement process, evidence as to the accuracy with which coders can apply the categories or rating scales which make up the system is mandatory. Without evidence of this kind there can be no confidence in the measure--even if evidence as to its adequacy, i.e., its relevance, representativeness, fidelity, etc., and its reliable administration are available. As with construct validity, evidence as to coder reliability must be empirical in nature. This is to say that while much of the evidence that supports the adequacy of a measure or the reliability of its administration can be logical or descriptive in nature, this is not the case with respect to coder reliability. Hard data has to be presented, <u>before observation for purposes of data collection begins</u>, that coders can independently and accurately assign categories as the measurement system dictates. <u>Moreover, evidence of this kind must be presented for each individual category or rating scale in the system</u>. It is not enough to present a single measure of the accuracy with which an observer can apply the total category system or the total set of rating scales used.

The need for reliability estimates on individual category or rating scales can be illustrated by a circumstance that occurs frequently when using category sets as part of a measurement system. It is possible to show, for example, that observers can be accurate with five categories in a system and inaccurate on ten, yet still have an over-all accuracy measure that appears to be relatively high. This can come about simply by the five categories on which they are accurate accounting for 90% or so of the interacts recorded, i.e., they are the "high frequency" categories, and if they are accurate on these they can be totally inaccurate on the ten other categories and not have it make a great deal

ERIC Full Text Provided by ERIC

of difference in their over-all reliability measures. Concretely, if the five categories on which they are highly accurate have a combined frequency of 180 and the ten categories on which they are inaccurate have a combined frequency of 20, it is possible for a 90% accuracy figure to be evidenced on the overall system even though the observers are totally accurate in their application of two-thirds of the categories in the system. This is an extreme example but it is for exactly this reason that observer reliability data must be available for each category that makes up an observational system.

How does one proceed to demonstrate that coders can "independently and accurately" assign categories or rating scale placements as the system dictates to that which is being observed? The basic element in the procedure involves the comparison of the records made by two observers observing the same behavior simultaneously but independently, i.e., without discussing or in any other way comparing their records. Also, the two sets of observations should be made from approximately the same vantage point. . .the observers should be close to one another in physical proximity, and represent reasonably adequate samples of the behavior being observed, e.g., a 20-minute sample on three separate days. With independent records and an adequate sample of behavior as a point of departure the calculation of coder reliability can be undertaken.

The most desirable procedure to be followed in establishing coder reliability is to make an item by item comparison of the content of the two records for their agreement or disagreement. This requires a detailed analysis of the two records, on a time line, and for the users of some systems this represents a level of detail beyond the capacity of their resources. It also requires a high level of specification in the determination of "categorization" errors in contrast to

"timing" errors, and the translation of these data into relatively complex reliability formulae. Because of the limits of space and time the details of these calculations cannot be dealt with here. For those who desire such detail see Schalock and Micek, 1968.

In addition to the presentation of reliability data for individual coders on each of the categories or rating scales used in a judgment-dependent measurement system, one further requirement exists: administrative and coder reliability need to be established separately for each setting in which the measurement system is to be applied. This is the case if one observes first in the classroom and then in the home, or first in the primary grades and then the upper grades, or even if one first codes video tapes of a particular situation and then makes live observations of it. Whenever a task or setting varies significantly from that for which reliability of coding was established, reliability must be established for the new application. The need for independent reliability checks in differing situations can be illustrated by again citing experience gained in getting ready for the comparative study of mother-child interaction in the home and laboratory (Schalock, 1956, pp. 58-61).

> Having demonstrated reliability in the therapy situation, the next step was to observe a mother and child in the play room. It was not supposed that there would be any great difficulty in obtaining reliable observations in this situation because of its similarity to the therapist and child situation, that is, the mother and child would be operating in the same physical setting as had the therapist and child. However, upon observation of the mother and child in the play room, observer agreement proved to be much lower than that obtained on the therapist and child. The therapist-child observations were all made on the same therapist who was for the most part dealing with children who were severely disturbed and whose interactive behavior was ordinarily quite limited. The behavior of the therapist was also quite consistent toward all of the children seen in therapy. These factors, the consistency and limited range of behavior, enabled the observers to become very familiar with the types of behavior the therapist would employ and made it possible, as it were, to anticipate the types of behavior that would accompany a given situation.

Mother-child interaction in the laboratory differed from this pattern considerably. The mother and child exhibited a much wider range of interactive behaviors, and for the most part these occurred within the framework of a generally higher activity level. This necessitated the use of categories that had not appeared in the therapist-child situation. The increased level or speed of the interaction resulted in problems in recording and timing that had not previously been experienced. In order to demonstrate reliability on the observation of the mother and child in the play room, the observers once again had to undertake the process of refinement and redefinition of categories. In addition to this, however, there was the need to ncrease the overall level of observing ability in order that this more complex and active interaction might be reliably observed. Two months were devoted to further observational training and refinement of categories before an acceptable level of reliability was established on mother-child interaction in the play room.

Observations were then immediately undertaken on the mother and child in the home. To familiarize the observers with the problems associated with home observations, several practice sessions were employed before actual reliability tests were made. The problems encountered in these first home observations were felt to be primarily a function of the more difficult observational setting rather than the inadequacy of the categories, the level of interaction, or the unfamiliarity of the observers with the categories.

While observing in the home, the mother and child were encouraged to continue as freely as possible in their customary routine of behavior, even if that routine carried them out of doors. With such an orientation on the part of the mother and child, the observers frequently had to observe from a standing, mobile position in order that they might follow the shifting scene of interaction. To make this mobile recording possible, the recording blanks were placed on a clipboard that had a stopwatch attached. The observers could then hold the board and record while moving. This arrangement, however, did create problems. At first the observers experienced considerable difficulty in managing the clipboard while writing. They also found that the position of the stop watch was inconvenient to easy timing since it was attached to the clip on the board necessarily in such a way as to force a side-angle view of the second hand while recording.

Closely associated with the more difficult recording process was the problem of simultaneous observation from similar perspectives, since dissimilar observational positions would give rise to the perception of different behaviors. To keep errors of this kind at a minimum, the observers remained in close physical contact throughout the observation. If the scene of interaction changed, the observers would follow simultaneously while remaining in close physical contact.

One other problem that was evidenced more clearly in the home situation than in the play room was that of observer fatigue. With experience, however, it became possible to observe for a forty-five minute period without experiencing severe fatigue.

Only a short period of training was needed by the observers to overcome the problems encountered in the practice home observations. After the practice observations in the home, arrangements were made with families entirely unknown to the observers for purposes of final reliability measurement. It required only a week to establish satisfactory observer reliability for the home observations.

In conclusion, it is not possible to assume reliability of observation in any situation in which reliability has not been specifically demonstrated. A change in participants or a change in the interactional setting will produce differing interactions, and in order to use data derived from these interactions with any degree of confidence, reliability must be demonstrated for that specific situation.


5.7

PART VI: NOTES ON THE DEVELOPMENT OF EDUCATIONAL OUTCOME MEASURES

As indicated elsewhere the development of trustworthy measures requires knowledge and skill that is relatively specific to each class of measure available to the behavioral scientist, and the chapter cannot accommodate the detail that would be required to prepare the reader to be able to develop all such measures. As a consequence instrument development procedures are illustrated only for one class of measure that the educational researcher or evaluator ultimately must use, namely, that of educational outcomes. Outcome measures were selected for illustrative purposes because of their pervasiveness and significance.

ERIC Full Text Provided by ERIC

Steps In the Development of Low Inference Measures

Most educational outcomes (knowledge, skill, sensitivities) can be assessed by means of low inference measures. While a rather wide range of standardized achievement measures exist (see Buros, 1965), most measures that are used in assessing educational outcomes are of the informal, teacher-made variety. Considering the almost limitless number of educational objectives open to pursuit, and the part that individual teachers can play in their selection, this probably will continue to be the case for many years to come. With this as a frame of reference the discussion which follows is addressed to the development of teacher-made low inference measures.

Within this focus four topics are considered: 1) planning the measure, 2) developing the measure, 3) trying out the measure, and 4) evaluating the measure. While more classroom teachers rely heavily upon paper and pencil measures of educational outcomes the procedures and attributes reviewed within these topics are applicable generally across all classes of measures within the behavioral sciences (see Appendix A). For persons wishing concrete, specific guidance in the development of paper and pencil classroom tests, five recently published books provide excellent references: Sax, G. The Construction and Analysis of Educational and Psychological Tests: A Laboratory Manual (1962); Ahmann, J. S., and Grock, M. D. Evaluating Pupil Growth (1963); Stanley, J. C. Measurement in Today's Schools (1964); Gronlund, N. E. Measurement and Evaluation in Teaching (1965); and Ebel, R. L. Measuring Educational Achievement (1965). The pros and cons of various item types, including essay items, how these can best be developed, how to administer, score and use test, etc. are discussed in these references in detail.

Planning the Measure. Given an educational outcome to be measured, and some general guidelines as to the indicators that are acceptable as evidence of the realization of that outcome, there are three tasks to be accomplished in the test planning phase: 1) a decision as to the fidelity of the specific items to be used in the measure, that is, the extent to which the response to items within the measure will require the performance of behavior representative of the real-life performance of the objective in contrast to the performance of behaviors which are only related to it, e.g., building a house vs. describing how one would be built or acting considerately vs. describing what acts of consideration involve, 2) a decision as to the representativeness of the items to be used in the measure, that is, the number and range of situations to be sampled by the items in which the behavior is reflected, e.g., does one ask to have one house of one kind built (to test for transfer?), and 3) the weight to be given to each item. The question of fidelity, of course, revolves around the issue of inference: is one willing to accept as evidence of the accomplishment of a terminal objective something less than the actual performance of that objective? Is one willing to accept a behavior supposedly related to the terminal objective as evidence of the objective, and if so how close or how distal can the relationship be and still be acceptable? Is one willing to accept as evidence of considerateness toward parents, for example, a verbal statement to the effect that the examinee is considerate? Would observation of considerateness toward animals or children be any better? Would considerateness toward peers or a teacher be better still? These are the kinds of issues involved in the question of fidelity and they are crucial in the process of achievement measurement.

The question of representativeness revolves around the issue of generalizability: how many and what kind of situations must be sampled or observed in order to feel confident that the criterion behavior is sufficiently well established that it is a functional part of the behavioral repertoire of the examinee? The question of item weighting is simply a matter of deciding whether some items will be more informative of criterion performance than others.

Unfortunately, there are no firmly established rules to guide decisions on any of these three questions. The level of fidelity desired in an instrument is a matter of personal preference (though one probably should be able to defend it) interacting with the reality demands of a situation. It is sometimes difficult to set up a testing situation which has architectural students building houses. The extent of evidence required as to the representativeness of the performance measure is also largely a matter of personal preference and reality demands: generally speaking, the broader the sampling of items the better the evidence, but also the higher the cost. Establishing rules for weighting item responses is also arbitrary, though in most short answer tests of educational achievement equal weight is given to all items. In measures involving essay questions, or products, or interpersonal behavior, the matter of assigning item weights is equally arbitrary. The single rule that could be thought to operate in all these decisions is that of logic: the fidelity of items, the representativeness of situations sampled and the assigning of item weights must somehow meet the demands of elementary logic. If a decision on any of these matters wrenches one's credulity, it is likely that even this simple criterion has been violated.

The lack of established rules for making planning level decisions should not be taken to mean that these are insignificant decisions. Indeed, of all the

ERIC

decisions made in relation to measurement they are perhaps the most basic, for all else stems from them.

Developing the Measure. Given clarity as to the fidelity of items to be used in a measure and the range of situations to be sampled by those items, one is then ready to proceed with the development of the measure. Three steps are involved in this process: 1) specifying the set of operations that are to be used in assessing the status or condition of the objective (the development of an "item" pool), 2) specifying the conditions under which these operations are to be made (the development of a set of directions) and 3) specifying the rules by which numerals are to be assigned to these operations (the development of a "scoring" system). In deciding upon the set of operations to use (the "item types") it is necessary to look to the full range of measurement methodologies available to the behavioral scientist (see Table 1). In so doing the aim is to find the methodology that provides the most appropriate set of operations for the assessment of a given objective at a given level of fidelity. Thus, if the objective being measured is considerateness toward others, and the level of fidelity is set at "identical elements", i.e., it requires the performance of the concrete behaviors specified in the objective, the measurement methodology most appropriate to the task would probably be systematic observation. Other appropriate methodologies would include the interview, simple observation and contrived observation. Perhaps the least appropriate methodology would be that involving teacher-made paper and pencil tests! Examples of "item types" when using systematic observational procedures include the recording of all instances of helping another when help is sought, helping another when help is not sought, "surprising" another with unexpected gifts, thoughtfulness, etc.

ERIC

The selection of a methodology, and the various "item types" within it, represents the first important consideration in item development. Once this is done it can be combined with the specification of situations to be sampled by the items, e.g., the considerateness of the examinee at home with her mother, her younger brother and her father and her considerateness at school with her "three best friends". Once this is done the actual development of items can get underway. Generally speaking, a measure should contain as many items as is practically possible, for the wider the sample of items within situations the more representative the responses to the measure should be. Ordinarily, it is wise to develop a larger pool of items than ultimately will be needed.

In combination, the selection of an appropriate measurement methodology and the selection of appropriate item types within it provide a basis for assessing the <u>relevance</u> of a measure to that which is to be measured. (It will be recalled that the concept of relevance is comparable to the earlier concepts of face or content validity--see Standards, 1966). As was the case with planning decisions, the single rule that may be thought of as operating in development decisions is that of logic: if a decision as to measurement methodology or item type stretches one's credulity as they pertain to a particular objective then it is likely that the criterion of logic has been violated.

Since paper and pencil tests are widely used by educators, and are appropriate for the measurement of many of education objectives in the cognitive domain the paragraphs which follow illustrate application of the principles outlined above to the development of such tests. In following this discussion it is to be realized that all of the planning and development

decisions that have been reviewed thus far apply to the development of all measures within all measurement methodologies.

As with any other methodology it is wise in developing paper and pencil measures to develop a larger pool of items than ultimately will be needed. Also, it is wise to employ more than one type of item, e.g., completion, true-false, multiple choice, ranking. Items should be phrased so that the content of the statement rather than its form will determine the answer, that is, avoid telltale words such as "always", "never", "entirely", or absolutely", which usually lead to an answer being false, or such words as "may", sometimes", and "as a rule" which are most often associated with true statements. All items of a particular type should be placed together in the test.

A particularly critical issue for the builder of classroom paper and pencil tests is that of item difficulty. Obviously, if an item is so easy that everyone taking the test answers it correctly, then it is of no value in discriminating between those who have more and those who have less of the property being measured. The same problem exists if an item is so hard that no one answers it correctly. Yet for purposes of morale, the reduction of test anxiety, etc., it may be good policy to put in a few items at the beginning of a test that are extremely easy, and for purposes of maximum discriminatory power it may be desirable to put in a few items at the end of the test that few are likely to get right. As yet there are no agreed-upon rules to govern treatment of this dilemma, but a commonly appearing rule of thumb suggests that most items in the test should be of approximately 50 percent difficulty, that is, approximately half of the group being tested should know the answer. This rule evolves from experimental work which shows that under these conditions

a test has maximum discriminative power. In practice, however, test makers generally produce test items having a wide range of difficulty with only the average level of difficulty being approximately 50 (Guion, 1965, p. 200).

In advance of giving a test one can never be sure, of course, how difficult a particular item actually is. A rather complicated procedure labeled "item analysis" is required in order to gain such information. The steps involved in this process are outlined in the next section of the paper.

After an item pool has been established, the two remaining instrument development tasks can be undertaken, namely, the specification of the conditions under which the items are to be administered and the specification of the rules by which mumerals are to be assigned to that which is assessed by an item. In the example where measurement involved the observation of a girl relating to others, specification of the conditions under which the items are to be administered might include such statements as "after dinner alone with her mother in the kitchen" or "at bedtime with her brother upstairs." Directions for taking a paper and pencil test also must be clearly stated. In general, instructions for taking a test should be so clear that the least able student in the class knows what he is expected to do, even though he can't do it.

Rules for assigning numerals to that which is being assessed ("scoring" rules) vary widely with measurement methodology, item type, and weighting decisions. For example, when preconceived category sets are used in systematic observation procedures the scoring rule is usually 1 or 0, that is, a category appeared or it didn't. When rating scales are used the scoring rules may be something to the effect that "when the behavior in question occurs frequently, check scale position 3, when it occurs seldom or not at all check position 1; when it occurs at a frequency somewhere between 3 and 1, check position 2."

Most short answer paper and pencil tests also provide a score of 1 or 0, although some involve a +1 or a -1 with a -1 serving as a penalty for guessing. Formulas do exist that permit correction of "chance" or "guessing" on short-answer paper and pencil measures (Guion, 1965).

Trying out the Measure.[1] Since it is impossible in advance to know how good a measure is, or to know which items are good and which are poor, a tryout of the measure is essential. Two rules need to govern the tryout: 1) every reasonable precaution should be taken to insure the best of measurement conditions, and 2) the time allowance for obtaining the measure should be generous. Since both rules mean somewhat different things for different measurement methodologies, e.g., in systematic observation the setting being observed needs to be "natural", attention is to be directed to the normal course of events, and observer influence needs to be at a minimum and in paper and pencil testing the situation needs to be quiet, attention has to be directed to a specific task and teacher influence (presence) needs to be high, administration and timing rules will not be discussed here. For those interested in administration, timing, etc. in relation to most of the methodologies in the behavioral sciences see Festinger and Katz (1953), Lindzey (1954), Mussen (1960), and Kerlinger (1965). For those interested in the administration of paper and pencil measures in the classroom see Stanley (1964) and Ebel (1965).

---

1 Trying out a measure before it is to be used for evaluation purposes is based upon the assumption that a measure is to be developed that will be used more than once. If this is not the case the test should be read or the measure tried by at least one other person before it is administered formally.

ERIC

A good deal can be learned about a measure from simply attending to the response of examinees to it. For example, in an interview situation one can obtain a fair idea as to which items are clear and which are confusing, or which are probing "sensitive" areas and which are not, simply by observing the respondent's reaction to the items. Similarly with a paper and pencil measure: the overt reactions of students to the test will often suggest whether it is understandable or not, whether it is hard or easy, whether it is relevant to the objectives of the class or irrelevant to them, whether there is too much or too little time, etc. While a measure will not stand or fall upon information as informally obtained as this, it is nevertheless information that is worth obtaining.

Evaluating the Measure. In addition to the informal assessment of a measure, as described above, a formal assessment may be made once the measure has been administered to a number of people that are representative of the target audience specified for the measure. The nature of the evaluation depends, however, upon the use to which the measure is to be put. For example, if a test is to determine mastery to an absolute criterion, and the criterion involves passing all items, then one looks only at the number of items answered correctly; if it is to serve a diagnostic function, then interest is not so much in how many but in which items were passed or failed. If a test is to distinguish between students on that which is being measured, such as academic achievement or a set of motor skills, then one looks at each test item in terms of its ability to do so. Since measures taken in the classroom most commonly serve this last function, attention will be directed in the present paper to it. For reasons of space, the discussion will take as a point of focus only a

ERIC

short-item paper and pencil measure of achievement. The discussion can be generalized, however, to the evaluation of the measurement methodologies.

In evaluating a test for its effectiveness in discriminating between students, the range and distribution of the total test score is only a rough indication of its adequacy. Since a test is made up of items, a total test score is dependent upon the discriminating power of those items; if only part of the items are discriminating then the test score is reflecting the work of only those items. It is possible, through item analysis procedures, to determine which items are and are not discriminating. The value of an item analysis is that it permits one to know which items to eliminate in the next edition of the test. (Obviously, in using this procedure, an item analysis has to be made each time new items are added). By following item analysis procedures one can continuously increase the power of a test to discriminate, as well as increase one's confidence that individual items do in fact discriminate across various class groups.

The simplest procedure for testing the discriminating power of an item is to determine the number of correct responses to the item by the students who rank in the highest 27 percent of the class <u>on the test as a whole</u>, and to compare this with the corresponding number in the lowest 27 percent of the class. The items for which the number of correct responses of the high group <u>most</u> exceeds that of the low group are most discriminating; those in which the number of correct responses of the high group falls behind that of the low group, and those in which the numbers are the same, are not discriminating. These are the items that should be rewritten or discarded. The basic assumption underlying this procedure is that the initial total test score is a relatively

accurate measure of the behavior under consideration for all items subsequently kept and added are based upon that score. To many, this is a difficult assumption to accept.

Steps in the Development of High Inference Measures

As indicated previously, the characteristics of and procedures involved in high and low inference measures are remarkably similar. So too are the steps involved in their development. In fact, the development of high inference measures parallels completely the development of low inference measures; difference between the two appears only in the added requirement of evidence of construct validity for high inference measures. Operationally, this means that there must be some empirical evidence that the measure is in fact measuring that which it is supposed to be measuring.

On the surface, such a difference may seem relatively inconsequential. In reality it is major, for obtaining evidence as to construct validity requires a full scale research study. Two kinds of empirical evidence are admissible in support of the construct validity of a measure: 1) correlation with another measure that is known to be a valid measure of the property under consideration (concurrent validity), and 2) the experimental verification of hypotheses which involve the concept being measured and which have used the test being considered in the research that has provided that verification (theoretical validity). Either or both kinds of evidence provides confidence that the measure "is in fact measuring that which it is supposed to be measuring." Since obtaining evidence of this kind requires a great deal of energy and a fairly high level of research skill, most classroom teachers will be unable to manage it. Without it, high inference measures are suspect, and

should not be used. Practically, this means that most high inference measures available to the educator have to be developed by measurement specialists.

While this may sound somewhat constraining (and it is) it in no way means that such measures are unavailable to the educator. Indeed, there are hundreds of such measures, and they are relatively easy to obtain. O. K. Buros' *Mental Measurements Yearbooks* (cf. 1959, 1965) is of great value in this respect, for he provides in a single volume a comprehensive listing of all published tests in nearly all fields of education, along with excellent reviews of the specific strengths and weaknesses of each. Some form of reliability and validity data are usually included in these reviews, although it is not always easily translatable. Along with the *Yearbooks* Buros has published a 479 page index to *Tests in Print* (1961). With this reference source it is possible to locate rather easily the yearbook in which a particular test has been reviewed. To provide some notion of the range of subject areas covered by Buros the classification of tests included in the 1959 *Yearbook* is included in Appendix B.

5.8

ERIC

References

Ahmann, J. S. and Grock, M. D. Evaluating pupil growth Allyn Bacon, 1963.

Bellack, A. A., in collaboration with Hyman, R. T., Smith, Frank L., Jr., and Kliebard, H. M. The language of the classroom: meanings communicated in high school teaching. Part II. U.S. Office of Education, Cooperative Research Project No. 2023. New York: Institute of Psychological Research, Teacher's College, Columbia University, 1965.

Buros, O. K. (Ed.) The fifth mental measurements yearbook. Highland Park, N. J.: Gryphon Press, 1959, 1965.

Buros, O. K. Tests in print. Highland Park, N. J.: Gryphon Press, 1961.

Campbell, D. T. and Fiske, D. W. Convergent and discriminate validation by the multitrait-multimethod matrix. Psychology bulletin. 1959, pp. 56, 71-105.

Ebel, Robert L. Measuring educational achievement. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Festinger, L. and Katz, D. Research methods in the behavioral sciences. N. Y.: Dryden, 1953.

Gage, N. L. (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.

Gagne, R. The conditions of learning. N. Y.: Holt, Rinehard & Winston, 1964.

Gronlund, N. E. Measurement and evaluation in teaching. N. Y.: MacMillan, 1965.

Guion, R. M. Personnel testing. N. Y.: McGraw Hill, 1965.

Kerlinger, F. N. Foundations of behavioral research. N. Y.: Holt, Rinehard & Winston, 1965.

Lindquist, E. F. (Ed.) Educational measurement. Washington, D. C.: American Council on Education, 1950.

Lindzey, G. (Ed.) Handbook of social psychology, Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

Mussen, P. H. (Ed.) Handbook of research methods in child development: N. Y.: Wiley, 1960.

Sax, G. The construction and analysis of educational and psychological tests: a laboratory manual. Madison, Wisconsin: College Printing and Typing Co., 1962.

Schalock, H. D. Issues in the conceptualization and measurement of teaching behavior. Paper presented as part of a symposium on the Prediction of Teaching Behavior, American Education Research Association, New York, February, 1967. (Mimeographed)

Schalock, H. D. Situational Response Testing: An Application of Simulation Principles to Measurement. In Paul A. Twelker (Ed.) Instructional simulation: a development and dissemination activity. U.S. Office of Education, Bureau of Research No. 7-I-045. Monmouth, Oregon: TEACHING RESEARCH, Oregon State System of Higher Education, 1969.

Schalock. H. D. Smith, C., Micek, S. and Vogel, Frances. An Overview of the TEACHING RESEARCH System for the Description of Teaching Behavior in Context. Revision #1, 1969. (Mimeographed)

Standards for educational and psychological tests and manuals. Prepared by a joint committee of the American Psychological Assoc., the American Educational Research Association, and the National Council on Measurement in Education. Washington, D. C.: American Psychological Assoc., Inc., 1966.

Stanley, C. Measurement in today's schools. (4th ed.), Englewood Cliffs, N. J.: Prentic-Hall, 1964.

Taba, Hilda, Levine, S. and Elzey, F. F. Thinking in elementary school children. U.S. Office of Education, Cooperative Research Project No. 1574. San Francisco State College, 1964.

Webb, E. J., Campbell, D. T., Schwartz, R. D. and Seechest, L. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

ERIC Full Text Provided by ERIC

SUGGESTED READINGS
ON
THE NATURE OF MEASUREMENT IN THE BEHAVIORAL SCIENCES

1. For the mathematical foundations of measurement see:

Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.) Handbook of experimental psychology. N. Y.: Wiley, 1951, pp. 1-49.

2. For good introductory references to the concept of measurement see the following references:

Kerlinger, F. N. Foundations of behavioral research. N. Y.: Holt, Rinehard and Winston, 1965, Chs. 1, 3, and 23.

Lorge, I. The fundamental nature of measurement. In E. F. Lindquist (Ed.) Educational measurement. Washington, D. C.: American Council on Education, 1950, Ch. 14.

Peak, Helen. Problems of objectives observation. In L. Festinger and D. Katz (Eds.) Research methods in the behavioral sciences. N. Y.: Dryden, 1953, Ch. 11.

3. For those interested in measurement from the point of view of the philosophy of science, see these:

Campbell, N. R. Foundations of science: the philosophy of theory and experiment. N. Y.: Dover, 1957.

Ellis, B. Basic concepts of measurement. Cambridge, England: Cambridge Univ. Press, 1966.

Margenau, H. The nature of physical reality, N. Y.: McGraw-Hill, 1950.

Torgerson, W. Theory and methods in scaling. N. Y.: Wiley, 1958.

Ross, S. Logical foundations of psychological measurement: a study in the philosophy of science. Copenhagen; Scandinavian Univ. Press, 1964.

APPENDIX A

MULTIPLE CRITERION MEASURES FOR EVALUATION OF EDUCATION OUTCOMES*

I. Indicators of Status or Change in Cognitive and Affective Behaviors of Student in Terms of Standardized Measures and Scales (Obtrusive Measures)

Standardized achievement and ability tests;

Standardized self inventories designed to yield measures of adjustment, appreciations, attitudes, interests, and temperament;

Standardized rating scales and check lists for judging the quality of products in visual arts, crafts, shop activities, penmanship, creative writing, exhibits for competitive events, cooking, typing letter writing, fashion design, and other activities;

Standardized tests of psychomotor skills and physical fitness.

II. Indicators of Status of Change in Cognitive and Affective Behaviors of Students by Informal or Semiformal Teachermade Instruments or Devices (Obtrusive Measures)

Incomplete sentence technique: categorization of types of responses, enumeration of their frequencies, or rating of their psychological appropriateness relative to specific criteria.

Interviews: frequencies and measurable levels of responses to formal and informal questions raised in a face-to-face interrogation.

Peer nominations: frequencies of selection or of assignment to leadership roles for which the sociogram technique may be particularly suitable.

Questionnaire: frequencies of responses to items in an objective format and numbers of responses to categorized dimensions developed from the content analysis of responses to open-ended questions.

---

* From a paper by Metfessel, N.S. and Michael, W. B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Paper read at the 1967 AERA conference in New York. (Mimeographed).

Self-concept perceptions: measures of current status and indices of congruence between real self and ideal self--often determined from use of the semantic differential or Q-sort techniques.

Self-evaluation measures: student's own reports on his perceived or desired level of achievement, on his perceptions of his personal and social adjustment, and on his future academic and vocational plans.

Teacher-devised projective devices such as casting characters in the class play, role playing, and picture interpretation based on an informal scoring model that usually embodies the determination of frequencies of the occurrence of specific behaviors, or ratings of their intensity or quality.

Teacher-made rating scales and check lists for observation of classroom behaviors: performance levels of speech, music, and art; manifestation of creative endeavors, personal and social adjustment, physical well being.

III. Indicators of Status or Change in Student Behaviors Other than Those Measured by Tests, and Observation Scales in Relation to the Task of Evaluating Objectives of School Programs (Nonobtrusive Measures)

Absences: full-day, half-day, part day and other selective indices pertaining to frequency and duration of lack of attendance.

Anecdotal records: critical incidents noted including frequencies of behaviors judged to be highly undesirable or highly deserving of commendation.

Appointments: frequencies with which they are kept or broken.

Articles and stories: numbers and types published in school newspapers, magazines, journals, or proceedings of student organizations.

Assignments: numbers and types completed with some sort of quality rating or mark attached.

Attendance: frequency and duration when attendance is required or considered optional (as in club meetings, special events, or off-campus activities).

Autobiographical data: behaviors reported that could be classified and subsequently assigned judgmental values concerning their appropriateness relative to specific objectives concerned with human development.

Awards, citations, honors, and related indicators of distinctive or creative performance: frequency of occurrence or judgments of merit in terms of scaled values.

ERIC

Books: numbers checked out of library, numbers renewed, numbers reported read when reading is required or when voluntary.

Case histories: critical incidents and other passages reflecting quantifiable categories of behavior.

Changes in program or in teacher as requested by student: frequency of occurrence.

Choices expressed or carried out: vocational, avocational, and educational (especially in relation to their judged appropriateness to know physical, intellectual, emotional, social aesthetic, interest, and other factors).

Citations: commendatory in both formal and informal media of communication such as in the newspaper, television, school assembly, classroom, bulletin board, or elsewhere (see Awards).

"Contacts": frequency or duration of direct or indirect communications between persons observed and others.

Disciplinary actions taken: frequency and type.

Dropouts: numbers of students leaving school before completion of program of studies.

Elected positions: numbers and types held in class, student body, or out-of- school social groups.

Extracurricular activities: frequency or duration of participation in observable behaviors amenable to classification such as taking part in athletic events, charity drives, cultural activities, and numerous service-related avocational endeavors.

Grade placement: the success or lack of success in being promoted or retained; number of times accelerated or skipped.

Grade point average: including numbers of recommended units of course work in academic as well as in non-college preparatory programs.

Grouping: frequency and/or duration of moves from one instructional group to another within a given class grade.

Homework assignments: punctuality of completion, quantifiable judgments of quality such as class marks.

Leisure activities: numbers and types of; times spent in; awards and prizes received in participation.

Library card: possessed or not possessed; renewed or not renewed.

Load: numbers of units or courses carried by students.

Peer group participation: frequency and duration of activity in what are judged to be socially acceptable and socially undesirable behaviors.

Performance: awards, citations received; extra credit assignments and associated points earned; numbers of books or other learning materials taken out of the library; products exhibited at competitive events.

Recommendations: numbers of and judged levels of favorableness.

Recidivism by students: incidents (presence or absence or frequency or occurrence) of a given student's returning to a probationary status, to a detention facility, or to observable behavior patterns judged to be socially undesirable (intoxicated state, dope addiction, hostile acts including arrests, sexual deviation).

Referrals: by teacher to counselor, psychologists, or administrator for disciplinary action, for special aid in overcoming learning difficulties, for behavior disorders, for health defects or for part-time employment activities.

Referrals: by student himself (presence, absence, or frequency).

Service points: numbers earned.

Skills: demonstration of new or increased competencies such as those found in physical education, crafts, homemaking, and the arts that are not measured in a highly valid fashion by available tests and scales.

Tardiness: frequency of.

Transiency: incidents of.

ERIC

APPENDIX B

CLASSIFICATION OF TESTS IN THE FIFTH MENTAL MEASUREMENTS YEARBOOK (1959)

Section VI: Experimental Design

Editorial Foreword

The chapter of experimental design authored by Cathy Kielsmeier and the editor was constructed for the unstatistical or anti-statistical reader. It presents eleven major experimental designs in terms of information that can be gleaned from such designs, under ideal conditions, and information that cannot be obtained from these designs.

The yield of each design is based on a six-category taxonomy of information. The taxonomy, "The Peapicker System," is presented first and then each of the eleven designs is analyzed in terms of this six-category system. The chapter is limited to the bare bones of design. The alleged yield of each design is based upon assumptions that accompanying measures, experimental methodology, etc., are relatively ideal.

As indicated in portions of the text, the chapter itself is now part of a more embracing instructional system which includes a number of exercises and simulation activities designed to achieve skills that probably cannot be attained by reading alone.

After presenting the rationale for the six-category taxonomy, each design is presented in terms of:

a. its structure

b. its information yield and non-yield

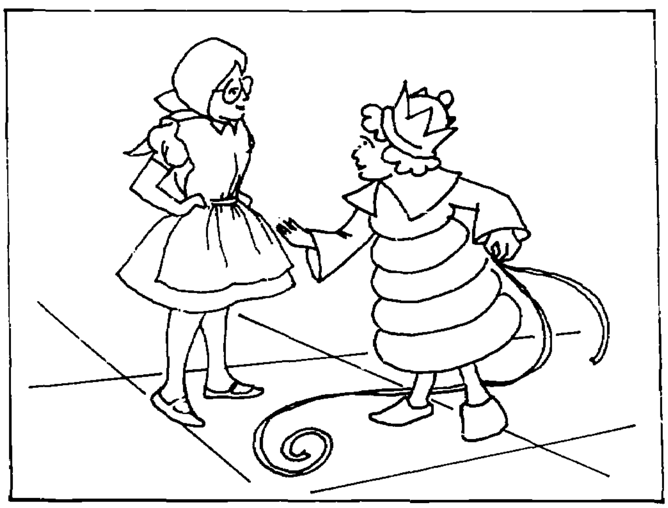
c. a concrete example of the design with the yield identified.

A simplified, and highly artificial, way of computing the relative costs of each design is also presented. While the estimators used are only moderately realistic, they offer the reader a basis to compare the relative usefulness of each design.

VI

The chapter owes much of its original impetus to the Campbell and Stanley approach as presented in Chapter 5, Research on Teaching, N. L. Gage, editor. However, it represents several distinct breaks with that tradition.

This chapter presents a number of plans for constructiong experiments to yield certain kinds of information. The plans presuppose that some sort of measurement will be used to gather in information and that some kind of data analysis will be used to interpret it; but does not delve into either of these topics. The plans, or experimental designs can be used in either instructional development or more general research problems. For example, one of the simpler designs, i.e., a simple pre-test of subjects and a post-test after the instruction is used in most systematic attempts to improve instruction.



"A pawn goes two squares in its first move, you know. So you'll go very quickly through the Third Square...the Seventh Square is all forest--however, one of the Knights will show you the way--and in the Eighth Square we shall be Queens together..."

The plans represent some relatively straightforward ways to gain the maximum amount of desired information. Each plan has advantages and disadvantages in terms of information yield and cost. The reader hoping to find some really "good" or "bad" designs will be disappointed. In the author's view, a design's appropriateness depends upon the information needed and the resources available.

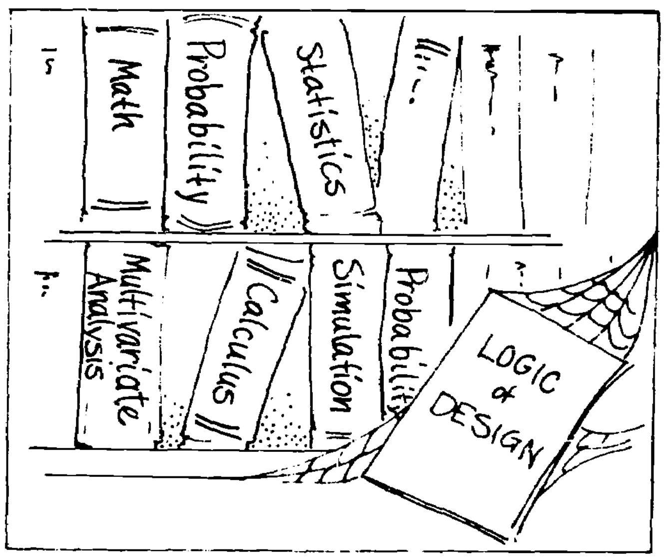
EXPERIMENTAL DESIGN

Introduction

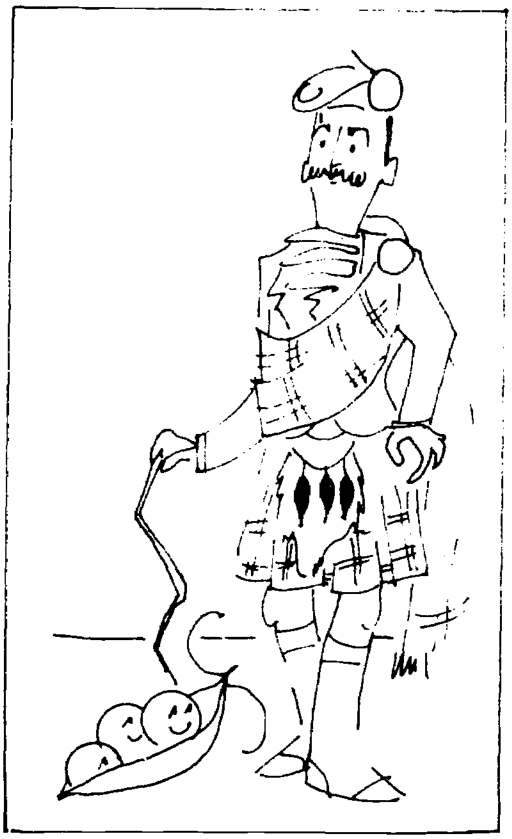
This outline has been constructed to assist students with "Experimental Design." Instruction in design is typically written from a watery mathematical perspective. While a share of such instruction is couched in English, the reader is repeatedly punctured with allusions of the superiority of such mathematical techniques.

Standard instruction in experimental design attempts to combine instruction in computation, analysis of data, the statistical assumptions on which such analysis rests, with learning to build and use appropriate experimental designs.

By contrast, the present approach is couched entirely in the English language. It emphasizes the basic logic of design in terms of what a design is capable of yielding in information, and what it cannot yield. And, the approach places an emphasis upon estimating the relative cost of a design. To achieve this focus, we have sacrificed excursions into computational procedures, formulae and statistical inferences.



This approach is an attempt to build upon existing language strength in the learner's own culture; that is, his facility with his native language,



English. The remarkably robust mother tongue has successfully encompassed more than one way of thinking. To what degree it can fairly contour the ideas and relationships of Experimental Design remains open.

We have grouped the major classes of information potentially yielded by experimental design into six categories. The first letters of these six categories of information form the word P-PICER. As a mnemonic device, we might call this <u>pea picker</u>, reminiscent of Tennessee Ford's the old "pea picker" greeting. The Pea-picker classification of information yield, our concern with relative costs, and the historical fact that experimental designs owe their development to the science of growing plants more than any other activity has led us to our Scotish Pea-Picker analogy.

The type of design that concerns us here refers to a plan for collecting information from an experiment. The experimental aspect refers to a planned interference in the natural stream of events by

the researcher. He does something more than carefully and passively observe what is occurring anyway. The experimenter is an intruder. In the experimental designs we will discuss, some active manipulation occurs, i.e., some students are administered treatment, etc.

This emphasis upon an experiment reflects the higher regard generally given to information so derived. There is a good rationale for this. Much of the substantial gain in knowledge in all sciences has come from actively manipulating or interfering with the stream of events. There is more than just observation of a natural event. A selected condition or a change (treatment) is introduced. Observations are planned to illuminate the effect of any change in conditions. However, if one were patient enough (infinitely patient) all experiments might occur naturally. The basis of inquiry logic is really no different between a "natural" and a contrived experiment. Thus, extrapolation from experimental design, similar planning strategies, i.e., designs, can be used across a range of non-experimental problems and projects.

The importance of experimental design also stems from the quest for inference about causes or relationships as opposed to description. Researchers are rarely satisfied to simply describe the events they observe. They want to make inferences about what produced, contributed to, or caused events. To gain such information without ambiguity, some form of experimental design is ordinarily required.

As a consequence, the need for using rather elaborate designs ensues from the possibility of alternative relationships, consequences or causes. For example, Treatment A may have caused observed Consequence O, but plausibly the consequence may have derived from Event E instead of the treatment or from Event E combined with the treatment. It is this pursuit of more

ERIC

refined and unambiguous relationships that leads to the need for carefully planned designs.

The kinds of planned manipulation and observation called experimental design often seems to become a bit complicated. This is unfortunate but necessary, if we wish to pursue the potentially available information. The kind of information we are usually pursuing in the behavioral sciences and education is often masked by noise. Our manipulations or treatments rarely have such clear-cut effects on subsequent behavior of the subject as, for example, the French guillotine. No complex designs have even been thought necessary to explore the post-treatment behavior of guillotined subjects.

The effect of one potential causal agent on your behavior while reading this paragraph may well be masked by the effects of hundreds of others. A carefully planned design could tease out the relative effect of some of these causal agents.

This plan which we call a design is an essential part of research strategies but not all. The design itself entails:

1) Selecting or assigning subjects to groups or experimental units,
2) selecting or assigning units for specific treatments or conditions of the experiment (experimental manipulation),
3) specifying the order or arrangement of the treatment or treatments,
4) specifying the sequence of observations to be taken.

By convention, the problems of design do not ordinarily include details of sampling, selection of measurement instruments, selection of the research problem, the nuts and bolts of procedure required to actually do the study, etc.

Reader Prerequisites

This chapter is intended for the learner who wishes assistance in selecting or constructing an experimental design. The chapter is not intended to be profound nor extensive in the field of design. No mathematical, statistical, scientific prerequisites are required. The English language and logical skills typically attained by the eighth grade are sufficient.

We will introduce the reader to the design approach based on classes of information which a given design may or may not yield. We will present some simple, and arbitrary, means of calculating the relative cost of a given design. Then we will examine each of eleven generally used, and useful, designs in terms of their information yield and cost.

The eschewing of statistics in this chapter should not be interpreted as the author's personal distaste. We enjoy an occasional sampling distribution as well as anyone.

Objectives

The following objectives refer to the entire package, of which this chapter is a portion. Those marked * are probably appropriate for the reader.

A. Cognitive

*1. The participants will be able to correctly identify, i.e., name each of eleven standard experimental design forms from examples. Presented with a random ordered set of sixteen or more examples of each of these designs, they will be able to name correctly each design with 100 percent accuracy.

2. Participants will be able to enumerate the major research design questions each of the above designs will answer and those major questions it will not answer. Having first studied a prepared list of such questions accompanied by the strengths and inadequacies of each design the participants will be able to recall such answers and non-answers when given either the name or an example of each of the eleven major designs. Participants will be able to recall such strengths and weaknesses with 90 percent accuracy.

3. Given a typical instructional design problem, participants will be able to construct, in a 20 minute period, a useful design. They will be able to enumerate the main questions the design will answer, those questions it leaves unanswered, and the relative cost of the design.

4. Given an example of a typical design <u>inadequate</u> for the problem for which it was designed, participants will be able to construct a superior alternative and enumerate the additional information yield and the relative cost of each design.

5. Given any two standard approaches to the same instructional problem, that is, two different designs, the participant will be able to describe the advantages and disadvantages of each alternative.

*6. Participants will be able to recall four threats to internal validity. After recalling each threat, he will be able to list one or more design variations which cope with that threat.

7. Participants will be able to match abstract definitions and examples with the names of 50 important concepts and principles in experimental design.

8. Participants will demonstrate verbal fluency in appropriate use of the above concepts and principles.

9. Given a problem which demands the construction of a <u>hybrid design</u> not ordinarily found in a basic design text, but one that requires a combination of two or more standard designs, the participant will create such a solution.

10. Participants will be able to read typical experimental research reports and:

    (a) Analyze the design
        (1) name it
        (2) enumerate questions answered and not answered

    (b) Construct an alternative design which will answer more questions either:
        (1) with increased cost
        (2) with no substantial increase in cost.

B. <u>Affective</u>

1. The participant will, using typical attitude measures, rate the area of experimental design in a highly favorable light.

2. Participants will display not only favorable attitudes, but highly confident behavior and initiative in attacking design problems.

3. Participants will be able to accept a relatively simple design that answers the major questions required by a problem within the time/cost constraints necessary; and, will not feel guilty or professionally inferior because this design will not cope with other questions.

Considerations in Design Selection

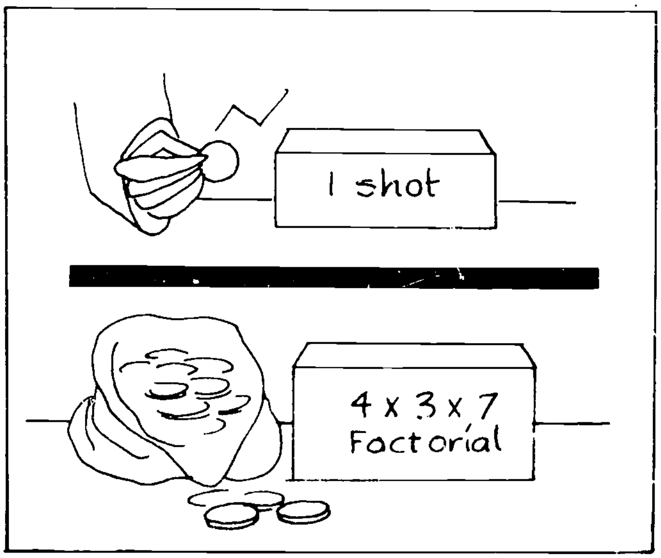
The selection of a specific type of design depends primarily on both the nature and the extent of the information we wish to obtain. Elaborate designs, usually involving a number of "control groups," offer more information than a simple group design. If "more information per project" were the sole criterion for selection of a design, we would be led to more and more complex designs. However, not all of the relevant information which may be needed, even in the most circumscribed area, can be derived from any given design. Part of the information will be piggy-backed into the study by assumptions, some of which are explicit. Other information derives from a network of knowledge surrounding the project in question. Theories, accepted concepts, hypotheses, principles and empirical evidence from related studies contribute. To the extent that this knowledge is already available, the task of extracting the exact information needed to solve any given research problem is circumscribed.

Furthermore, collecting information is costly. The money and staff resources available have some limits. Subjects are usually found in finite quantities only. Time is a major constraint, to mortals. The information to be gained has to be weighed against some estimate of the cost of collection. This rationale points up two ways of checking potential designs:

1. What questions will the design answer?

   To do this, I assume we must also be able to specify many of the questions the design won't answer, as well as ones it will answer.

ERIC



This should lead, I hope, to a more realistic approach to designs than is usually given. Some innocent and useful designs have been labeled "poor," and by implication, downright nasty, because they are relatively simple and will not answer some questions. Yet they may provide clear and economical answers to the major questions of interest. Cumbersome designs are not as useful for some purposes.

2. What is the relative information gain/cost picture?

We are not suggesting the use of any specific formula or strategy for deriving cut-off points. Just take a close look at the probable cost before selecting a design. In the folloiwng exercises we have used some arbitrary cost figures, just to get the reader sensitized to this dimension.

The Information Desired (Questions to be Answered)

Typically, in education and the behavioral sciences, the information desired is based upon some variant of the question: "What do people do under these conditions?" Accurately observing and recording indices of the behavior in question is the problem of measurement. Design becomes increasingly important as questions begin to emphasize determinants of behavior, and which factors are not appreciably related. If further questions are asked, e.g., to determine which of several antecedents of the behavior

experimental design, the group or groups. We might symbolize this as follows, a selected group, a treatment, an observation.

$$\boxed{GP} \quad \boxed{T} \quad \boxed{O}$$

If we assume this represented a study, and the horizontal dimension from left to right represented time, we could read the above diagram as selecting a group, applying a given treatment, and observing some behavior of the group after the treatment. This symbolizes a simple research design often called the "one-shot."

The group in an experiment which receives the specified treatment is called the <u>TREATMENT GROUP</u> or the experimental group. However, the term <u>CONTROL GROUP</u> refers to another group assigned to the project but <u>not</u> for the purpose of being exposed to the treatment. Thus, the performance of the control group usually serves as a baseline against which to measure the effect of the full treatment on the treatment group.

There are a few more terms that researchers continually use that you should also be able to quote with precise abandon. For example, a <u>VARIABLE</u>.

A variable refers to almost anything under the sun. There are only two kinds of stuff in the world for researchers: variables and constants. As a result, almost any concept, or thing, or event they are interested in, <u>that varies</u> or <u>can be made to vary</u>, and that is related to their research can be called a variable. Researchers pay particular attention to variables which they are manipulating in the study as well as unwanted variables that may influence the results, (much to the concern of the researcher). If the treatment itself represents a cohesive set of ideas such as "reinforcement" or "feedback," it is often referred to as a variable.

EXTRANEOUS VARIABLES (external to the experiment), are such things as the maturation of the subject, social calamities, marked changes in the community, etc., that may influence the subject's behavior in addition to the treatment.

A variable of specific experimental interest is sometimes referred to as a FACTOR, especially if it is assumed to underlie a complex form of observed behavior. Ordinarily, the term is used in one of two contexts. One occurs when an experiment involves more than one variable. These variables are often identified as factors and are labeled "Factor A," and "Factor B," etc. For example, a study might be designed to investigate the effects of (1) small group discussions and (2) some form of individualized instruction. Small group discussion might then be referred to as one factor, or Factor A, while individualized instruction would be Factor B.

The other main use of the term, factor, refers to a statistical technique called "factor analysis," by which variables which seem to have similar effects are grouped together into a category and referred to as a factor. This essentially creates a new construct, "the factor," which then may be presumed to underlie the original individual variables.

LEVEL refers to the degree or intensity of a factor. Any factor may be presented in one or more of several levels, including a zero level.

RANDOMNESS refers to the property of completely chance events that are not predictable (except in the sense that they are random). If they are truly random, examining past instances of occurrence should give you no clues as to the future occurrences. Thus, if we were to predict outcome from perfect pairs of dice rolled in an unbiased way, (which are random events), previous rolls give no clues.

ERIC

Randomness becomes important in the design of the experiments primarily in the assignment of subjects to groups. Researchers feel more secure about the results of their studies if subjects have been randomly assigned to groups. RANDOM ASSIGNMENT tends to spread out differences between subjects in unsystematic (random) ways so that there is no tendency to give an edge to any group.

Randomization, or random assignment, refers to a technique of assignment or ordering such that no consistent or systematic effect in the assignment is tied in with the method. Elimination of such systematic influence upon assignment or selection allows for chance assignment. Approved ways of generating chance assignments involve tables of random digits numbers. However, typically, researchers frequently resort to simple counting off, flipping a coin, and other short cuts.

Another major way of selecting subjects is simply to use intact groups: such as all the students in a given classroom. Researchers are usually worried whether the students were assigned to one classroom or another in a non-random way or whether some subtle factors were operating to exert a bias of selection factors in the assignment to the groups.

EX POST FACTO refers to causal inferences drawn "after the fact," for in the ex post facto study, the causal event of interest has already happened. The investigator then examines the event and tries to ascertain the causes of it. These are known as non-experimental studies and are often contrastual with experimental studies. A typical example is giving people a questionnaire about events that have happened sometime in their past life, such as interviewing divorced people, and from their statements making inferences about what caused the marital breakdown.

ERIC

VARIANCE refers to the variability in any event. If you use a fine enough measuring device, you can find differences between any two objects or events or even between a given object from one point in time to another. As more and more precise instruments are being used in investigating human behavior, investigators become confronted by seas of variance. The problem then often becomes an analysis of this variance into different components of sources. Formal attacks upon this occupy a large part of what is referred to as the statistics used to analyze experiments.

The inside logic of an experiment is referred to as INTERNAL VALIDITY. Primarily, it asks the question: Does it seem reasonable to assume that the treatment has really produced the measured effect? Extraneous variables which might have produced the effect with or without the treatment are often called "threats to validity."

EXTERNAL VALIDITY, on the other hand, refers to the proposed interpretation of the results of the study. It asks the question: With what other groups could we reasonably expect to get the same results if we used the same treatment? If Treatment X resulted in increased school attendance for first graders in your study, could you logically say it would result in increased attendance with high school students?

BLOCKS usually refers to categories of subjects within a treatment group. For example, we might divide the group into high and low anxious students based on the Trait-State Anxiety Inventory: In a two-group study this could be done for both groups. The advantage is to enable us to discover how the treatment affects each of the anxiety blocks. For example, it might depress performance

ERIC

of the high anxious and increase performance of the low anxious. This would be an interaction between treatment and subject characteristic.

Treatment Gp.	Blk	(Hi Anx)
	Blk	(Lo Anx)
Control Gp.	Blk	(Hi Anx)
	Blk	(Lo Anx)

INTERACTION refers to variables in the treatment which may interact with each other. It may make a difference whether a variable is used by itself, with another, or with different levels or degrees of another. For example: (1) Increasing the reward for problem solving may increase performance if subjects are under low stress; or (2) it may decrease performance if subjects are under high stress.

In this instance, the effect of reward may depend on whether or not stress accompanied the reward. Thus, the effect of stress would depend on whether or not reward is present; the effect of reward would depend on whether or not stress is present. Then, these two factors would interact in their influence upon the observed behavior. This is a two-factor or first-order interaction.

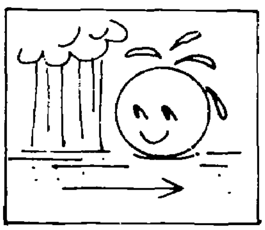
Higher order interactions are possible. One factor may depend on the presence or absence of two other factors; a second-order interaction.

THE HAWTHORNE EFFECT: This refers to the behavior of interest being caused by subjects being in the center of the experimental stage, i.e., having a great deal of attention focused on them, being in the midst of some novel,

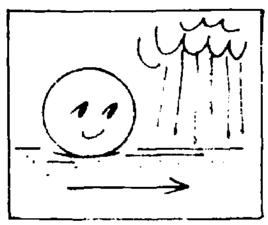
and presumably helpful experience, etc. The effect usually manifests itself as a spurt or elevation in performance, unconnected with the special treatment used, but resulting from the subjects being singled out as a "special" treatment group. This is a threat to the internal validity of the experiment, and often an important factor to control.

Information to be Considered by the Designer of Projects

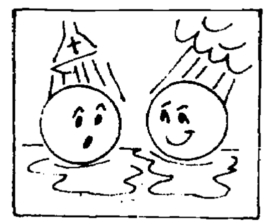
There are six major classes of information with which a designer must cope. These relate to:



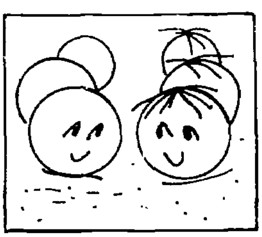
Post-treatment behavior (P_1)



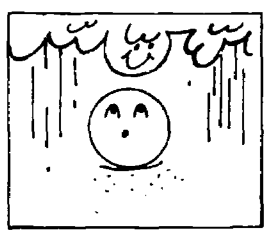
Pre-treatment behavior (P_2)



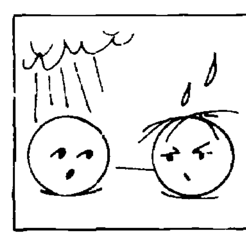
Internal threat to validity (I)



Comparable groups (C)



Experimental errors (E)



Relationship of the treatment (R)

Within the six major classes of information, P_1-P_2ICER, the following more specific issues may be identified: (see Figure 1, page 17, for complete list).

Figure 1. Experimental Design Summary Sheet
The "Pea Picker" System of Design Yield

P_1-P_2ICER:

P_1 - Post-treatment Behavior (results)

P_2 - Pre-treatment Behavior (pre-test data)

I - Internal Threat to Validity

C - Comparable Groups

E - Experimental Errors

R - Relationships of the Treatment

TYPES OF DATA YIELD

(P_1)

1. P_1-1...behavior immediately or shortly after treatment
2. P_1-2...a comparison of post-treatment behavior between experimental and control groups
3. P_1-3...a comparison of post-treatment behavior between experimental groups or blocks
4. P_1-4...long-term effects with continuing treatment and periodic observations
5. P_1-5...long-term effects without continuing treatment but with observation(s)

(P_2)

6. P_2-1...behavior immediately or shortly before treatment
7. P_2-2...comparing pre-treatment to post-treatment behavior
8. P_2-3...a comparison of pre-treatment behavior between different groups of subjects
9. P_2-4...a comparison of the differences between pre-treatment and post-treatment behavior among groups of subjects
10. P_2-5...the effect of the pre-treatment observation on subsequent behavior of the subject

Figure 1. (continued)

(I)

11. I-1...the subjects exhibited behavior because of some event <u>other than the treatment</u>, e.g.;

    a. learning from some other experience
    b. maturation or development
    c. a change in attitude, emotion, etc.
    d. a physical change

12. I-2...the subjects could, or would perform the behavior without the treatment

(C)

13. C-1...were the groups comparable before the treatment

    a. did they differ in some essential past experiences
    b. did they differ in some traits, or intelligence or personality factors

14. C-2...did the groups receive a comparable degree of experiences during the time of the study (except for differences in treatment)

(E)

15. E-1...were the results due to the Hawthorne phenomenon

16. E-2...were the results due to the personal influence of the experimenter or other project staff

17. E-3...did some trivial aspects of the treatment account for the outcome behavior, e.g., assembling the subjects in groups, reading instructions to them, etc.

(R)

18. R-1...did the treatment interact with subject characteristics, so that subjects with different characteristics reacted differently

19. R-2...how does the treatment interact when combined with other sorts of treatment

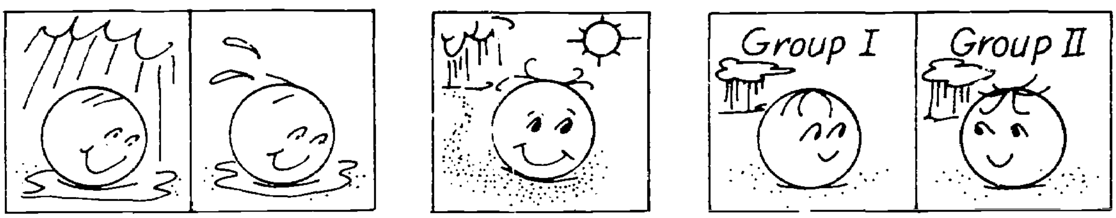
20. R-3...does the treatment contain different factors which may operate differently on the subjects

21. R-4...what are the effects of different levels or degrees of the treatment

22. R-5...what are the effects of different orders or sequences of various treatments

P_1 Post-treatment behavior

In a typical experiment, this is the behavior, the class of information of primary interest. How did the subjects behave after the treatment? All designs shed some light on this class of information, the results of an experiment. Usually only immediate or short-range results are obtained. However, long-term effects (a year or two after the treatment conditions) are frequently not yielded. More complicated kinds of information derive from, and concern questions of comparing post-treatment behavior between groups who have had various kinds, levels, or even absences of treatment. Thus, five categories of post-treatment behavior can be identified:

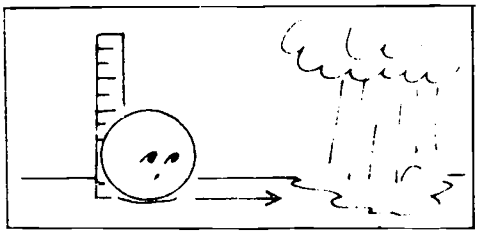


1. Behavior immediately or shortly after treatment
2. A comparison of post-treatment behavior between experimental and control groups
3. A comparison of the post-treatment behavior between experimental groups or blocks
4. Long-term effects with continuing treatment and periodic observations
5. Long-term effects without continuing treatment but with observation(s)

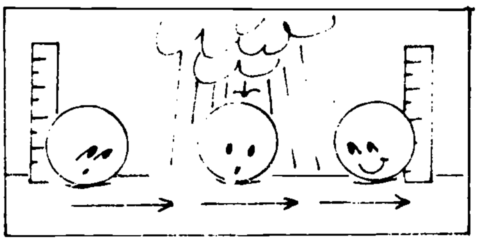
P_2 Pre-treatment behavior

Information concerning pre-treatment behavior requires some observation, a test or measure, to be administered before the experimental manipulation.

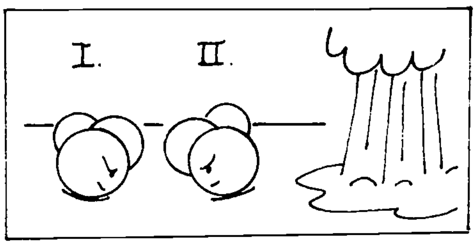
Without such observations, the design itself will not answer any questions about the subjects before the experimental conditions have been introduced. Such information, however, may be accrued from general knowledge and/or other studies. Direct acquisition of this information adds to the cost of an experiment. Furthermore, it may introduce a confounding effect. That is, sometimes the pre-treatment observation influences the subsequent behavior of the subjects. When the study is over, it may not be clear whether the behavior was due to the treatment, the pre-treatment observation or measure, or both. Several classes of pre-treatment information can be acquired:



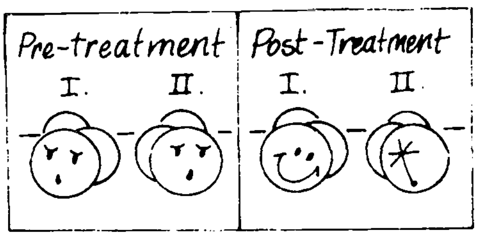
1. Behavior immediately or shortly before treatment



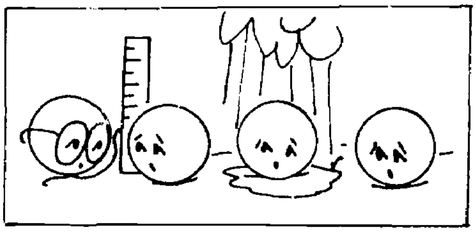
2. Comparing pre-treatment to post-treatment behavior



3. A comparison of pre-treatment behavior between different groups of subjects



4. A comparison of the differences between pre-treatment and post-treatment behavior among groups of subjects

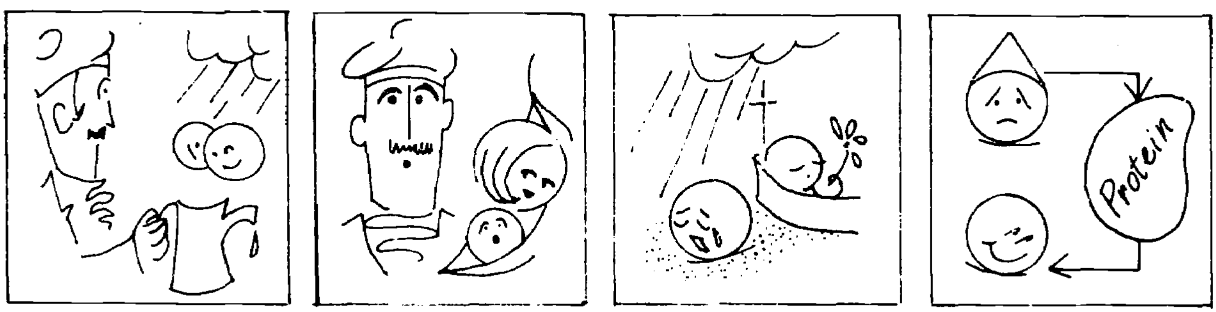


5. The effect of the pre-treatment observation on subsequent behavior of the subject

I <u>Internal threats to the validity of a design</u>

This class of information refers to some rival hypothesis that threatens clear interpretation of the experiment. A common clot of rivals threatens most experiments, particularly those using human subjects. Typically, the rival hypothesis asserts that something <u>outside</u> of the experiment proper produced the behavior of interest. Thus, some observed behavior may have occurred due to informal learning otuside the experiment itself, maturing of the subjects with consequent change in their action, etc. To discover whether or not such rival events exert an influence, the designer must usually provide for one or more control groups. Typically, internal threats include:

1. The subjects exhibited behavior because of some event <u>other than</u> the treatment, e.g.:



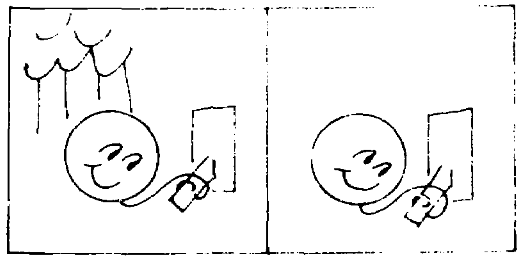
a. learning from some other experience

b. maturation or development

c. a change in attitude, emotion, etc.

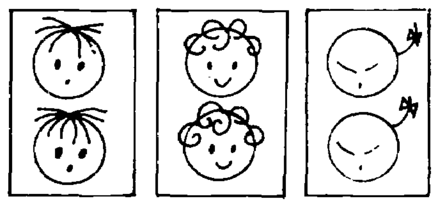
d. a physical change

2. The subjects could, or would, perform the behavior without the treatment.



C Comparable groups

This class of information, available only when two or more experimental units or groups of subjects are used, deals with whether the subjects in the different units were about the same in relevant attributes before the treatment, and during the treatment period, except for the treatment condition itself. If the designer cannot provide information as to the comparability of groups, he must be prepared to admit the possibility that the groups differed in some essential aspect which produced the results observed. Equating the groups by some pre-test or random assignment are the two major techniques of providing this information. Thus, there are two types of comparability information:



1. Were the groups (either experimental or control) comparable before the treatment?

   (a) Did they differ in some essential past experience(s)?

   (b) Did they differ in some traits, or intelligence, or personality factors?

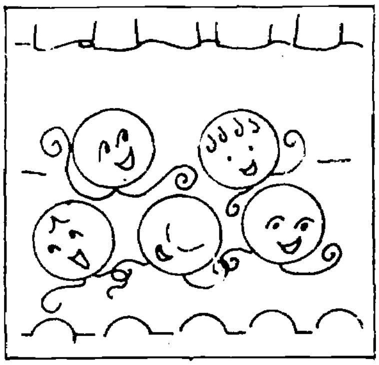
2. Did the groups receive a comparable degree of experiences during the time of the study (except for differences in treatment)?

E Experiment errors

Experiment error refers to some unwanted side effect of the experiment itself which may be producing effects rather than the treatment. The

ERIC

Hawthorne Phenomenon mentioned earlier, i.e., subject reactions when they are obviously given research attention and a presumably preferred treatment, are



a continuing source of experiment error in education research. There are two major ways used to acquire information about the possible presence of the Hawthorne effect: (1) provide for a placebo treatment group which gets the attention, but not essential attributes of the treatment; and (2) continue the treatment over prolonged periods. (This presumes that the Hawthorne effect is short-lived.)

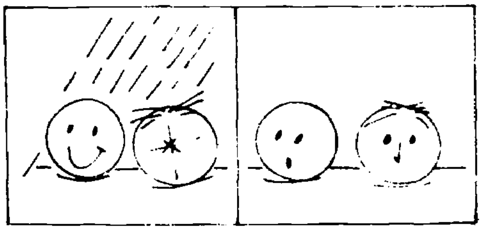
Information on personal or experimenter influence (rather than the treatment) usually requires, for design, providing additional treatment groups or sub-groups, each with a different experimenter or instructor.

1. Was the effect due to the Hawthorne Phenomenon?
2. Was the effect accounted for by the influence of the experimenter personally, or other project staff?
3. Did some trivial aspects of the treatment account for the behavior, e.g., assembling the subjects in groups, reading instructions to them, etc.

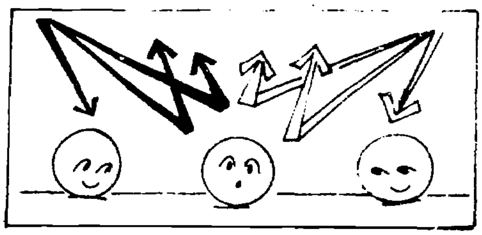
R Relationships of the treatment

This class of information deals with the possible interaction of the treatment effects with: different kinds of subjects, other treatments, different factors within a complicated treatment, different degrees of intensity, repeated applications or continuation of the treatment, and different sequences or orders of the treatment or several treatments.

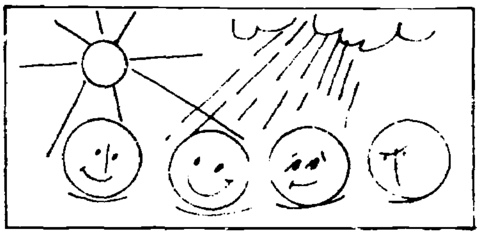
Typically, information of this sort is acquired from blocking, from factorial design, and various repeated measures designs.



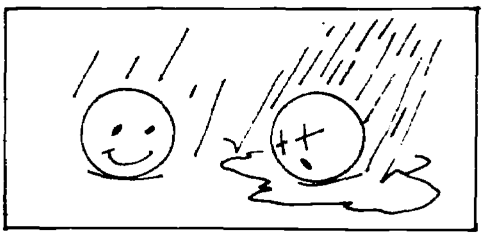
1. Did the treatment interact with subject characteristics, so that people with different characteristics reacted differently?



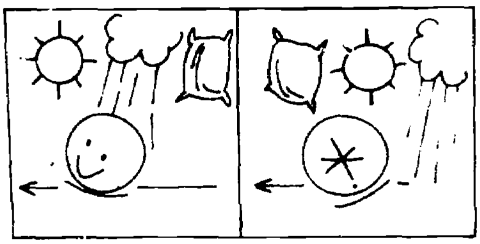
2. How does the treatment interact when combined with other sorts of treatment?



3. Does the treatment contain different factors which may operate differentially on the subjects?



4. What is the effect of different levels or degrees of the treatment?



5. What is the effect of different orders or sequences of various treatments?

These appear to us to be the major classes and sub-classes of information which should be considered by the designer of a study. Other subordinate classes could be considered. Statistical regression, often cited as a major threat to the validity of an experiment, has been subsumed in this classification under the comparability of groups. The statistical regression threat refers to the effect of selecting a group on the basis of an extreme score on a measuring instrument. However, that extreme score was probably based on some chance or unknown factors. Given a repetition of the measure the group would probably score somewhat higher or lower but toward the average in the absence of any treatment. Thus, low scores go up, high scores go down.

To summarize the six classes of questions:

P_1 Post-treatment behavior

P_2 Pre-treatment behavior

I Internal threats to validity

C Comparable groups

E Experiment errors

R Relationships of the treatment

As a mnemonic device, again think of Tennessee Ernie Ford's old term "Pea Pickers." The initials of these six classes of questions are: P_1-P_2 I C E R (pronounced "pea picker"). The eager learner will make some effort to commit these classes of questions to memory. Pea-picker. Shouting it aloud often helps your memory, and that of your colleagues.

DESIGN COST ESTIMATES

Since it will cost in time or money to do almost any part of a study, we have assigned, somewhat arbitrarily, values to every component of a design.

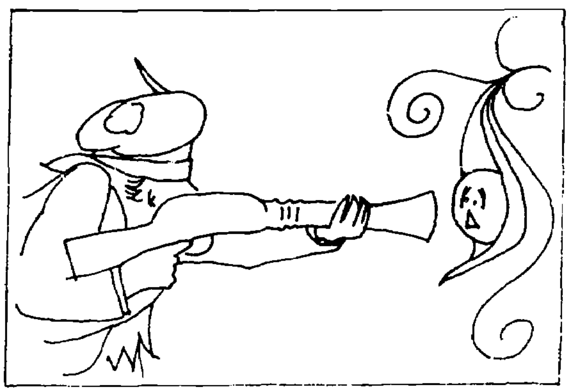
ERIC

Activity	Symbol	Arbitrary Time and Resource Cost
Selection of each group or experiment unit	GP	5 points (each)
Random assignment to a group	R	5 points (each)
Blocking subjects, or other variables, into sets	BK	5 points (each)
Presenting a treatment to a group	T	10 points (each)
Observing	O	5 points (each observation of a group or experimental unit)
Subjects		1/2 point for each subject (thus, for 30 subjects the cost is 15 points)

One of the reasons behind the use of the above estimates is that they are moderately realistic and easily learned in the more active design games within the total package. Now...

In the perspective of the Pea-picker Information Categories and some trial cost estimates, let us examine several standard designs. (See Figure 2, page 47 for a summary chart of all eleven designs.)

The One-Shot Design



A design in which a group of subjects are administered a treatment and then observed. In experimental research, an experimental treatment should be

ERIC

given to subject(s) and observation(s) made. (In ex post facto research, an example of this might be the case study.)

Example:

The Labile-Lingual Self-Teaching textbooks for Esperanto will be used for a newly formed class in Esperanto. Results on the final exams will be used as an assessment of the effectiveness of the texts as a teaching device.

Thus, one group will be given one treatment, and one "observation" will be made (results on the final). This might be diagrammed as follows:

GP	T	O

<u>COST</u>: assuming 20 subjects = 10 + 5 + 10 + 5 = 30 points

Usually, with this design, an intact group of subjects is given the treatment and then observed. No attempt is made to randomly assign subjects to the groups, nor does the design provide for any additional groups as comparisons.

The One-Shot Design is highly useful as an inexpensive measure of a new treatment behavior of the group in question. If there is some question as to whether <u>any</u> expected effects will result from the treatment, for example, a new system of instruction might well be a bomb, then a one-shot may be an economical route.

In cases where other studies, or the cumulative knowledge in the field provide information about either pre-treatment baseline behavior, the effects of other kinds of treatments, etc., the experimenter might sensibly decide that it is not necessary to undertake a more extensive design. Simplicity,

ease, and low cost represent strong potential advantages in the oft-despised one-shot.

This design answers only one question and that is in reference to post-treatment behavior, P_1-1. It will describe the information about the behavior of the subjects after treatment, shortly after treatment.

If another observation is made, sometimes subsequent to the treatment, then the design might also yield information about the longer-term behavior of the subjects after treatment. Thus, at most, the One-Shot Design, even extended, could cope with only two of the twenty questions, these being the first two on the list.

You could summarily dismiss P_1-2 through 5, and the other five classes of questions perhaps by putting in the "not answered" column information of Class P_2, Class I, Class C, Class E, and Class R.

<u>One-Group Pre-Post Design</u>

In this design, one group is given a pre-treatment observation, the experimental treatment, and a post-treatment observation. However, their post-test (after treatment) measures are compared with their pre-test (before treatment) measures. Each test provides the experimenter with an observation, so that in this design, <u>two</u> observations are being compared.

Example:

The Sex Role Tolerance Test, which assesses attitudes toward female

peers, will be given to a group of sixth-grade boys. For the next four years, this group will be exposed to sensitivity training as part of the Social Studies Curriculum. At the beginning of tenth grade, this group will again be tested with the same test, and scores assessed to determine if there has been a shift in a positive direction toward female peers as a result of the special training.

Thus, one group will be "observed" (via test scores) before <u>and</u> after the experimental curriculum. We could diagram this:

$\boxed{GP}$ $\boxed{O}$ $\boxed{T}$ $\boxed{O}$ <u>COST</u>: (20 subjects)

10 + 5 + 5 + 10 + 5 = 35 points

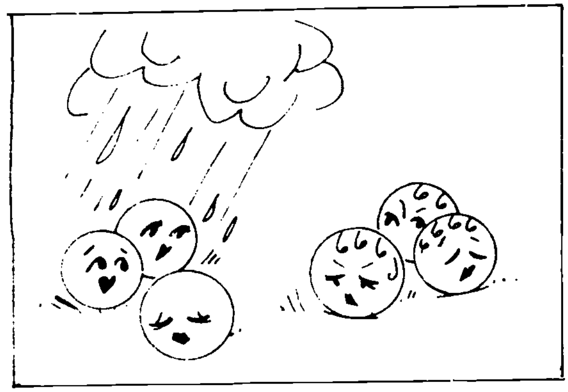
The usefulness of this design is similar to that of the one-shot, except that an additional class of information is provided, i.e., pre-treatment behavior. This design is typical of that used in instructional systems development. If learning gains on measures of increase in competencies following treatment are a major concern, the design economically provides information.

This design will answer the same question as the one-shot design (P_1-1) so that not only the post-treatment behavior of the subjects is answered, but it will also answer some questions in the pre-treatment behavior. If you determine the behavior shortly before treatment, you can compare this to the post-treatment behavior. It seems to answer three of the twenty questions: P_1-1, P_2-1, and P_2-2.

You could summarize information classes I, C, E, and R as not answered. It would not answer questions P_1-2 through 5, or P_2, pre-test behavior, which relate to comparisons of groups at the start, the pre-post difference between groups, or the effect of the pre-treatment observation.

ERIC

Static Group Design



In this design, two intact groups are used, but only one of them is given the experimental treatment. At the end of the treatment, both groups are observed to see if there is a difference between them as a result of the treatment.

Example:

The entire senior class from two small high schools will be used for the study, Nueve Riche Preparatory, and Peoria Podunk High. The Magnificent-Media-Math curriculum will be given to the N.R. Prep students for two semesters. An achievement test will be given to both groups of seniors in May and the results compared in order to assess the effective-ness of M-M-M. This could be diagrammed:

$\boxed{GP}$	$\boxed{T}$	$\boxed{O}$	COST: (20 subjects, 10 per group)
$\boxed{GP}$		$\boxed{O}$	10 + 10 + 10 + 5 + 5 = 40 points

This design may provide information on some rival hypotheses. Whether it does or not, depends on the initial comparability of the two groups and whether their experience during the experiment, differs in relevant ways only by the treatment itself.

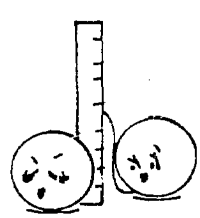
Whether the groups were comparable or not is crucial in determining the extent of information yielded by this design. The design is typically used to compare two or more kinds of instruction. Intact classrooms are often used as the unit or group when random assignment seems practically impossible.

If the designer cannot, on the basis of information outside the experiment itself, assume the comparability of the groups, then the design yields only information regarding P_1-1 and P_1-2.

Under the "not answered" column, it will not grapple with the questions P_1-3 through 5, nor will it answer questions dealing with pre-treatment behavior (P_2 questions). It will not handle Class C questions of comparable groups, Class E questions of experiment errors, nor questions of Class R concerning relationships of the treatment.

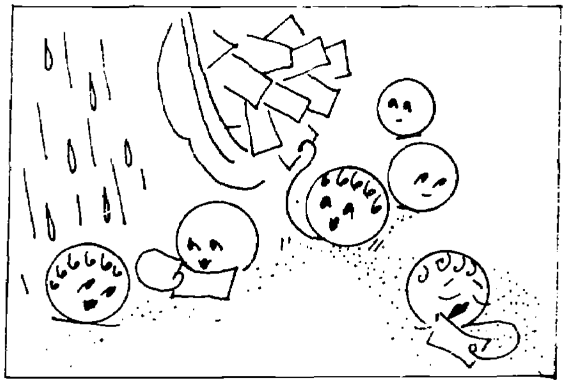
It appears to answer some Class I questions because of the possible control proved by the non-treatment groups. It appears that you might be able to control for possible general learning outside of the treatment, or maturation, etc. However, since intact groups were used without random assignment of subjects to groups, matching pre-test, etc., this is probably an inadequate answer to Class I questions.

Spurious claims are often made for the static group design based on the unwarranted assumption that the groups were really comparable. Such unwarranted claims elevate the supposed information yield to that of a randomized group design. Again, on the basis of the <u>design alone</u>, such interpretation is untenable. The design itself will not yield this much information. However, the design <u>plus</u> additional information can achieve such a yield.



The Random Group Design

This design is similar to the Static Group Design except that an attempt is made to insure similarity of the groups before treatment begins. Since it is difficult to have exactly similar people in each of two groups (unless you separate identical twins), the design works toward a guarantee of an equal amount of strengths and weaknesses in each group by assigning people to groups at random. If you do this you are likely to have similar numbers of fast and slow readers in each group, high and low socioeconomic status in each group, similar numbers of boys and girls in each group, etc., (in the long run).



Example:

The subjects for this study will be all students enrolled in 7th period PE at West Junior High. This will include 40 boys from boys' PE and 30 girls from girls' PE. They will be randomly assigned either to the control or the experimental group by mixing all their names in a hat and pulling them out at random, assigning the first one drawn to Group 1 (experimental), the second one drawn to Group 2 (control), the third to Group 1, etc.

The experimental group will spend their entire PE period studying the new text, "Scientific Principles of Succeeding in Basketball." The control group will receive their traditional PE instruction in basketball. At the end of four weeks, both groups will be tested on performance in basketball skills, such as dribbling, shooting baskets, etc. The ratings for both groups will be compared to see which group is superior in performance.

Using the symbol R to stand for the process of random assignment, we can diagram the processes of this experiment:

$\boxed{R}$ $\boxed{GP}$ $\boxed{T}$ $\boxed{O}$

$\boxed{R}$ $\boxed{GP}$ $\phantom{\boxed{T}}$ $\boxed{O}$

<u>COST</u>: (70 subjects)
35 + 5 + 5 + 5 + 5 + 10 +
5 + 5 = 75 points

This design might be called the real work-horse. Probably appropriate across a broader range of investigations in the behavioral sciences and education than any design, it can compete cost-wise. It provides fairly clear-cut information as to the relationship between treatment and post-treatment behavior. As this is often the major, and freqeuntly the sole question of interest, the Randomized Group Design is the appropriate selection.

It provides no information about pre-treatment behavior. As a consequence, there is no need to control for possible sensitizing effects of any pre-treatment measure.

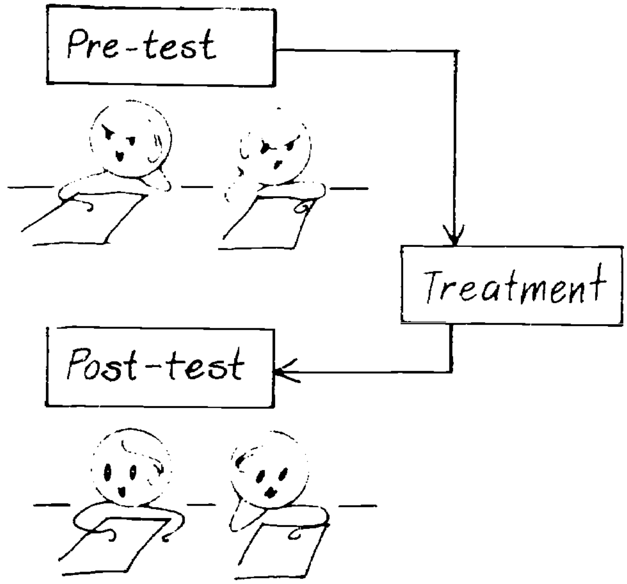
This design will answer questions about post-treatment behavior: questions P_1-1 and P_1-2 (the two questions answered by the Static Group Design). The groups are randomized so that this design will cope with the internal threats to validity. Subject changes due to other causes should affect the control group so that a comparison of the post-treatment behavior should reveal any differential effects of the treatment.

Class C, comparable group questions (C-1 and C-2) are also answered in the randomization provided that there are no probable differences between the groups entering the experiment except for the treatment.

The design will not cope with Class P_1-3, 4, and 5; P_2's questions of pre-treatment behavior; Class E questions; or Class R questions.

Pre-Post Randomized Group Design

This design adds a pre-test to the previous design as a check on the degree of comparability of the control and experimental groups <u>before</u> the treatment is given.



Example: All 7th grade students will be randomly assigned to Section A or Section B of the required art course. At the beginning of the course both groups will be pre-tested with the Psychedelic Art Awareness Test. Section A (Miss Smith's classes) will be given the experimental curriculum "Psychedelic Art in the 20th Century," while Section B (Mr. Jones' classes) will be given the standard required course in art. At the end of the semester, both groups will again be given the Psychedelic Art Awareness Test, to see if the new curriculum has achieved its objective of making students more aware of psychedelic art forms in their own environment. This could be diagrammed:

$\boxed{R}$	$\boxed{GP}$	$\boxed{O}$	$\boxed{T}$	$\boxed{O}$
$\boxed{R}$	$\boxed{GP}$	$\boxed{O}$		$\boxed{O}$

<u>COST</u>: (40 subjects)
20 + 5 + 5 + 5 + 5 + 5 + 5 +
10 + 5 + 5 = 70 points

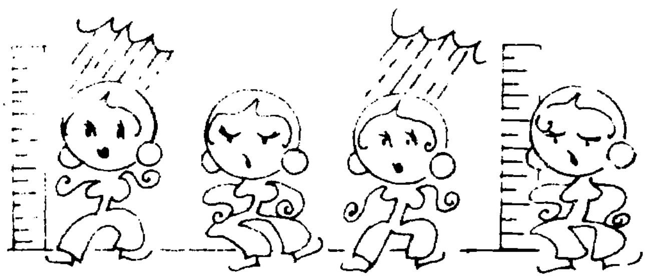
This yields information of P_1-1, P_1-2 as to post-treatment behavior (after treatment) and a comparison of post-treatment behavior between groups. It also answers most P_2 questions on pre-treatment behavior, questions P_2-1 through 4. It answers most of the Class I questions, that is, threats to

internal validity. It handles the Class C questions, the groups are comparable because they are randomized.

This design does not answer P_1-3, P_1-4 or P_1-5; nor the Class E questions relating to experiment errors; nor the Class R questions (relationship of the treatment); nor P_2-5, the effect of the pre-treatment observation on subsequent behavior of the subjects.

The Solomon Four Group Design

This more elaborate design attempts to control for the possible "sensitizing" effects of the pre-test by adding two more groups who have not taken



the pre-test at all. In the previous example given for the Pre-Post Randomized Group, it was not possible to control for the possibility that the Psychedelic Awareness Test may have made both groups acutely aware of the art forms around them so that both groups might score differently on the post-test as a result of having taken the pre-test. The actual effect of the experimental curriculum would be difficult to assess under those conditions.

Example: The Acne Attitude Test which measures degree of acceptance toward acne will be used to assess the effectiveness of the new Health and Hygiene text, "Skin Care, Facts and Fallacies," in changing attitudes toward acne in 7th grades. Before school begins, all 7th graders will be randomly assigned to one of four groups:

Group I - will receive the pre-test, the experimental text and a post-test.

Group II - will receive only the pre-test and the post-test.

Group III - will not receive the pre-test, but will use the experimental text, and be given a post-test.

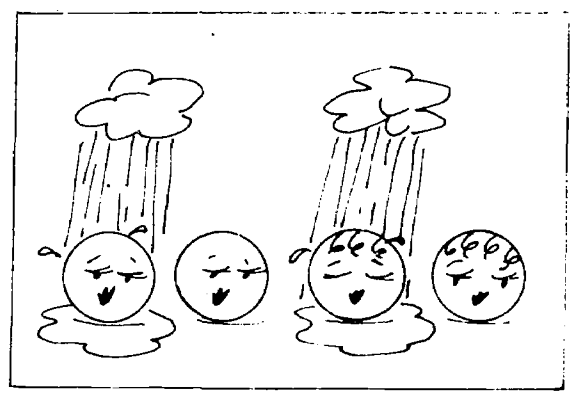
Group IV - will be given the post-test only.

The effect of the new text can then be determined independently from the possible sensitizing effects of the pre-test. This design could be diagrammed:

R	GP	O	T	O
R	GP	O		O
R	GP		T	O
R	GP			O

<u>COST</u>: (80 subjects)

40 + 20 + 20 + 10 + 20 + 20 = 130 points



The increased costs in this design must be justified by the need for information concerning possible effects of the pre-treatment observation.

The Solomon Design answers the P_1 questions, P_1-1, P_1-2, and P_1-3 (post-treatment behavior). It answers the P_2 pre-treatment questions, including the effect of the pre-treatment observation. It handles threats to internal validity, the Class I questions. It also handles the Class C questions because the groups are randomized. It does not handle P_1-4 and 5, nor the Class E questions, nor the Class R questions.

<u>The Randomized Block Design</u>

This design is of particular value when the experimenter wishes to determine the effect of a treatment on different types of people <u>within</u> a group.

For example, does the experimental method help one kind of student learn, but hinder another kind of student? Does it help the slow learner, but slow down the bright one? Does it work well f‹ a very well-adjusted student, but result in emotional upset for the very anxious student?

When this effect does occur, your experimental treatment may not seem to be having any effect, because the gain for one type of student is neutralized by the losses occurring in another type of student. This design is helpful in discovering the actual effect of your erimental treatment on different types of students.

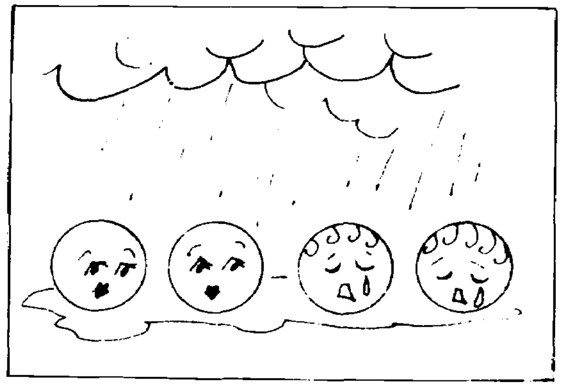
To achieve this, you will be dividing your experimental pool into "types" of students, or to use the technical phrase, into "blocks" of students.

Example: An experiment was designed to assess the effect of the Programmed Course in Auto Mechanics on both shy and extroverted boys. Using scores from the Shy-Outgoing Personality Test (which was given to all students the previous semester), students will be assigned (on paper) to either the "shy block" or the "extroverted block." Then students from each block will be randomly assigned to either Group I, the control group, or Group II, the experimental group.

Group II, the experimental group, will then be given the Programmed Course in Auto Mechanics, while the control group will receive the standard instruction in the same course.

At the end of the course, results on the final exam will be compared to see (1) if there is any difference between the control and experimental

groups, (2) if there is any difference between the scores of the shy and the extroverted boys using the programmed text, and (3) if there is an interaction between shyness-extroversion and the method of instruction



This could be diagrammed:

BLK T O

R GP

BLK T O

BLK O

R GP

BLK O

<u>COST</u>: (40 subjects)

20 + 10 + 10 + 20 + 20 + 20 = 100 points

Typically, this design refers to blocking or grouping of subjects with similar characteristics into treatment sub-groups. The group to be used in an experiment is usually given some pre-treatment measure, or previous records are examined, and the entire group is blocked or sorted into categories. Then equal numbers from each category are assigned to the various treatment and/or control groups.

While blocking according to subject characteristics is most typical of this design, blocking could be based on other relevant attributes.

For example, if subjects are to be treated during different times of the day, such as morning and afternoon, we might block a morning and an afternoon group within each treatment condition.

The importance of the design lies in the probability that the variable upon which the blocking is based may interact with the treatment. Frequently, no overall treatment effect is observed because subjects with different characteristics react differentially to treatment. If they were blocked on the appropriate attributes, differential treatment effects would be revealed.

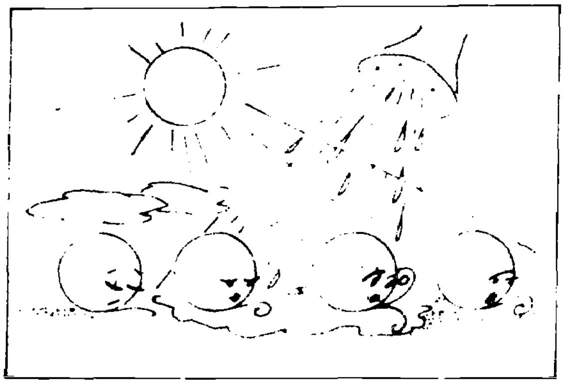
The Randomized Block Design will handle the P_1 questions, P_1-1, P_1-2 and P_1-3. It does handle the Class I questions. It handles the Class C questions of comparable groups. It does not handle the P_1-4 and P_1-5 or the P_2 questions of pre-treatment behavior. It does not handle Class E questions. It does handle one, and only one, of the Class R questions, R-1, the relationship of the treatment to subjects characteristics, by virtue of blocking.

The Factorial Design

Recall that in a blocking design, the subjects were assigned to different groups on the basis of some of their own characteristics such as high or low intelligence, etc.

Sometimes we wish to assign different variations of the treatment as well, and the procedure is similar. For example, we may wish to try two kinds of treatments varied in two ways. (This would be called a "2 x 2" factorial design.)

Some factorial designs include both assignment of subjects (blocking) and several types of experimental treatment in the same experiment. When this is done it is considered to be a factorial design. For illustration here, only the treatments will be varied.



Example:

An experiment will investigate the effect of praise and reproof on performance of sophomore classes in Essay Writing. Students will be required to turn in an essay per week. No grades will be given, but the teacher will either praise the essay (Condition A_1) or express disapproval of the essay (Condition A_2). Praise or reproof will be either <u>written</u> on the essay (Condition B_1) or given to the student <u>orally</u> (Condition B_2).

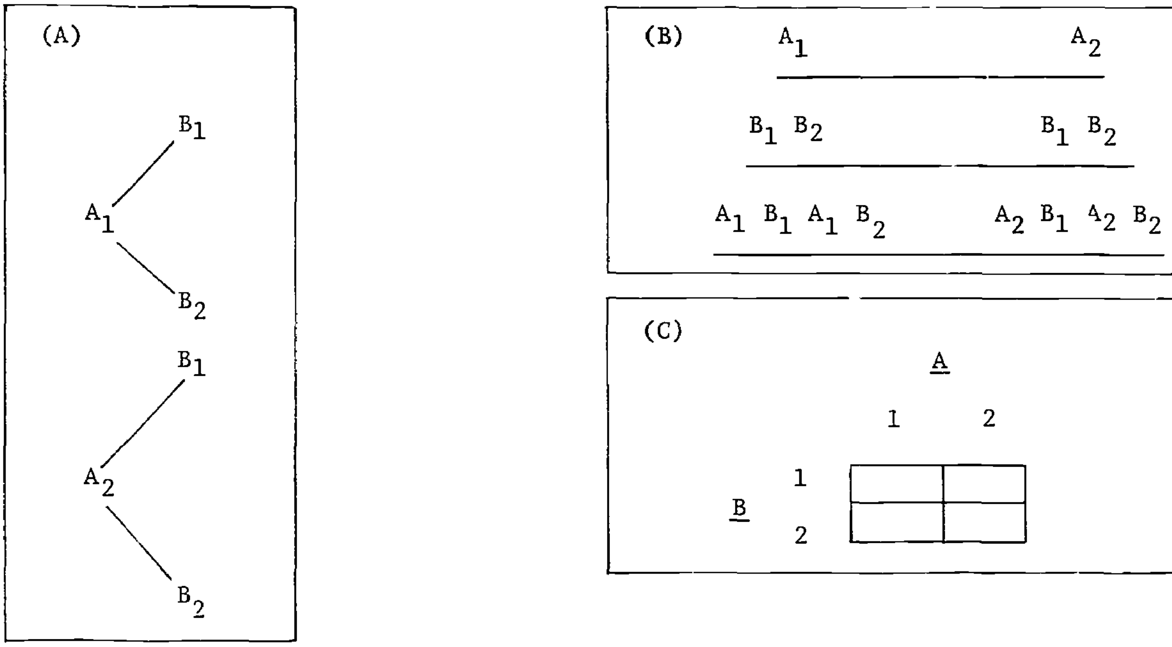
Students will be randomly assigned to one of four groups:

Group I	Praise written	(A_1 B_1)
Group II	Praise oral	(A_1 B_2)
Group III	Reproof written	(A_2 B_1)
Group IV	Reproof oral	(A_2 B_2)

The final exam will be an essay to be written in the class. It will be assessed to determine the effect of the experimental treatments. This can be diagrammed in different ways:

2 kinds of treatments (factors): A and B

Each varied in 2 ways (levels): 1 and 2



R	GP	T A_1 B_1	O
R	GP	T A_1 B_2	O
R	GP	T A_2 B_1	O
R	GP	T A_2 B_2	O

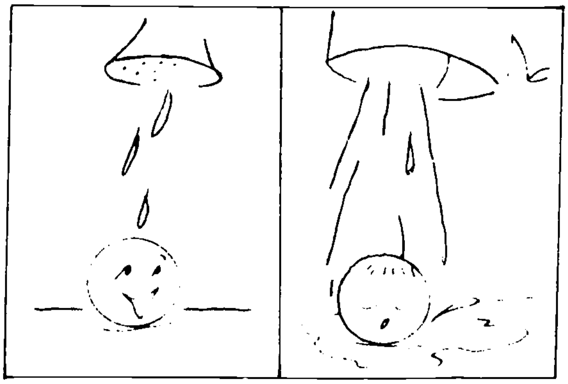
COST: (40 subjects)

20 + 20 + 20 + 40 +

20 = 120 points

The Factorial Design, as we are describing it, is really a complete factorial, rather than an incomplete factorial of which there are several kinds. The factorial is appropriately used when we wish information

concerning the effects of different kinds or intensities of treatment. The factorial provides relatively economical information not only about the effects of each treatment, level or kind, but also about <u>interaction</u> <u>effects</u> of the treatment.



In one "2 x 2" factorial such as that given in the example above, information can be gained about the effects of each of two treatments, the effect of each of two levels within each treatment, and the interaction of the treatments. If all these are questions of interest, the factorial design is much more economical than running separate experiments.

The factorial handles some of the same classes of information as the previously described randomized group design; P_1-1, P_1-3, and Class C questions. Little support for Class I questions may be obtained, but some weak inferences might be drawn. This design also answers Class R questions, R-2, R-3, and R-4 concerning relationships among treatments, factors, and levels. It does not answer P_1-2, P_1-4, P_1-5, or P_2 questions; Class E questions; or questions R-1 and R-5.

Factorials can run to enormous and uninterpretable proportions. "2 x 3 x 4's" are not uncommon. One of the authors worked on a project involving a "4 x 3 x 4 x 4" factorial. Higher order interactions, i.e., that variable D's influence depends on variable C's interaction with A, B, and E (including all permutations of this relationship) present a

puzzle to most researchers.

<u>The One-Shot Repeated Measures Design</u> (1)

This design, or variations of it, is used to assess the effects of a treatment with the same group or the same individual over a period of time. A measure, or observation is made more than once to assess the effects of the treatment.

Example:

Does the All-New-Innovative-Reading Method result in increased reading achievement in deprived children? This method will be used with all the children in Mrs. Jones' first grade class beginning in September. The Reading Achievement Test will be given in October. The method will be continued throughout the school year with an additional Reading Achievement Test being given in February and in June. Reading Achievement Test scores will be examined to determine if there has been a significant gain in achievement due to the use of the new method.

This design can be diagrammed:

$\boxed{GP}$ $\boxed{T}$ $\boxed{O}$ $\boxed{T}$ $\boxed{O}$ $\boxed{T}$ $\boxed{O}$

<u>COST</u>: (20 subjects)

10 + 5 + 10 + 5 + 10 + 5 + 10 + 5 = 60 points

This design is an extension of the simple one-shot and adds only information regarding the effects of repeated or continued treatment. Often an economical trial balloon, the design can acquire high yield when other extra-design sources of knowledge can be related to it.

This design handles only the questions related to Class P_1 or post-treatment behavior, questions P_1-1 and P_1-4. It answers questions about behavior shortly after treatment and the longer-term effects related to subsequent treatments. It might handle one Class R question, R-5, relationships of the treatment, in that the effect of the repeated treatments may be observed. It will not answer P_1-2, P_1-3, P_1-5; nor any Class P_2, I, C, or E questions; nor question R-1 through 4.

Randomized Groups Repeated Measures Design (2)

A variant of the previous design in which two or more experimental methods are compared and repeatedly measured or observed. Although we are using two groups for illustration, any number may be used.

Example:

You now wish to compare the All-New-Innovative-Reading Method with the newly published Plenty-Phonics-Phor-Phine-Phirst Graders Reading Method. Therefore, in this experiment, you randomly assign all first graders to either Group I (the All-New Method) or Group II (the Plenty-Phonics Method). You test both groups with the Reading Achievement Test in October, February, and June, to compare and assess the two methods. We can diagram this design as follows:

R	GP	T	O	T	O	T	O
R	GP	T_2	O	T_2	O	T_2	O

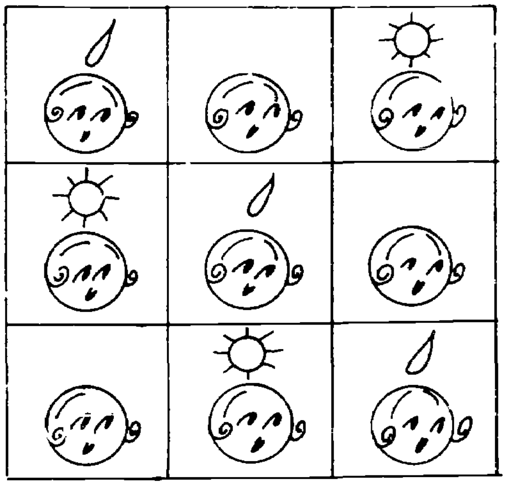
COST: (40 subjects)

20 + 10 + 10 + 20 + 10 + 20 + 10 + 20 + 10 = 130 points

This design answers P_1 classes of information, P_1-1, P_1-3 and P_1-4. It does handle Class C questions. The groups are comparable because of randomized assignments. Some light is shed upon the comparability of repeated treatments, question R-5.

It fails to handle P_1-2, P_1-5, or Class P_2, pre-treatment behavior. It does not handle well Class I questions, internal threats to validity, in the present example. However, the addition or substitution of an appropriate non-treatment group would yield Class I information. It fails to handle Class E questions and most of Class R.

The Latin Square Design



You may wish to use several different treatments in the same experiment, for example, a programmed text, a film, and a field trip. If you suspect that the order of presentation of these different treatments may affect the outcome, i.e., that more information is gained from the text before a field trip than after, you can enlarge a repeated measures design by including the three treatments in varying order.

ERIC

Example:

A new class is being organized to teach classroom teachers about Mental Retardation. They will be tested (observed) three times during the course on their knowledge of retardation. The experimental treatments will be T_A, a film on mental retardation; T_B, a programmed text on retardation; and T_C, a field trip to an institution for the retarded.

All subjects will be randomly assigned to one of three groups which will have information presented in one of three orders:

Group I	T_A T_B T_C	film, text, trip
Group II	T_B T_C T_A	text, trip, film
Group III	T_C T_A T_B	trip, film, text

After each experience, the teachers will be tested on their knowledge of retardation and results will be compared to see which order of presentation has been most effective.

This can be diagrammed:

R	GP	T_A	O	T_B	O	T_C	O
R	GP	T_B	O	T_C	O	T_A	O
R	GP	T_C	O	T_A	O	T_B	O

COST: (60 subjects)

30 + 15 + 15 + 30 + 15 + 30 + 15 + 30 + 15 = 195 points

A researcher might want to explore all possible combinations in this case. They might be diagrammed in six orders.

Group I	A	B	C	film, text, trip
Group II	B	C	A	text, trip, film
Group III	C	A	B	trip, film, text
Group IV	A	C	B	film, trip, text
Group V	B	A	C	text, film, trip
Group VI	C	B	A	trip, text, film

Obviously, the cost would be doubled.

This Latin Square Design will yield P_1-1, P_1-3, C-1, C-2, and R-5. It will not provide information on P_1-2, P_1-4, P_1-5, nor questions in the I and E classes. Questions R-1 through 4 are not answered.

The Latin Square represents a double blocking: rows and columns. Typically, however, one or more of these is based on time.

There are a number of possible variants of the Latin Square. Frequently, in psychological and physiological research, individual subjects form the rows; time, the columns; and different treatments, the cells. In this type of Latin Square Design, an observation is taken at each treatment cell. When this is done and the observations between treatments are to be compared, clear interpretation may be difficult because of possible carry-over effects, that is, earlier treatments affect later ones.

It does seem to the authors that this use of the Latin Square is not of great value in educational research. Perhaps a more appropriate application would be one in which tested subjects are given different orders of treatment, that is, groups of subjects comprise the basis for rows, columns are time periods and treatments are the cells. This allows the product

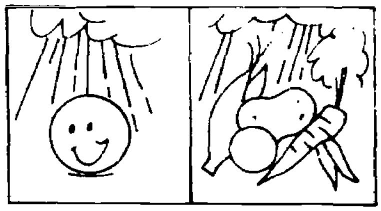
designer to capitalize on the possibility of carry-over effects. Observations may be made at the end of the sequence, during the sequence, and compared across groups. A number of varying applications of the Latin Square are given in Edwards (1968), and Cox (1958).

A hampering restriction on the use of Latin Squares is that rows, columns, treatments (or whatever designates cells) must all equate. Procedures for selection of Latin Squares and descriptions of 2 x 2, 3 x 3...7 x 7 squares are given in any conventional design text.

(See Figure 2 on the following page.)

6.1

<u>The External Validity Problem</u>



Questions of a different sort than we have faced arise from our need to generalize from a limited set of observations. No one, even the project director, is long interested in observations that in no way extend beyond his particular project. Most seekers of knowledge display this propensity to generalize from one observation to an eternal truth. Psychoanalysts, behavior modification specialists, and barroom raconteurs have cultivated the skill to its highest degree.

Generalizability depends on whether the observed behavior (O) is representative of the people, surrounding conditions, and the treatments to which we now wish to extend it. Classes of questions include:

Did some early procedure in the project affect the subjects so that their later behavior was, in part a result of that?

Figure 2. EXPERIMENTAL DESIGN SUMMARY SHEET

SUMMARY OF DESIGN YIELDS AND COSTS IN UNITS

TYPE OF DESIGN	INFORMATION YIELD	STRUCTURE AND COST
#1 ONE-SHOT	P_1-1	(NOTE: The figure 20 refers to 20 subjects) 20 + GP + T + O 10 + 5 + 10 + 5 30 units
#2 ONE-GROUP, PRE-POST	P_1-1 P_2-1 P_2-2	20 + GP + O + T + O 10 + 5 + 5 + 10 + 5 35 units
#3 STATIC GROUP	P_1-1 P_1-2	20 + GP + T + O 10 + 5 + 10 + 5 20 + GP + O 10 + 5 + 5 50 units
#4 RANDOM GROUP	P_1-1 P_1-2 I-1 I-2 C-1 C-2	20 + R + GP + T + O 10 + 5 + 5 + 10 + 5 20 + R + GP + O 10 + 5 + 5 + 5 60 units
#5 PRE-POST RANDOMIZED GROUP	P_1-1 I-1 P_1-2 I-2 P_2-1 P_2-2 C-1 P_2-3 C-2 P_2-4	20 + R + GP + O + T + O 10 + 5 + 5 + 5 + 10 + 5 20 + R + GP + O + O 10 + 5 + 5 + 5 + 5 70 units

Figure 2. (continued)

#6 SOLOMON FOUR GROUP	P_1-1 P_1-3 P_1-2 P_2-1 I-1 P_2-2 I-2 P_2-3 P_2-4 C-1 P_2-5 C-2	20 + R + GP + O + T + O 10 + 5 + 5 + 5 + 10 + 5 20 + R + GP + O + O 10 + 5 + 5 + 5 + 5 20 + R + GP + T + O 10 + 5 + 5 + 10 + 5 20 + R + GP + O 10 + 5 + 5 + 5 130 units
#7 RANDOMIZED BLOCK	P_1-1 P_1-2 P_1-3 C-1 I-1 C-2 I-2 R-1	BLK + T + O 20 + R + GP + 5 + 10 + 5 20 + 5 + 5 + BLK + T + O 5 + 10 + 5 BLK + O 40 + R + GP + 5 + 5 20 + 5 + 5 + BLK + O 5 + 5 140 units
#8 FACTORIAL	P_1-1 P_1-3 C-1 C-2 R-2 R-3 R-4	20 + R + GP + T(AB) + O 10 + 5 + 5 + 10 + 5 20 + R + GP + T(AC) + O 10 + 5 + 5 + 10 + 5 20 + R + GP + T(DB) + O 10 + 5 + 5 + 10 + 5 20 + R + GP + T(DC) + O 10 + 5 + 5 + 10 + 5 140 units
#9 ONE-SHOT REPEATED MEASURES	P_1-1 P_1-4 R-5	20 + GP + T + O + T + O + T + O 10 + 5 + 10 + 5 +10 + 5 +10 + 5 60 units

ERIC

Figure 2. (continued)

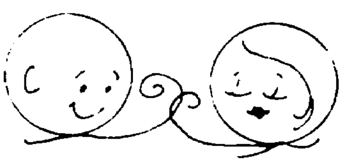
#10 RANDOMIZED GROUPS REPEATED MEASURES	P_1-1 P_1-3 P_1-4 C-1 C-2 R-5	$20 + R + GP + T + O + T + O + T + O$ $10 + 5 + 5 + 10 + 5 + 10 + 5 + 10 + 5$ $20 + R + GP + T_2 + O + T_2 + O + T_2 + O$ $10 + 5 + 5 + 10 + 5 + 10 + 5 + 10 + 5$ 130 units
#11 LATIN SQUARE	P_1-1 P_1-3 C-1 C-2 R-5	$20 + R + GP + T_1 + O + T_2 + O + T_3 + O$ $10 + 5 + 5 + 10 + 5 + 10 + 5 + 10 + 5$ $20 + R + GP + T_2 + O + T_3 + O + T_1 + O$ $10 + 5 + 5 + 10 + 5 + 10 + 5 + 10 + 5$ $20 + R + GP + T_3 + O + T_1 + O + T_2 + O$ $10 + 5 + 5 + 10 + 5 + 10 + 5 + 10 + 5$ 195 units

Were the subjects themselves a representative sample of the general population of people?

Was there some atmosphere about the project or surrounding setting that would dispose the subjects to act differently from non-project people?

Was the treatment accompanied by any personal interaction that may be somewhat peculiar to the project or the people or the experimenter involved?

The designer needs to face this line of questioning. You do not need to build enduring generality into each study. Unfortunately, every penny abortion project in the country tends to be accompanied by assertions of sweeping truths over space, time, and peoples. Just try to clarify where the results of your observations may be legitimately extended and where they can not yet legitimately be extended.



The reader may increase this set of eleven standard designs by variations and hybridization among the designs. We will describe an additional set of designs in a later version of this text. Most of the additional standard experimental designs are mutations or fractional replications of the eleven designs enumerated. For reasons of economy, when the entire design cannot be implemented, a fractional factorial, etc., is used.

The following list enumerates those of general use:

1. Incomplete Block Design

A fractional randomized block design in which the number of treatments exceed the number of experimental units available for the block.

2. Youden Square Design

   Resembling a Latin Square but a design in which the number of treatments exceeds <u>either</u> number of rows or columns --but not both.

3. Lattice Square Design

   Resembling a Latin Square, but in which number of treatments exceeds <u>both</u> the number of rows and number of columns.

4. Fractional Factorial Design

   A truncated factorial containing only a subset of all possible factor combinations.

5. Graeco-Latin Square Design

   Use of another variable to group the experimental units in three ways, as compared to a Latin Square which groups them in two ways, e.g., row and column classification. Typically, letters from the Greek alphabet are combined with letters from the Latin alphabet. Any system can be used, however. The process can be extended to 3, 4, 5 or any number of "alphabets."

6. Split-Plot Design

   Used when treatment, usually factor combinations, exceeds the available experimental units. One factor is held

constant throughout an entire block or superordinate grouping. Clear information on that factor is sacrificed to achieve other information.

7. Covariance Design

   More a statistical procedure than a design, the technique uses some concomitant measure which may be correlated with post-treatment observations. Post-treatment observation scores are adjusted to remove variance associated with the concomitant variable. The effect is similar to blocking.

8. Time Series Design

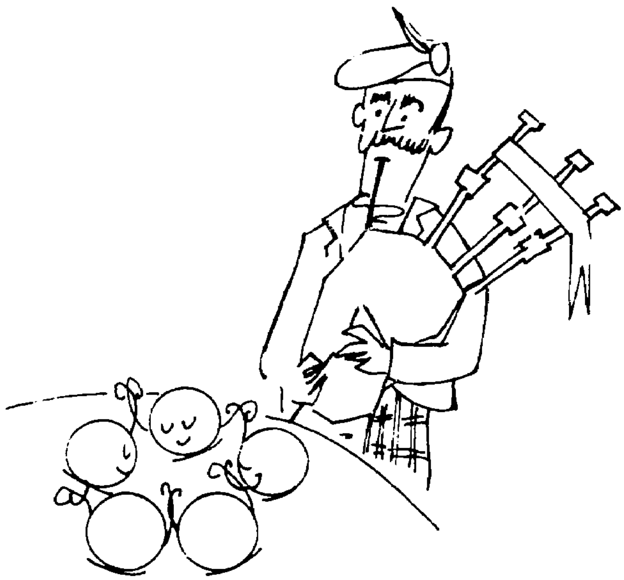
   A variant of repeated measures design in that more observations than treatments are given, e.g., O O O X O O. This emphasizes the collection of pre- and post-baseline information.

Our intent in this text is that the reader acquire not only the objectives enumerated in the front of the chapter, but a <u>confident approach</u> toward the subject of experimental design. Such a feeling, reality based, has seemed to us more important at the start than computational approaches, etc.

The reader should be alerted that our approach to design is characterized by a higher risk-taking orientation than is usual in design and statistics instruction. We have placed primary importance on identifying a design that will produce the kind of information desired--within cost constraints. That the learner's use of a design, and subsequent analysis, may violate some standard formal assumptions, e.g., statistical, has not concerned us greatly.

Violation of statistical assumptions is a recurring plaint of the defensive variety of statistician. Such cries of violation are about as appropriate as complaints of virginal violation from the working staff of a house of prostitution. All sorts of violations continuously take place in both fields--without complaint. In fact, neither could operate effectively otherwise. Practical minded practitioners who face up to this often use the term "robust" to characterize ability to recover from, or bear up under, violation. (Edwards, 1968)

The attached reference lists identify some standard sources for additional information. If within the areas of design described you wish to develop further competencies, the accompanying workbook contains a brief set of exercises. Further competencies may be gained by taking the complete Thrifty Pea Picker s Guide to Experimental Design.



6.2

References

Campbell, Donald T. Reforms as experiments. American Psychologist, April, 1969, 409.

By one of the more astute writers on design, the article illustrates the use of varying repeated measure approaches to complex social problems.

Campbell, Donald T., and Stanley, Julian C. Experimental and quasi-experimental designs for research on teaching. In Gage (Ed.), Handbook on research on teaching. Chicago: Rand McNally & Co.,1963.

Also available as a separate paperback. Minimizes statistics, emphasizes threats to validity of each design. By all means, read this one.

Edwards, Allen L. Experimental design in psychological research. New York Holt, Rinehart & Winston, Inc., 1968.

Long a classic text in the behavioral sciences, a combination of logic, statistics, and computational procedures. Probably easiest to read of any of the standard graduate texts.

Finney, D. J. The theory of experimental design. Chicago: The University of Chicago Press, 1960.

Written primarily for the statistician, however, much of the text can be read with little more than an introductory statistics background. We found the chapters on Efficiency of Experimentation and Economics of Experimentation particularly interesting.

Kerlinger, Fred N. Foundation of behavioral research. New York: Holt, Rinehart & Winston, Inc., 1964.

A voluminous book attempting to cover all aspects of research at an introductory level. The 85-page section on design is easy to read and replete with detailed examples which assist the beginning reader. Written from the "good scientist" moralistic standpoint.

Lindquist, E. F. Design and analysis of experiments in psychology and educationBoston, Mass.: Houghton Mifflin Co., 1953.

The classic approach, quite similar to Edwards'. The exemplary designs and problems stress educational research.

McGuigan, F. J. Experimental psychology: A methodological approach. (2nd ed.) New Jersey: Prentice-Hall, Inc., 1968.

Extremely clear treatment and comparisons of several basic designs. The author's explanation of control procedures, while extending beyond the field of design, is a major contribution.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Although the author asserts an emphasis on the logic of design, the text frequently reflects more concern with computation and formula. A more thorough and rigorous treatment than the typical behavioral science or education text. Assumes a middling competency in statistics.

Section VII: Data Analysis

Editorial Foreword

In this chapter James Beaird presents a useful guide for the statistically illiterate reader, a guide to the selection of appropriate data analysis tools. These recipes for the use of 18 analysis techniques require no previous statistical background. However, the chapter is tightly organized and closely written, and for the naive student it may be slower going than some of the other chapters.

The author attempts to instruct the reader in how to ask and answer four successive questions. The branching path so followed leads to an appropriate statistical technique for data analysis. The four basic questions which form this decision tree are:

1. What is your question? (three alternatives)
2. What level of measurement has been used to collect the data? (three alternatives)
3. Are the samples independent or not? (two alternatives)
4. Are the samples related or independent? (two alternatives)

The essence of the chapter is neatly summarized in the tables on pages 15 and 16. The editor found scanning these tables useful before and after reading the text.

Throughout the chapter brief verbal explanations of several crucial measurement and statistical concepts are given. These may be of assistance. However, the reader can effectively use the guide whether or not he has a fair grasp of the concepts involved.

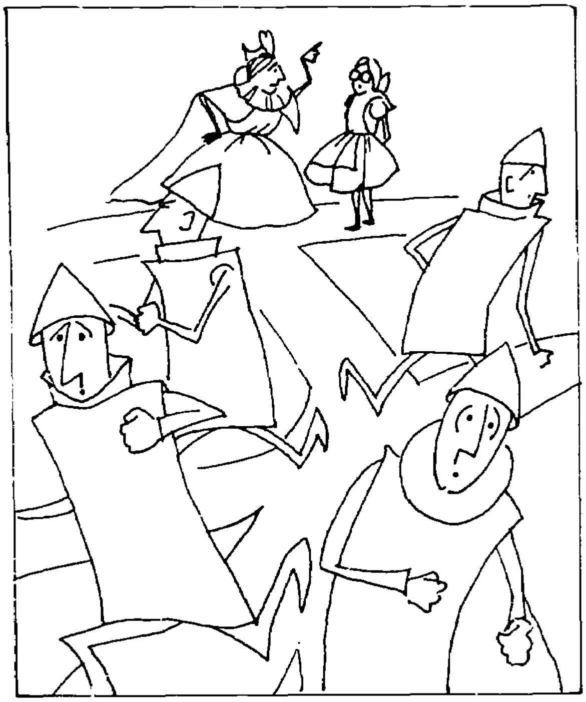
The author did not intend, nor should the reader assume that the techniques suggested are the only ways to answer or analyze a particular

VII

question. However, elaboration of alternative techniques would require a much extended chapter. The reader may get the impression, therefore, that the chapter straps him in and hurtles him down a track with no choice of switching or destination. The reader is correct.

However, I know of no other source where a novitiate to statistics can obtain a guide covering so many classes of data problems with such little effort.

"Either you or your head must be off."



VII

ERIC

DATA ANALYSIS

James H. Beaird

Introduction

Objectives

Perhaps the easiest way to get into a list of what behaviors we would expect the learner to have after completing the study of the contents of this chapter would be to first specify what the chapter does not intend to develop in the learner. It is not the intent of this chapter to develop even the rudiments of competence in using any of the statistical analysis tools. It is not the intent of this chapter to develop the basic skills in probability theory to such a point that the student is then better equipped to initiate a study of analysis techniques. It is not the intent of the chapter to cover all of the myriad of statistical techniques which might be applied in educational research settings.

It is the intent of the chapter to acquaint the student with some of the considerations that he must give to the selection of an analysis technique appropriate for the study he is initiating and/or critiquing. It is the purpose of the chapter to provide the student with a ready made reference and procedures for using that refer e such that he can identify by name the statistical tool appropriate for many of the research situations faced by educators.

In somewhat behavioral terms the following are objectives for this chapter:

1. Given a description of a research problem the student will correctly answer the question, what is my research question?, by supplying one of the following alternatives: compare, relate, or describe.

2. Given a description of the measurement tool and/or technique utilized in the research question, the student will correctly identify the level of measurement suggested by supplying one of the following alternatives: nominal scales, ordinal scales, interval scales, or ratio scales.

3. The student will correctly identify data resulting from the use of interval or ratio scales as being Class A data.

4. The student will define data resulting from the use of ordinal scales as being Class B data.

5. The student will define data resulting from the use of nominal scales as being Class C data.

6. Given a description of a research situation the student will correctly identify the number of samples used in that research situation.

7. Given a description of a research situation the student will correctly determine whether or not the samples are related or independent.

8. The student will be able to describe the rationale underlying the necessity for using null hypotheses in comparative studies.

9. The student will state that the formal null hypothesis contains within it the basic assumptions underlying that hypothesis.

10. The student will differentiate between experimental studies and causal-comparative studies and will state the formal null hypothesis appropriate for each type of study.

11. Given a description of a research situation, the student will identify by name the statistical tool appropriate for the treatment of data coming from that study.

Orientation

In determining the appropriate analysis technique one must consider the following factors:

1. What is my research question?

2. What is the nature of my data?

3. How many samples do I have?

4. Are my samples related or independent?

Three types of research questions will be considered: (1) Experimental or Comparative, (2) Relational, and (3) Descriptive. For each type of question one of several techniques may be used for analysis depending on other factors, e.g., the level of measurement available, the number of samples, the independence of the samples. Let's consider each of these in turn.

7.1

Levels of Measurement

Four levels of measurement may be identified. These levels are defined in terms of the scaling characteristics of quantifiable descriptors (numbers) each provides. The four levels are: nominal, ordinal, interval and ratio.

Nominal Scales. Nominal scales are those in which the numbers are merely substitute names for the objects being scaled. In the usual quantitative sense the number has no meaning. That is, the number does not indicate that the object has more or less of a characteristic than does any other object. Probably the most common use of a nominal scale are the numbers assigned to various members of an athletic team. Another commonly used system of nominal numbers is the Dewey Decimal System employed by many libraries.

Ordinal Scales. The lowest level of scales in which numbers have quantitative meaning are ordinal scales. As in nominal scales these numbers also differentiate between the objects being scaled; however, in addition the numbers indicate which of two objects have more of the characteristic. In some states, for example, schools are classified according to size of enrollment. The numbers assigned to the classes indicate a difference in size, e.g., Class I schools are larger than Class II schools, etc. If I were to order the institute members

according to their height, assigning the number 1 to the tallest, 2 to the second tallest, etc., the set of numbers would be an ordinal scale.

Interval Scales. Interval scales possess all of the characteristics of nominal and ordinal scales, i.e., they use numbers to name objects and the numbers have quantitative meaning, e.g., a value of 10 may mean that the object has more of the characteristic than does the object labeled 9, or 8, or 3. One important additional characteristic is added to the interval scale and that is that consecutive numbers are equally spaced along the scale, i.e., the intervals between the numbers are equal. This means that the distance between 9 and 13, for example, is equal to the distance between 7 and 3. The Fahrenheit thermometer is an interval scale. Is the centigrade thermometer? On an interval scale is 8 twice as large as 4? Why?

Ratio Scales. The highest order of scales is the ratio scale which adds a "zero point" to the highest interval scale. This addition is important since it permits us to say that one score is many times greater than another, e.g., 16 is twice as large as 8 or 4 times greater than 4 or 1/3 as large as 48. Measures of distance or weight are ratio scales.

Most measures of psychological attributes fail to meet the criteria of ratio scales. In fact they rarely are sophisticated enough to be interval scales. They often are more than ordinal. Therefore, the data levels will hereafter be referred to as being:

Class C: Nominal only

Class B: Ordinal only

Class A: Better than ordinal


7.2

Number of Samples

Comparative studies may be made using two or more samples. The distinction must be made between situations in which just two samples are being compared and in which more than two samples (any number of samples) are being compared.

Let's consider two sample studies first. Any time subjects of one type are compared with subjects of another type, a two sample comparison is involved. This implies to the investigator that he must choose one group to represent one type of subjects and another group to represent the second type of subjects. Comparisons might also be made of subjects training under one set of conditions with those who trained under another set of conditions.

Graphically, two sample studies may be portrayed as follows:

1. MEN vs. WOMEN

2. SIGHTED vs. BLIND

etc.

When more than two samples are involved several conditions must be considered. Suppose in a given company we wish to compare attitudes towards militant civil rights activities held by (1) management personnel, (2) supervisory personnel, and (3) nonsupervisory personnel. Obviously we would be required to select three samples (one for each classification of personnel). A similar situation might be the comparison of effectiveness of three teaching techniques.

In each case the categorization was made along a single factor or dimension, e.g., type of worker, or teaching method. Graphically such studies

would be portrayed as follows:

1. Categories in one dimension

Teaching Method

a.

A	B	C

School

b.

Engineering	Education	Arts & Sciences

Suppose in the second example, we suspected different learning capabilities for boys and girls. We may then add a second dimension, i.e., a dimension for sex which has two categories. Addition of this dimension increases the number of samples required. In this example we have three categories of method and two categories of sex, and must, therefore have six samples (3x2).

There is theoretically, no limit to the number of dimensions which might be included in a study. The point to be made here is that the number of samples increases by a multiple of the number of categories in each added dimension.

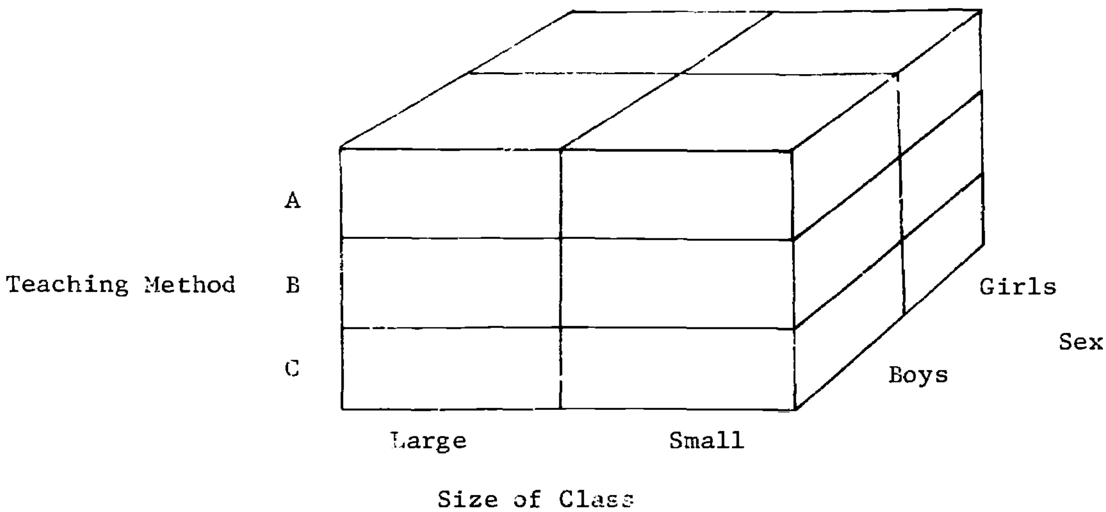
Graphically, studies in two or more dimensions might be portrayed as follows:

a.

Sex		A	B	C
	Boys			
	Girls			

ERIC

b.



<u>Independence of Samples</u>

Samples fall into one of two categories--they are either related or are independent. Independent samples are those which were randomly or psuedo-randomly selected. This means that any case within a population of cases had an equal opportunity to be selected. For most studies utilizing parametric statistical procedures, it is assumed that the samples were randomly selected from infinitely large, normally distributed populations.

Technically speaking, it is difficult to justify the randomness of samples used in many educational studies. Psychologists, for example, often will assign various sections of a course they are teaching to different treatment groups and conduct their study as though each section was a random sample drawn from some hypothetical, infinitely large population. I am certain that this practice stimulates all sorts of anxiety, wrath, and scorn in the statistical purists. Such practice, however contaminated, seldom results in <u>decisions</u> different from those which might result from using completely random samples.

In my opinion, the three assumptions (randomness, infinite populations, normally distributed) can be horrendously violated without affecting the decision to be made.

Related samples are those which do not, for various reasons, meet criteria of randomness. An example of related samples is the comparison of pre-and post-training scores for a group of subjects. Obviously post training performance can be expected to be related to initial performance. Similarly if all subjects in a study are subjected to all treatments, the samples are related. Analysis techniques differ for studies using related and independent samples.

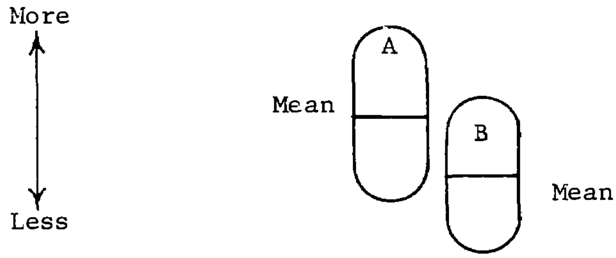
<u>A Note on Hypotheses</u> 7.3 7.4

Much of what we are about in this business of research is involved with hypothesis generation and testing. When we develop a new instructional technique, approach to group therapy, or measurement tool, we do so with the intent and hope that it will in some way do the job better than has been done previously and we so hypothesize. Likewise we generate hypotheses from various theoretical origins and then set about to gather evidence to substantiate our hypothesis. Such hypotheses are almost always conceptualized and thought of positively. We hypothesize that Technique A is superior to Technique B, that students of <u>this</u> type will out perform students of <u>that</u> type. Such hypotheses are natural and expected and desirable because they are statements of our intendeds.

Unfortunately, such direct, positive statements of our intendeds cannot be verified directly with the statistical procedures available to us. We simply have no way to verify a statement that the effect of A is greater than that of B unless there exists a condition of invariance in our repeated observations of the effects of both A and B. It is, I believe, safe to conclude that in the empirical world of our endeavor, conditions of invariance are nonexistent.

Full Text Provided by ERIC

Therefore when we repeatedly observe the effects of A we find that our observations fall into some range; likewise, our observations of B's effects <u>vary</u> within some range. This necessitates our choosing some descriptor of the effects of A and B--usually some central observation (such as the mean) which is representative of all of our observations. Consider the following figure.



The ovals indicate variation of the observations of the effects of treatments A and B. We have hypothesized A to have more effect than B. Our means tend to substantiate our hypothesis, but certainly the overlap of the two sets of effects raises some doubts. Our question becomes one of asking "given this variation, how much difference must be observed between the two means before we can reasonably conclude that A has more effect than B?"

For various and sundry reasons we must alter our strategy and use a little logic. The strategy we must employ is to forego a hypothesis of inequality (effects of A are greater than effects of B) and replace it with a statement of equality (the effects of A and B are equal).

Given a hypothesis of equality we can gather evidence which permits us or fails to permit us to reject the hypothesis. Rejection of the hypothesis of equality permits the acceptance of our positive hypothesis, i.e., within probability limits the positive statement becomes tenable. On the other hand, hypotheses of equality are never accepted. Since the original intent was to

gather evidence which would permit rejection of the hypothesis of equality, having gathered that evidence and finding it insufficient in strength to so reject, we cannot then use it as evidence to accept the hypothesis.

Usually the statement of equality is made in the form of a null hypothesis, e.g., the differences between treatment groups are equal to zero. I would like to focus my remaining comments upon formalized statements of the null hypothesis.

Researchers rarely state their null hypotheses formally. This practice makes it difficult for the reviewer of the research to determine the appropriateness of the statistical tool employed in the research. The formal statement of the null hypothesis, while laborious, does include a complete statement of the assumptions made and thereby permits examination of the fit of the statistical model to the assumptions.

As we noted previously, when engaged in empirical research we usually concern ourselves with both the means and variances of groups of scores. Analysis techniques now employed involve both means and variances of samples. Some techniques assume that samples were drawn from single, homogeneous, populations of scores, while others assume sampling from two or more populations of scores. Some techniques address themselves to means <u>and</u> variances, while others are concerned with only means <u>or</u> only variances. Whatever the concern, the formal null hypothesis should so specify.

Let's consider one large class of studies which seek to determine the nature of differences--experimental studies. Given a large number of high school students, an investigator wishes to know whether a film he has developed has any impact on their attitudes toward Viet Nam. Randomly, he selects two groups of subjects, shows the film to one group and not to another. The appropriate formal null hypothesis should be:

ERIC

Null Hypothesis I. The differences between the means and variances of the samples are no greater than differences due to the vagaries of random sampling from a single, normally distributed, infinite population.

Keep in mind that this was an experimental study, i.e., the groups were alike at the beginning and were subsequently treated differentially. Experimental studies are not limited to two groups. Likewise Null Hypothesis I is appropriate for studies involving two or more groups.

Another large class of studies also involves comparisons; however, the comparisons are between groups which are known or suspected to be different at the onset of the study. Stated in another way, the groups differ but the difference is outside the control of the experimenter. For example one investigator might wish to study maze learning ability of white and hooded rats. Another might wish to know if Negro students and Caucasian students exhibit differential tonal acuity. Studies of this type are known as causal-comparative studies.

Null Hypothesis I specifically referred to samples randomly drawn from a single population. By no stretch of the imagination could we expect to randomly draw two samples from a single population of rats and have one pure sample of white rats and another pure sample of hooded rats. Thus Null Hypotheses I just does not fit.

For Causal-Comparative studies the appropriate formal null hypothesis is:

Null Hypothesis II. The differences between means of the two samples is no greater than the differences to be expected from the vagaries of random sampling from two normally distributed infinite populations with equal means.

In summary, Hypothesis I states that there is a single population and experimental studies start with a single population. On the other hand Hypothesis II states that there are two or more populations, and Causal-Comparative

ERIC

studies distinguish between two or more samples on _a priori_ bases. What could be nicer than that?[1]

Degrees of freedom. The confidence that can be placed in conclusions is related to the "degrees of freedom" associated with the analysis. The concept "degrees of freedom" is quite complex and will not be covered in any depth here. Suffice it to say that usually the number of degrees of freedom is one less than the number of observations, i.e., if you have 20 cases in your sample (n=20) the number of degrees of freedom is 19 (d.f.=n-1 or in this case d.f.=20-1=19). Maybe an example will help to clarify this.

Suppose we have a large bowl containing "peas", each of which has a number between 0 and 10 printed on it. Let's say that I draw two peas (n=2) such that their sum equals 10. With the limiting factor, their sum must equal 10, I have only one opportunity to select a pea randomly. That is to say I have lost one degree of freedom or one opportunity for random selection. For example if I randomly select a pea with a 3 on it, I cannot randomly dip in to get my second pea for it must equal 7, i.e., (10-3). I have lost some freedom -- one degree of freedom in fact.

It is sometimes pointed out that Hypothesis I is evaluated with more degrees of freedom than Hypothesis II, and therefore is preferable. That is getting the cart before the horse. It is true that more degrees of freedom are used with Hypothesis I. When it is assumed that each case comes from a single population every case can be used to estimate the population parameter. When it is assumed that each sample comes from a separate population only the cases within a sample can be used to estimate its population parameters. Since

---

1 Editorial Note: Answers exceeding 500 words are to be discouraged.

degrees of freedom are closely related to the number of cases which can be used for estimating parameters, Hypothesis I does indeed have a greater number of degrees of freedom. This is not justification, however, for using Hypothesis I, when it is not reasonable to assume that the samples represent a single population.

Unstated assumptions. As these formal hypotheses have been stated there are no assumptions left unspecified. If the null hypotheses are abbreviated to "There is no difference between samples," as sometimes is the case, the reader cannot be certain what the proper statistical test should be. If the reader must guess, it is best to guess that the research worker was assuming that random samples have been drawn from a single, normally distributed, infinite population. The four assumptions of the previous statement (random selection, single population, normally distributed population, infinite population) are the most common assumptions for testing hypotheses.

7.5, 7.6

Selecting the Analysis Tool

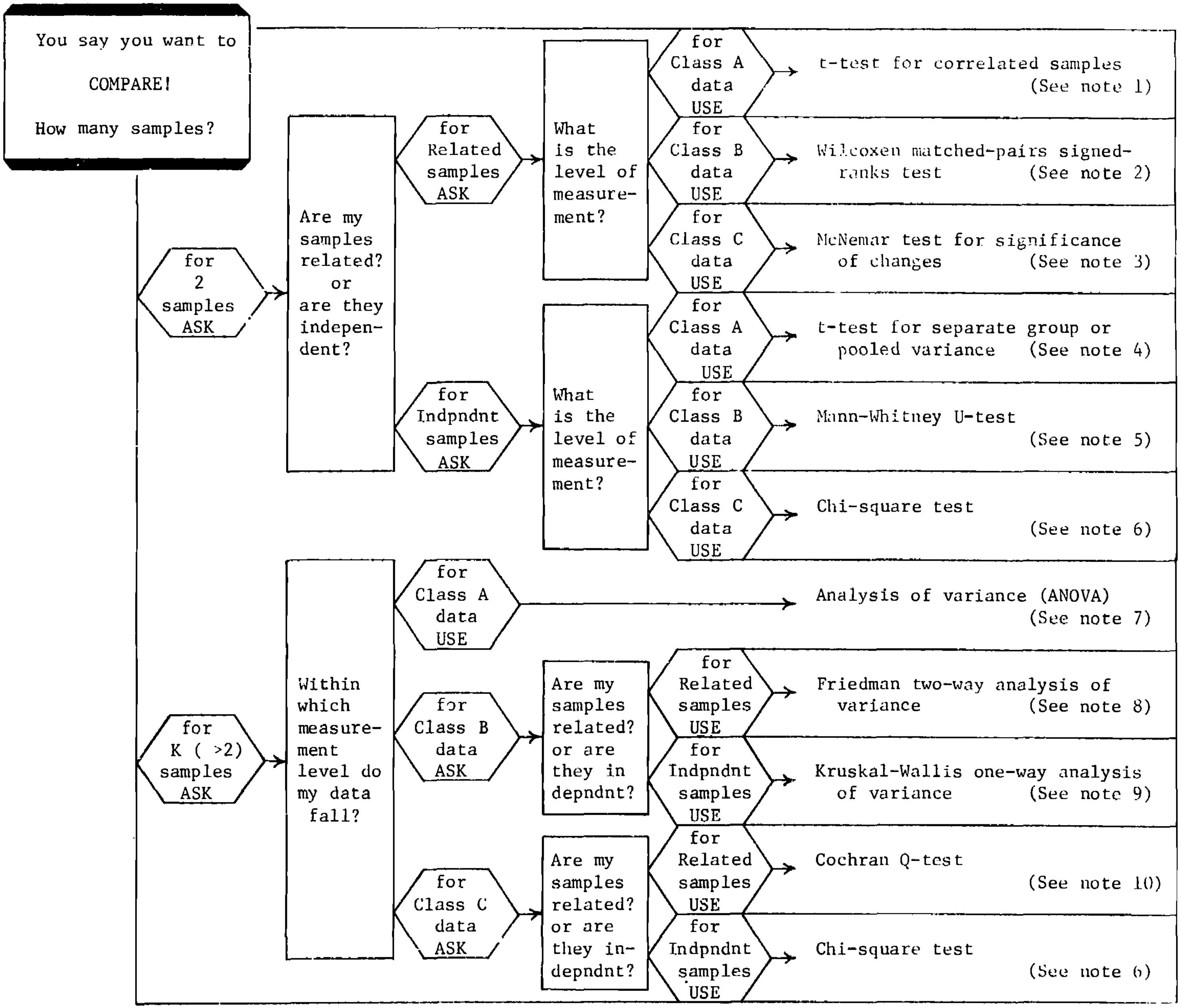
In the previous sections we considered questions which must be asked in order to select appropriate analysis tools, defined the parameters necessary to answer the questions, and discussed some considerations for hypothesis testing. Selection of the appropriate analysis tool is a matter of asking (and answering) a set of relevant questions. Recall that these questions were:

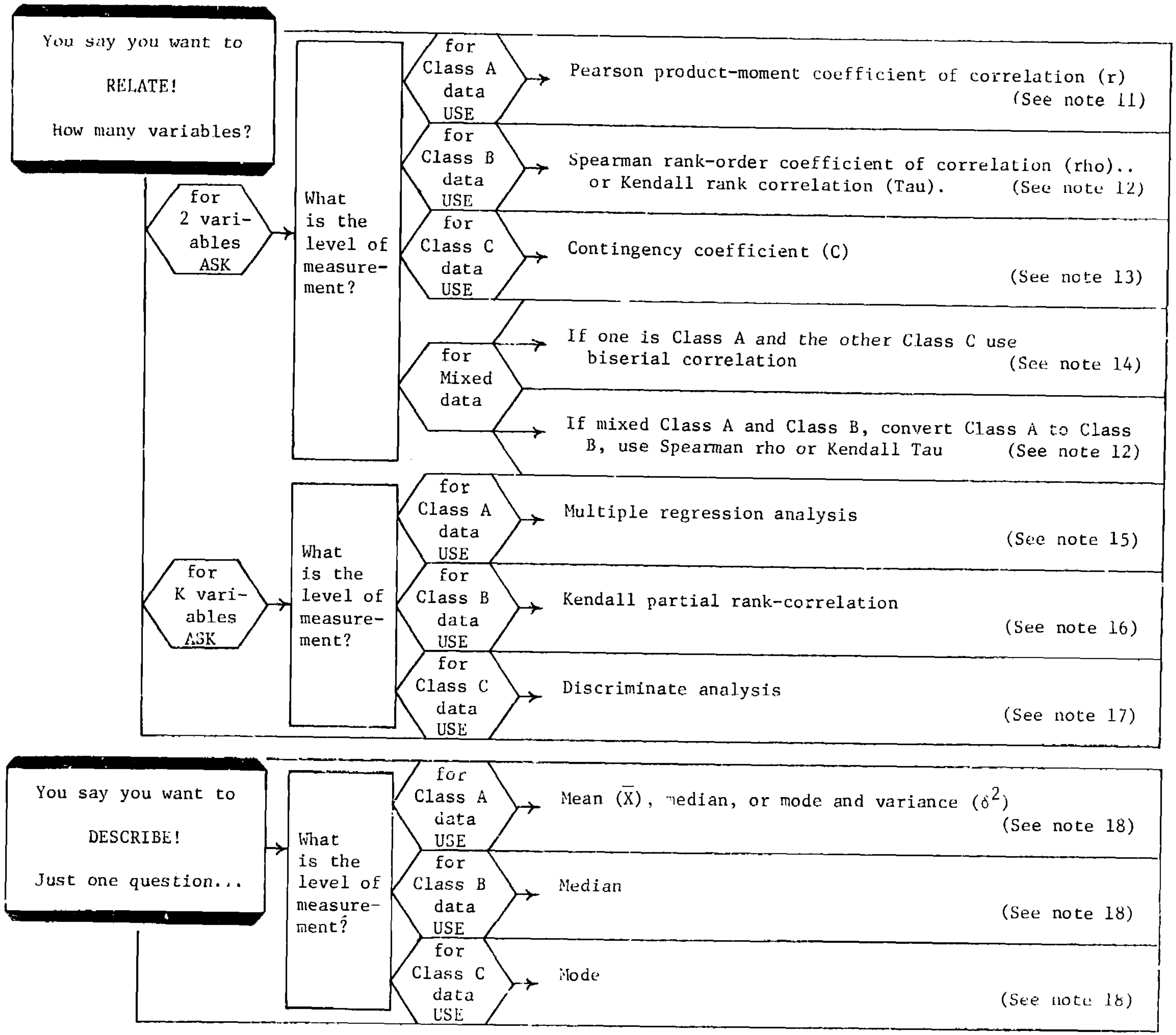
1. Do I want to compare, relate or describe data?
2. Does my data fall within Class A, Class B, or Class C?
3. How many samples do I have or how many variables do I wish to relate?
4. If I'm comparing, are my samples related or independent?

The following table will permit you to identify your analysis tool if you can answer each question presented. It will not tell you how to do the

computation or how to interpret the results. For this you must go to other sources, some of which are indicated in the "note" referred to in the table. It is suggested that you systematically study each "branch" in the table and read the note suggested. This will help you get a feel for some of the techniques available to you. The table, itself, may be used for future practical reference.

7.7


You say you want to
COMPARE!
How many samples?
for 2 samples ASK
Are my samples related? or are they independent?
for Related samples ASK
What is the level of measurement?
for Class A data USE
t-test for correlated samples (See note 1)
for Class B data USE
Wilcoxen matched-pairs signed-ranks test (See note 2)
for Class C data USE
McNemar test for significance of changes (See note 3)
for Indpndnt samples ASK
What is the level of measurement?
for Class A data USE
t-test for separate group or pooled variance (See note 4)
for Class B data USE
Mann-Whitney U-test (See note 5)
for Class C data USE
Chi-square test (See note 6)
for K (>2) samples ASK
Within which measurement level do my data fall?
for Class A data USE
Analysis of variance (ANOVA) (See note 7)
for Class B data ASK
Are my samples related? or are they independent?
for Related samples USE
Friedman two-way analysis of variance (See note 8)
for Indpndnt samples USE
Kruskal-Wallis one-way analysis of variance (See note 9)
for Class C data ASK
Are my samples related? or are they in-depndnt?
for Related samples USE
Cochran Q-test (See note 10)
for Indpndnt samples USE
Chi-square test (See note 6)


You say you want to
RELATE!
How many variables?
for 2 variables ASK
What is the level of measurement?
for Class A data USE
Pearson product-moment coefficient of correlation (r)
(See note 11)
for Class B data USE
Spearman rank-order coefficient of correlation (rho).. or Kendall rank correlation (Tau).
(See note 12)
for Class C data USE
Contingency coefficient (C)
(See note 13)
for Mixed data
If one is Class A and the other Class C use biserial correlation
(See note 14)
If mixed Class A and Class B, convert Class A to Class B, use Spearman rho or Kendall Tau
(See note 12)
for K variables ASK
What is the level of measurement?
for Class A data USE
Multiple regression analysis
(See note 15)
for Class B data USE
Kendall partial rank-correlation
(See note 16)
for Class C data USE
Discriminate analysis
(See note 17)
You say you want to
DESCRIBE!
Just one question...
What is the level of measurement?
for Class A data USE
Mean $(\bar{X})$, median, or mode and variance (δ^2)
(See note 18)
for Class B data USE
Median
(See note 18)
for Class C data USE
Mode
(See note 18)

Note #1

T-test for correlated samples. When you are making comparisons between two samples which are related and for which you have Class A data, the appropriate technique is the t-test for correlated samples. Recall that related samples are defined as those in which each observation in one sample is associated with a specific observation in a second sample. The appropriate analysis technique must, therefore, take this relationship into account, not only in terms of the gross computation involved, but also because the assumptions underlying the comparison are unique.

An example of a study in which related samples obtain would be a situation in which we wish to determine whether any change in creativity can be expected to result from the provision of training in problem solving techniques. Suppose we design our study by measuring pre-training levels of creative ability followed by specific training in problem solving techniques and a post-training measurement of levels of creative ability. Here it is obvious that some relationship between the pre- and post-training performance is to be expected. Assuming that our measures in creative ability yield Class A data, the t-test for correlated samples then becomes appropriate. Computational procedures to be employed with the t-test for correlated samples are reviewed very clearly in Wert, Neidt, and Ahmann (1954, pp. 141-142).

Note #2

Wilcoxen matched-pairs signed-ranks test. When a comparison is to be made between two related samples for which observations of the dependent variable can be identified as Class B data, an appropriate analysis technique is the

Wilcoxen matched-pairs signed-ranks test. Again the assumption is that each observation in one sample is paired specifically with an observation in the second sample. The assumption here is that the investigator can tell which of the two members of the pair have more of the characteristic and he can also determine the magnitude of these differences. Thus, the investigator is able to indicate differences between the pairs of observations and can order these differences in terms of their magnitude. It should be noted that when the above conditions obtain, the researcher is, in fact, converting Class A data to Class B data. In most cases this practice is defensible only if the researcher is unable or unwilling to make the assumption that his Class A data are not normally distributed. An example of the application of this analysis technique and the procedures for analysis are documented in Siegel (1956, pp. 75-83).

Note #3

The McNemar test for the significance of changes. When two related samples are being compared and the level of observed data is Class C, the McNemar test for the significance of changes is appropriate. This particular test is very appropriate for those types of studies in which a person is used as his own control and we are observing Class C or nominal data.

This approach might be used by a school superintendent who wishes to determine the impact a series of television spot announcements might have on voters' attitudes on an up-coming bond issue. Realizing that repeated broadcasts of the spot announcements would prove extremely costly, the superintendent could determine the impact these spot announcements might have by determining the attitudes held by a representative sample of viewers (as revealed by the voters indicating I would support the bond issue or I would vote against the bond issue),

broadcast the spot announcements for one week and then obtain the same attitude information from the same sample. The McNemar test for the significance of changes would be appropriate for him to analyze these data. Other kinds of before and after change situations following this model can utilize this technique. The rationale underlying this technique and the computational procedures to be followed are specified in Siegel (1956, pp. 63-67).

Note #4

T-tests for separate group variance or pooled variance. When comparison is made between two independent samples and the dependent variable is quantified in such a way that it yields Class A data, the t-test for separate group variance or pooled variance is appropriate. In most cases, probably, the t-test for separate group variance will be employed. This is especially true for causal-comparative studies. For experimental studies, pooled variance t-tests are usually more appropriate although the distinction becomes sticky and requires close attention to some specific details. Let's consider a couple of examples.

Let's say that we have devised a new technique for teaching reading that will lead to greater retention of facts. We have a "good" test for measuring retention that yields data at the Class A level. To determine whether or not our new reading instruction is superior to what we have been doing, we randomly assign our class to two groups, each of which will receive instruction under one of the techniques. The appropriate t-test is for pooled variance. The reason is, we have randomly selected two groups (samples) from a single population and have then added an experimental variable. We are trying to determine if the two samples could be expected to have been drawn from a single population. When

the samples have been drawn from a single population it is appropriate for us to "pool" all scores from both samples to compute the standard error of the differences between means, which is the denominator in the computation of the ratio.

When the assumption cannot be made that we have drawn samples from a single, homogeneous randomly distributed, population then we must use other procedures for estimating the standard error of the difference between the means. Let's look at one of these cases.

In another study, let's say that we wish to determine whether sex is a possible determinant of quantitative skills in children. Randomly selecting a sample of boys and a sample of girls from a population of tenth graders (age 15) we again may use a t-test for determining whether the difference in the mean quantitative score obtained by the two samples is in fact different. This time, however, the t-test for separate group variance is appropriate. Reasonably, we are sampling from two separate populations and are assuming that the population means are equal.

The computational procedures to be employed with these t-tests may be found in several sources. Winer (1962) is an excellent source for presenting the rationale. The symbolism used by Winer is somewhat difficult for the statistically naive researcher. A clear exposition of the computational procedure may be found in Wert, Neidt, and Ahmann (1954, pp. 129-137).

Note #5

The Mann-Whitney U-test. When a researcher wishes to compare differences in some characteristic exhibited by two independent samples and has measurement which we would classify as Class B data, the Mann-Whitney U-test is appropriate.

Siegel (1956) describes this as one of the most powerful of the non-parametric tests and a most useful alternative to the t-test when assumptions necessary for employment of the t-tests cannot be reasonably made or when the measurement data are clearly not of the Class A level. It is an extremely useful analysis technique, appropriate for many research situations. The best reference for examples and computational procedures to be employed with the Mann-Whitney U-test maybe found in Siegel (1956, pp. 116-127).

Note #6

The Chi-Square test. When the data falls within the Class C level, the Chi-Square test is appropriate for testing differences between two or more independent samples. Essentially the Chi-Square test is a test of independence as well as difference. Let's look at a situation.

An investigator wished to determine whether college graduates were more likely to support school levies than were non-college graduates. His findings are recorded in the following contingency table:

	Grad	Non-Grad	Total
For	75	148	223
Against	32	111	143
Total	107	259	366

The obtained Chi-Square value of 3.70 was not sufficiently large to reject the hypothesis that voting behavior was independent of graduate status.

In another study the investigator devised five alternative uses for a set of materials, designed to teach a given behavior. After training he noted the

following distribution of those who did and did not exhibit the behavior.

	Method					
	A	B	C	D	E	Total
Did	16	8	8	19	21	72
Did Not	7	6	13	7	10	43
Total	23	14	21	26	31	115

Again the obtained Chi-Square value of 7.58 was insufficiently large to suggest that acquisition of the behavior was better for any given method.

The investigator then wished to know how the trainee's attitudes toward the methods differed. He asked them whether they were positive, negative or neutral in their attitudes toward the method and found the following distribution:

	A	B	C	D	E	Total
Positive	8	5	10	13	18	54
Negative	5	6	5	6	8	30
Neutral	10	3	6	7	5	31
Total	23	14	21	26	31	115

The obtained Chi-Square value was 7.50, again showing attitudes independent of method.

The limitation of Chi-Square is that only two dimensions may be categorized at any one time. Indeed, this limitation holds for all non-parametric techniques. And, while this limitaiton is not serious when comparing two samples, it becomes a very serious limitation when the investigator is attempting to control for or identify the contribution of several independent variables in his study.

Perhaps the best source for further study on the uses of Chi-Square, it's rationale and computational procedures is Siegel (1956). Many other references contain fine chapters for this particular technique. The text by Wert, Neidt, and Ahmann, (1954, pp. 146-171) has devoted a complete chapter to the uses of Chi-Square.

Note #7

Analysis of Variance. A single analysis tool for use in those compartive studies where there is a single dependent variable and one or more independent variables is the analysis of variance tool (ANOVA). This model is applicable for all data that would normally fall in the Class A level of measurement. It may be used any time that assumptions of independent samples drawn from normally distributed populations are met. Additionally, the tool is sufficiently robust to permit the analysis of data which might be derived in causal-comparative studies and, in fact, in that broad classification of studies in which there is a combination of causal-comparative and experimental comparisons.

Again, let's consider a research situation. Let's say that we have developed three specific strategies for teaching arithmetic problem solving skills. We wish to determine which of these three strategies or techniques, is, in general, superior in developing these skills within the fourth grade curriculum. From among the fourth grades in our school system, we select nine classrooms. The dependent variable, computational problem solving skill, is adequately measured with a testing device that yields Class A data. Essentially this is an experimental study in that we have identified samples of students drawn from a single population and have experimentally exposed each of our samples to a distinct experience. Upon initial analysis of our data we find

that there is insufficient evidence to reject the hypothesis that no differences in criterion performance obtains with reference to the three teaching techniques employed.

Based on evidence that boys and girls tend to have different levels of quantitative ability, we decided to reanalyze our data adding a second independent variable, sex. To accomplish this we separate the scores in each of our samples so that now instead of having three samples we have six samples, that is, two samples for each technique, one composed of boys and the other of girls. We now have a combination of an experimental study and a causal-comparative study in that the second independent variable, sex, is not under the control of the experimenter but may be highly useful in studying the effects of the experimental variable, instructional technique. With the analysis of variance model the addition of the second independent variable permits us to look closely at the first independent variable or main effect (instructional strategy) and compare the three instructional strategies. In addition, we can compare the performance of boys and girls independent of the instructional strategy to which they were exposed. Additionally, a third comparision can be made and that is of the interaction of the two main effects, sex and instructional strategy. Even when we are not particularly interested in the additional independent variable or variables and their interaction with the independent variable of primary concern, the analysis of variance model permits us to get a clear look at the main effect we are particularly interested in.

In the above example, let's consider for a moment that we were interested in more than the main effect for instructional strategy. Let's say that in the

ERIC Full Text Provided by ERIC

design of our experiment we had included main effects for instructional strategy, sex, ability level, and ethnic background. An experiment having four main effects has a potential of identifying fourteen different comparisons which can be made, four comparisons of main effects and ten comparisons of various interactions of these main effects. Given that we had three distinct instructional strategies, three ability levels that we were interested in, and three different ethnic origins, plus the two sexes, this experiment would have fifty-four groups of scores or in our previous terminology, fifty-four samples involved in the making of comparisons.

For each comparison that yields statistically significant differences it is essential that we identify where these differences primarily obtain. For example, we may find that the comparison of instructional strategy yields significant differences. We should try to determine if this difference is a function of one of the strategies being superior to the other two with the other two strategies being equally effective; or whether these were actually significant differences between all strategies. There are several techniques for studying these kinds of differences which obtain in analysis of variance experiments. Winer (1962) is one of the most complete references on the analysis of variance model and is very specific in describing the procedures for further analysis of significant results which obtain in such experiments.

The following references would be extremely helpful for further study of the analysis of variance model: Wert, Neidt, and Ahmann (1954), Winer (1962), Tukey and Green (1960), Millman and Glass (1968). As a caution, I would suggest that the Winer and the Tukey and Green references be used only when you have had considerable experience in using the analysis of variance model and have begun to read quite widely in this area.

Note #8

The Friedman Two-Way Analysis of Variance by Ranks. When you have two or more samples that are related and have data that falls at the Class B level the Friedman two-way analysis of variance technique is appropriate. Since in educational research situations we seldom find a situation in which we have related samples, this technique is more adaptive to animal laboratory research and agricultural research situations.

Let's try to adapt it to a situation involving human subjects in a somewhat educationally oriented situation. Let's assume we wish to determine which of three high jumping techniques (the scissors, the western roll, or the Fosbury Flop) produced the best performance after a short period of training (8 hours in instruction). Since we are working hypothetically in this situation anyway let's consider that we identify a sample of twenty sets of identical male triplets all of whom are twelve years of age. We randomly assign one triplet of each set to one of the jumping style training programs. Following eight hours of instruction we ask each triplet to engage in a high jumping contest independently so that we may determine the maximum height each triplet can jump. We may then order the performance of each set of triplets. Using these ordinal data we may then compare the performances achieved by triplets performing under each of the jumping conditions.

Again, Siegel (1956, pp. 166-172) is the appropriate reference for this analysis technique. The computational steps are clear and rationale and and examples are provided.

Note #9

The Kruskal-Wallis one-way analysis of variance. When comparing several independent samples, observing the dependent variable with a measure that yields

ERIC

Class B data, the Kruskal-Wallis one-way analysis of variance technique is appropriate. This technique is adaptable to a wide variety of research situations. The computational procedures are quite simple, and in those situations where the researcher is unwilling to make the assumptions about the normal distribution of scores for any other distribution of scores for that fact, he may convert apparently Class A data to rank order data and employ this technique.

Let's consider again a research situation in which the investigator is interested in the differences in attitudes towards campus disturbances held by the following types of people: college administrators, college faculty, college students, chamber of commerce members, and parents of high school students. To measure the attitudes toward campus disorders the investigator develops a semantic differential scale and administers this scale to representatives of the above mentioned groups. If you were to convert the scores obtained on the semantic differentials to ordinal data he could then employ this analysis technique.

This technique, as well as most other non-parametric techniques is well described in Siegel (1956, pp. 184-194).

Note #10

The Cochran Q-test. When utilizing Class C data for comparing more than two related samples the Cochran Q-test is appropriate. Siegel (1956) uses the following case as an example for which this analysis technique would be appropriate;

> "Suppose we were interested in the influence of interviewer friendliness upon housewifes' responses in an opinion survey. We might train an interviewer to conduct three kinds of interviews: Interview 1, showing interest, friendliness, and enthusiasm; Interview 2, showing formality, reserve, and courtesy; Interview 3 showing

ERIC

disinterest, abruptness, and harsh formality. The interviewer would be assigned to interview three groups of eighteen houses, and told to use interview 1 with one group, interview 2 with another, and interview 3 with a third. That is, we would have obtained eighteen sets of housewives with three matched housewives (equated on relevant variables in each set). For each set, the three members would be randomly assigned to the three conditions (type of interviews). Thus, we would have three matched samples (K=3) with eighteen members in each (N=18). We could then test whether the gross difference between the three styles of interviews influenced the number of "yes" responses given to a particular item by the three matched groups."

The rationale and computational procedure is well described in Siegel on pages 161-166.

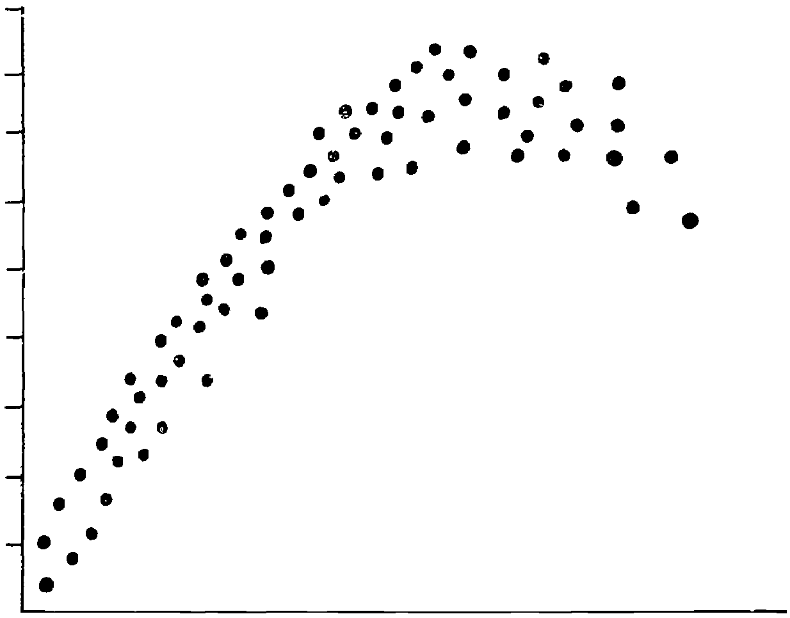
Note #11

The Pearson product-moment coefficient of correlation (r). When an investigator wishes to determine the extent of relationship between two characteristics, measures of both of which fall within the Class A level, the Pearson product-moment coefficient of correlation (r) is the appropriate tool in most cases. The most crucial limiting factor, which might alter the use of this approach is the colinearity of the two variables. If the relationship of variables has been studied previously and evidence of linearity is present, use of r is generally safe. Without such evidence the careful investigator should first produce a scatterplot of data, studying the plot to determine any obvious deviations from linearity. If the scatterplot provides rather clear evidence of curvilinearity, alternative correlational techniques should be used (Wert, et al., 1956).

An investigator wished to determine the relationship between quantitative aptitudes of mathematics instructors and achievement of their students. His first analysis of this relationship indicated little relationship (r=16). The scatterplot for his data is shown on the following page.

Class

Achievement

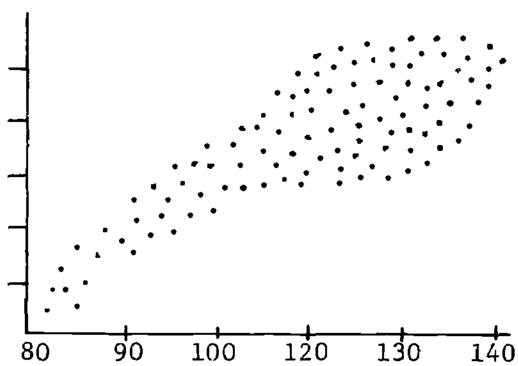


Instructor Aptitude

Would he be justified in concluding that little relationship was present?

Another investigator, in reviewing creativity research, found the statement that a rather clear relationship between creative behavior existed for those with I.Q.'s of 120 or lower but beyond that point little relationship was present. The following scatterplot might illustrate such a finding.

Creativity



Intelligence (I.Q.)

Note #12

The Spearman rank-order coefficient of correlation (Rho) or the Kendall rank-correlation (Tau). When data variables yield Class B data, the Spearman (Rho) is the equivalent to r. It requires that data variables be ordered and is sufficiently robust to accept ordered ties. Excessive numbers of ties tend to attenuate Rho and should, therefore, be adjusted. Procedures for adjusting for excessive numbers of ties are well specified in most references describing these techniques. An alternative method to the Spearman rank-order coefficient of correlation is the Kendall (Tau). Tau is a less frequently used statistic but is acceptable and to some researchers is considered desirable. References to Rho are made in Wert, Neidt and Ahmann (1954, pp. 87-89) and Siegel (1956, pp. 202-213. Siegel aslo contains reference to the Kendall-Tau statistic covering this on pp. 213-223).

Note #13

The contingency coefficient. When both variables that are being related are a dichotomous Class C variable, an extension of the Chi Square Technique yields a measure of relationship C. This, for example, could be applied to the previously stated Chi Square example of the difference of voting behavior of college graduates and non-graduates. (See Note #6) In that analysis we found that we could not reject a hypothesis of independent voting behavior and graduate status. We could further compute a contingency coefficient (C) to indicate the degree of relationship between the two variables. If we were to do that with the data previously recorded we would find that C equals .14. The procedures for computation and the rationale for it are contained in Siegel on pp. 196-202.

Note #14

Bi-Serial correlation. Occasionally an investigator wishes to determine the relationship between two characteristics, one of which is Class A and the other a dichotomous Class C characteristic. The appropriate descriptor of this relationship is the bi-serial r. An example might be the relationship between intelligence (Class A variable) of unemployed adults and their political party preference (Democrat or Republican, a Class C variable). Most references to this technique concern themselves only with those cases where the Class C variable is dichotomous. The technique may be expanded along the Class C variable, however, to include 3, 4, 5, or more categories. It is a direct extension of the Pearson product-moment coefficient of correlation. Bi-serial correlation is based upon the assumption that the dichotomous characteristic is not a true dichotomy, but is instead a normally distributed variable that can be observed no more accurately than in dichotomous terms. When operating with a true dichotomy, for example sex, a point bi-serial correlation is appropriate. Computational procedures for bi-serial correlation techniques are found in Wert, Neidt, and Ahmann, (1954).

Note #15

Multiple regression analysis. Many studies wish to determine the relationship between one variable and several other variables. The general form for such studies is a multiple regression model. It requires measures of all subjects on all variables and demands that the single dependent variable and most independent variables be of the Class A level. The regression model fits several types of studies. It yields a multiple coefficient of correlation, R; a prediction equation, partial coefficients of correlation (for example the prediction of grade point averages on the basis of motivation scores controlling

for intelligence) and finally permits the analysis of the relative contributions of each independent variable to the prediction of the dependent variable. The investigator is cautioned aginst using large numbers of independent variables in a study with this model. When the number of independent variables (predictors) equals or exceeds the number of subjects, the solution is overdetermined and R always is equal to unity. An accepted rule of thumb is to limit the number of predictors (k) so that k is equal to or less than $\frac{n}{2}$, where n is the number of subjects.

In the usual application of this model the relationship between the predictors and dependent variables is assumed to be linear. The model can be extended to non-linear cases quite readily. The extension to non-linear cases requires that the investigator make some assumption relative to the general shape of the curve that will fit his data, know the general form for the curve (remember your analytic geometry) and apply some relatively simple concepts of the calculus. These techniques are described in Wert, Neidt and Ahmann (1954) on pp. 226-255, and pp. 282-293.

Note #16

The Kendall partial rank correlation. It was mentioned previously in Note #15 that it is possible to determine the relationship between two variables holding the contribution of other variables constant. Siegel states:

> "In designing an experiment, one has the alternative of either introducing experimental controls in order to eliminate the influence of a third variable or using statistical methods to eliminate its influence. For example, one may wish to study the relation between memorization ability and ability to solve certain sorts of problems. Both of these skills may be related to intelligence; therefore, in order to determine their correct relation to each other the influence of differences in intelligence must be controlled. To effect experimental control, we might choose subjects with equal intelligence. But, if experimental controls are not feasible, then statistical controls can be applied."

Partial correlation is such a statistical control. When the data are of the Class B level the Kendall partial rank correlation technique is appropriate. This technique simply provides an interpretable coefficient of correlation, that is, an index that describes the amount of relationship that exists. The technique does not permit the development of prediction equations nor does it permit further analysis of the proportionate contributions yielded by each variable. These types of procedures are limited only to the more powerful parametric model described in the previous note. The Kendall procedure is described in Siegel (1956, pp. 223-229).

Note #17

Discriminant analysis. An extension of the multiple regression model yields a model for multiple prediction of a dichotomous Class C variable. The same caution exists for the discriminant analysis model as does for multiple regression. The same types of studies may be completed. The only difference is that the dependent variable for discriminant analysis is dichotomous rather than continuous. This model is especially useful for studies where we wish to predict whether a person will be a success or failure, will remain in school or will drop-out, or has a pattern of abilities best suited for one type of work as opposed to another. The model may be extended to those situations in which the dependent variable is observed in more than the two dichotomous categories. The technique is described in Wert, Neidt and Ahmann (1954) pp. 256-281.

Note #18

Measures of central tendency and dispersion. Descriptive studies are normally limited to utilization of statistics which yield two types of

information. The first is that of identifying some score which is representative of the scores being described. This statistic can be either the mean, median, or mode. For Class A data all are appropriate although the mean is probably most descriptive. The mean is, however, an inappropriate statistic to use for Class B or Class C data. In most cases, Class B data are best described using median and for all cases Class C data can be described only with the mode. These statistics (mean, median, and mode) are called measures of central tendency in that they represent some typical score about which all of the other scores cluster.

In many instances the measure of central tendency can be more readily interpreted when we provide a measure of dispersion which describes how extensively the scores which we have observed are scattered. Measures of dispersion are appropriate almost exclusively for Class A data. The most sophisticated measure of dispersion is the standard deviation. Less sophisticated but often useful measures of dispersion are the range, and the semi-interquartile range. References to these measures of central tendency and dispersion may be found in almost all standard statistical texts. Wert, Neidt, and Ahmann devote Chapters 2, 3, and 4 to these concepts. That is as good a place to start as any.

ERIC

References

Diamond, S. Information and error. New York: Basic Books, 1957.

This is one of the most readable introductory texts to statistical analysis. While its coverage of analysis techniques is limited, this reference does a fine job of providing the underlying rationale behind those techniques covered, and especially as they deal with simple statistical inference. The symbolism used in the text, while bulky for the somewhat sophisticated researcher, is well developed and probably very useful for the naive statistician.

Green, D. F., & Tukey, J.W. Complex analyses of variance: general problems. Psychometrika, 1960, 25, 127-152.

Millman, J., & Glass, G. V. Rules of thumb for writing the ANOVA tables. Journal of Educational Measurement, 1967, 4, 41-52.

The analysis of variance model is extremely complex but at the same time a very powerful analysis tool. The model has very many variations. Each variation has its own peculiar requirements. For the researcher who has, through training and experience, become quite familiar with analysis of variance techniques, this reference will become very useful. This reference when used appropriately will effectively eliminate the necessity for perusing a large number of statistical texts in an effort to find an example that matches his case precisely.

Senders, Virginia L. Measurement and statistics. New York: Oxford University Press, 1958.

This reference is especially useful for gaining an understanding of the various levels of measurement, i.e., nominal, ordinal, interval, and ratio. Senders begins with the definition of the scales and then bases the development of statistical techniques on the characteristics of these scales. No reference other than the Stevens' chapter in the handbook of Experimental Psychology treats the measurement scales as adequately as does Senders.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

This reference is without peer for its complete treatment of the nonparametric statistical analysis techniques. This is a "must" reference for any researcher who must work with data requiring these techniques and who must be responsible for his own analyses.

Green, D. F., & Tukey, J. W. Complex analyses of variance: general problems. Psychometrika, 1960, 25, 127-152.

This reference in combination with the Millman-Glass reference can prepare the researcher to such a level that he will avoid many commonly made errors in the usage of the analysis of variance.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, 1954.

While this reference is not recognized as one of the greater ones in the field of statistical analysis, it is written for the educator who needs the tools but does not require the depth of understanding into the rationale on which the tools are based. The coverage of analysis techniques made by this reference is as good as found in any reference. This is a good text for the beginner.

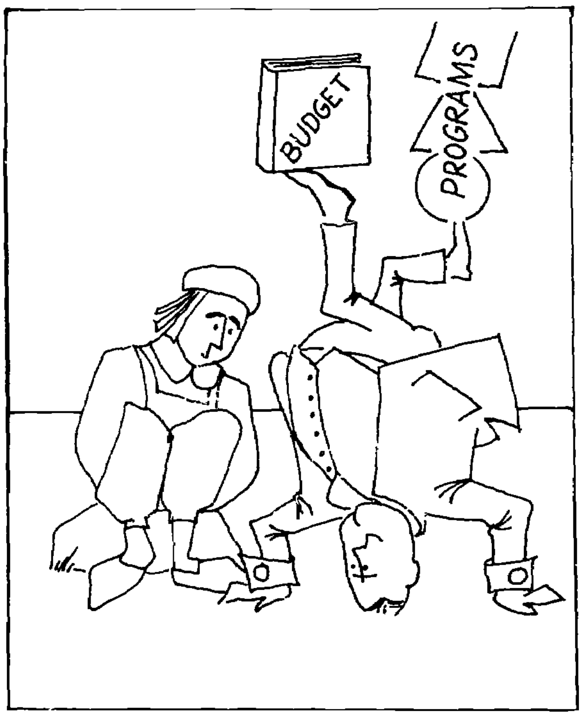
Winer, B. J. Statistical principals in experimental design. New York: McGraw-Hill, 1962.

This is an extremely fine reference for statistical analysis. The coverage is broad and the development of rationale is excellent. The reference is especially valuable for the thorough utilization of the analysis of variance model.

Section VIII: Administration

Editorial Foreword

In this chapter Jack Edling presents a brief and highly personal view on running a research organization. The editor can personally testify that Jack runs a good organization himself, or did until recently. I worked for him for several years before he decided to quit and become an investigator again. No causal relationship between my employment and his abdication has ever been demonstrated.



"You are old, Father William
One would hardly suppose
That your brain was as agile as ever.
Yet you balance ten programs
On the end of your toes.
What made you so·awfully clever?"

"I've done this for you for seven
years, and that is enough.
Now, don't give yourself management
airs.
Do you think I will listen all day
to such stuff?
Be off or I'll kick you downstairs!"

This short chapter flows along so easily that it is doubtful if the reader needs any guide. Jack's description of the types of "researchers" he has encountered is worth your admission price. The humorous, but slightly jaundiced view fails to reflect the esprit that Jack Edling was always able to bring out in staff--and which may be as crucial an attribute of management as any mentioned.

VIII

ERIC

The editor's principal regret about this chapter is that Jack could not be forced or cajoled into writing more.

VIII

ERIC

On the Administration of Research Agencies..

By Jack Edling

Introduction

In the novel Arrowsmith by Sinclair Lewis (1925) there is a description of a research laboratory which seems too true to be fiction. One episode relates how Gottlieb, the true scientist, and his bright young associates spend more time overcoming the deficiencies of poor administration than they do planning and conducting research. They all long for the day when Gottlieb will gain control of the laboratory. Finally, after years of agonizing frustration, he is made director. The story continues.. "So Max Gottlieb took charge of McGurk Institute of Biology, and in a month that institute became a shambles."

The problems associated with research administration are related to the dilemma of too little versus too much direction and organization. Most researchers want freedom to do what their brains tell them to do, and they want the kind of organization that will permit and support their demands. This sounds reasonable, but in actual practice it takes systematic procedures to support the demands of even a small number of researchers. So, some type of organization is created to provide desired services and an individual is placed in charge. Now the problems begin. If the person in charge is primarily an administrator, then systems are devised and supervised which permit rational, orderly, and economical support services. But, the researcher is not interested in the mechanics of the system, giving timely notice of demands, adhering to schedules, priorities, econo-

ERIC

mics, reports, etc. etc. He merely wants what he wants when he wants it. And he should have it - that is why the system was established. However, the typical administrator is more concerned with the system, its accountability, responsibility, fiscal integrity, etc. etc., than he is in what he considers to be the whim of some prima donna researcher. And he should have his concerns in these areas for these are the purposes for which he was engaged.

The typical response of researchers to "beat the system" is to get one of "their own" in the key administrative position. This tactic places them in a most favorable posture whenever there is a confrontation with support personnel of any type - be they technical, fiscal, logistical, or whatever. When "their man" is in control, the researcher can commit no evil - the system be damned. But, unfortunately, the victory of the researcher is short-lived because in such situations the technical, fiscal, and logistical services which he so desperately needs cease to function, or they function inadequately. When administrative services do not receive the attention and consideration they deserve and when they are not directed by a capable, sympathetic, and unbiased administrator, problems develop so rapidly that effective research soon becomes practically impossible.

The person who finds himself in the key administrative position in an educational research agency has more to concern himself with than settling disputes between or among researchers and support staff. His principal activities relate to personnel, organization, policy, administration, and finally facilities, including equipment and materials. Each of these

ERIC

topics require more elaboration than is provided herein, but the discussion may be useful for those just getting started.

Personnel

Experience has clearly demonstrated that an educational research agency can operate fairly successfully with practically any organization, policies, administration, facilities etc., if it has "good" people, but, that no organization, policy, administration, or facility can operate successfully without "good" people. Now, the only question is, who is a "good" person - or in the present instance, what are the characteristics of "good" and "poor" researchers and research support personnel for conducting educational research?

It is easiest, perhaps, to list types of educational researchers as they actually arrive in the typical agency. They all arrive with excellent references and vita, but they may soon be classified as being dominated by one or more of the following traits:

1. Most common is the genius. This type, usually fresh from the Ph.D. program or from a series of one year appointments, often does have a great deal of information or expertise in one or more tasks of the educational researcher. He is more like a highly skilled technician than he is like a scientist. He has answers - not questions. His interests are usually fairly narrow. He is often uncooperative and refuses to "waste his time" on problems which appear ambiguous or for which no clear-cut methodology appears "obvious". He has contempt for most of his potential colleagues because they, so obviously, are unqualified, etc.

ERIC

2. The "broadly-based" researcher. This type, usually highly verbal, can pontificate on any subject and produce on none. He has no expertise or skills involved in research. When his proposals are rejected, it is the biased nature of the reviewers. When his products are criticized, it is the assignment which was not sufficiently specified. When he is asked to resign, it is "time" - he was not given sufficient of it to demonstrate his many capabilities.

3. The conceptualizer. This type asks relevant questions, but he can never ask them with sufficient completeness to satisfy his compulsiveness for the perfect piece of research. So, he continues forever to revise and refine every concept with which he is confronted. He is always on the verge of initiating a proposal, an instrument, some data collection, an analysis etc., but each requires more thought - more conceptualization. Unfortunately, nothing ever happens.

4. The doer. This is the "action" researcher. Collecting data and running analyses are no problem. The fact that there is no research question related to his activity is of no real concern because he has rationalized his present behavior as merely "pilot" work for the better controlled and more adequate work which he will do when he has a better "feel" for all the "parameters and variables". He never gets around to anything but the "quick and dirty" because "thinking" is only a rationalization (an excuse) to avoid the real work of instrument building, data collection, statistical analyses, etc. So, much is accomplished in the way of activities and products, none with much meaning, direction or utility.

5. The "researcher-plus". This type is essentially a "good" researcher. He has all the qualifications necessary to make a substantial contribution to a research effort. Unfortunately, he has some additional characteristics not related to the research effort which negate all his desirable qualities. He comes in several subspecies. One is a political animal who has such power needs that in addition to doing research he is undercutting all his colleagues, plotting the "revolution" that will make him "supreme commander". Another is the psychopath, who takes anything from anyone, exploits or usurps anyone's budget, doesn't bother with personal obligations, considers any act of courtesy, compliance, or cooperation a rather stupid act, appropriate for subordinates but a detail well beneath his consideration and concern. Another is the egotist. In all instances, regardless of the trait, it is a personality flaw so obnoxious that in one way or another the individual is more of a liability than an asset. His administrators and colleagues spend more time avoiding, or excusing him than they do researching with him. None of us is perfect, and we must learn to tolerate idiosyncracies (I suppose) but there is a limit even to the most tolerant among us. The disruptive qualities of some men exceed their contributions to a research staff to such a degree that they must be considered lone researchers - not part of a team.

While the list may be extended, the point may now be made. It is this. Whether the man has the ability to contribute to research, but won't; or wants to, but can't, or is so insecure, compulsive, so disruptive as a person that he distracts from the total effort, in any instance he is something less than a "good" researcher. Some research administra-

ERIC Full Text Provided by ERIC

tors believe that they can either tolerate, balance off, or somehow change these liabilities into assets. My recommendation is to encourage the types of researchers described above to either practice alone or join those research administrators who practice a "hopeful" type of research administration. My experience leads me to believe that an organization of weak, deficient, and incomplete human beings is a weak and deficient organization. An organization of intellectually strong and complete men tends to make a more effective and more complete organization. I recommend giving up the activity of remaking or tolerating adult neurotics and concentrating on the task of giving new or additional technical competence to men who are essentially well, by mental health standards, i.e., inquisitive, secure, bright, cooperative, and who desire to produce new information. If some special competence is needed for a specific task, consider hiring a consultant or a genius (almost any type can be tolerated for a few days).

In personnel acquisition, obviously, specifications for the type of person desired are essential. But, too frequently paper qualifications are substituted for basic intellectual and personality characteristics. The "good" researcher who will work in a team effort must have technical competence <u>and</u> traits which make him compatible with the balance of the team. I have been impressed that even small amounts of creativity or insight require vast implementation efforts. Every research organization needs sufficient creativity and insight to give it direction, but many, if not most, are so constipated with "direction" there is no "movement". The other extreme is the diarrhea of research (dissertation type) which lacks not only direction, but several other essential ingredients.

Support staffs consist of specialists such as electronic technicians for building instrumentation, photographers, artists, computer programmers, reference librarians, data processors, maintenance men, business managers, fiscal officers, typists, accountants, public information types and a list of others. These people must not only be creative but understand the idiosyncracies of researchers and think like team players. As to the method to find those "good-bright, etc. people". I have never found a method except "trial and fire." Everyone's credentials are great. There is no one in the eyes of his major professor who does not have some excellent qualities, and the credentials always concentrate on strengths and overlook weaknesses. My advice is to have at least two colleagues recommend the person to be hired and give the person a fair trail (six (6) months maximum). Most people demonstrate their potentialities in a realtively short period. If they demonstrate the characteristics described above, my advice is to find them another position immediately (1st year). The longer a person is around the more difficult it is to get rid of him. Many if not most research shops are loaded with types one doesn't need. When two or more colleagues have been involved in the original approval no one need take complete responsibility for having made the initial mistake.

Organization

Assuming that men of technical competence and good will can be recruited, how then can they best be organized to produce new knowledge and/or products?

A basic consideration is funding source. If the research organization

is autonomous and obtains funding from a number of sources and is answerable to a given source only for a specific project, then one set of conditions prevail.

If the research organization is an agency of a public or private institution, another set of conditions prevail. In either instance a policy board or directorate is essential (at the present time) to provide input and guidance to a coordinator or administrator on operational matters.

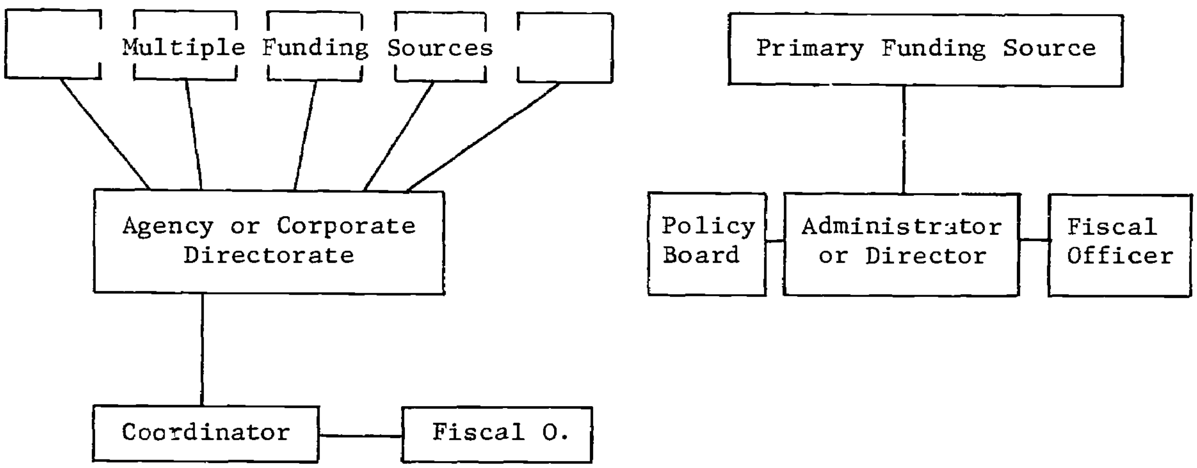


Figure 1. Organization to Provide Overall Coordination and Direction

A fiscal officer is essential (at the present time) to provide input and guidance on fiscal matters.

The directorate or policy board should represent every super or subordinate agency, group, or individual who has useful input for a coordinator or administrator, or who might benefit from participating in the ouput from the deliberations of such a group(s). The fiscal officer should participate in selected activities of such policy groups. These relationships are illustrated in Figure 1.

Groups of researchers having like interests or working on related problems should probably form some sub-organization to provide for internal coordination and additional relevant input to staff. They too require both operational and fiscal input. The administration and support staff also requires coordination and administration and should be given administrative identity comparable to research units. These relationships are illustrated in Figure 2.

Policy

Policies take time to develop, but in the long run can save time, misunderstandings, and ill-will. Educational research agencies should consider developing policies on topics such as:

a. How much of an individual researcher's time can be devoted to "extramural" activities such as personal consulting, writing, teaching, etc?

b. How much, when, and what regarding leave or vacation time? Sometimes researchers are known to "take-off" for weeks. When a man says he can't "create" on schedule, that he needs isolated "think-time" and his office isn't the place - what is the policy?

c. Relationship, care and treatment of human subjects (children, young people and adults).

d. Use of cars, equipment and related items of capital outlay.

e. Attendance at reguarly scheduled staff meetings.

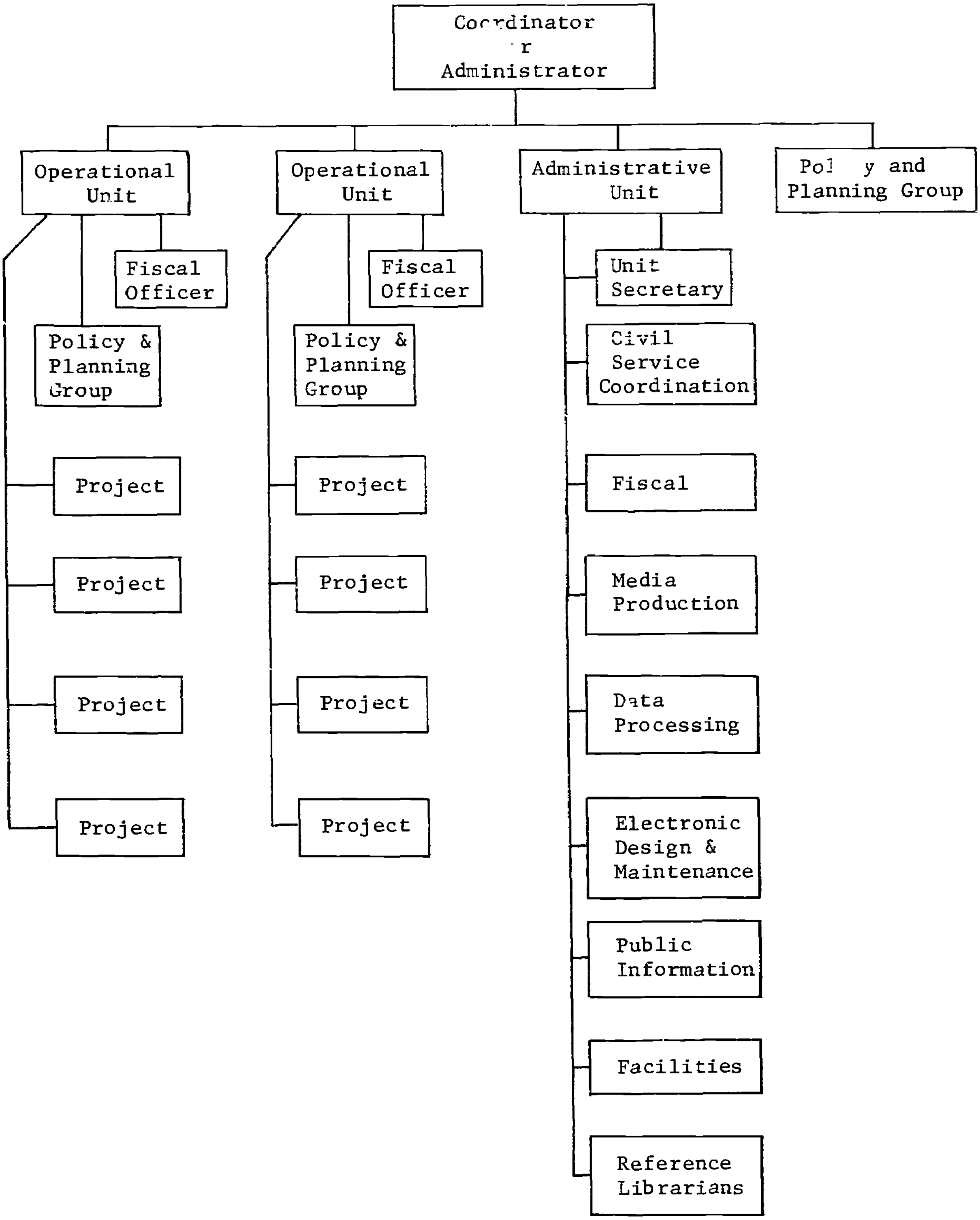


Figure 2. Organization to Provide for Administration of Multiple Projects

ERIC

f. Representing the institution and speaking for it (committing it) to other educational agencies.

g. Control of expenditures, e.g. telephone, mileage, materials and other budgeted items.

h. Advancement and salary increases.

Perhaps more important are operational policies such as the idea that each individual has 3 continuing responsibilities, i.e., 1) to learn to perform his responsibilities and to continually become more proficient, 2) to learn the responsibilities of the person to whom he reports and be prepared to assume his duties at any time, and 3) to prepare his subordinates or those who report to him to assume his responsibilities at any time.

Administration

The ultimate test of an effective organization is its functioning when the key leadership is absent. If the administrator's presence is required for the organization to function, he hasn't done his job. The administrator's role is to analyze what must be done into appropriate subtasks which will assure attainment of ultimate objectives. Organize the manpower, physical, and financial resources to get those tasks accomplished. Deputize the appropriate personnel to perform the various subtasks. Supervise each task to determine whether it has been accomplished adequately and if not, to take appropriate action to assure that it is accomplished.

There should always be back-up personnel (a reserve) for every task and as soon as a situation requires the commitment of the back-up

person, the administrator's first task is to identify and initiate training for a new back-up person.

In addition to basic principles of administration (such as those listed above) some consideration should be given to:

(a) adequate involvement of all those who have administrative responsibilities;

(b) the degree of decentralization and autonomy that can be given subordinate units;

(c) the kinds of decisions that various positions or offices are authorized to make, e.g., can a position only recommend or inform others of decisions previously made, or approve or veto proposed decisions, or actually authorize action;

(d) the total number of persons reporting to a superior;

(e) job descriptions;

(f) orientation and in-service training;

(g) individual capacity and work load;

(h) regularly scheduled staff meetings;

(i) channels of communication;

(j) security of information, confidential data, etc.;

(k) frequent and accurate fiscal summaries to project directors;

(l) bonding staff members who handle money;

(m) development of long-range goals;

(n) working relationships with other agencies;

(o) provisions for staff morale, e.g., working space, coffee, social events;

ERIC

(p) evaluation of progress toward goals and appropriate feedback arrangements.

Facilities
(Including Materials & Equipment)

The administrator is concerned both with the acquisition and utilization of space. In planning an educational research facility consideration should be given to requirements such as the following:

1. Administrative areas

   (a) Allowance for circulation and general services

   (b) Reception area

      1). Control center

      2). Coat room

      3). Lavatories

   (c) Administrative services

      1). Reception

      2). Secretarial space

      3). Director's office

      4). Storage rooms

      5). Staff coat room

2. Research areas

   (a) Allowance for circulation and general services

   (b) Office areas

      1). Lobby

      2). Receptionist-secretarial space

      3). Staff offices

ERIC

4). Conference room

5). Office supplies storage

(c) Laboratory areas

1). Research laboratory for each program

2). Auxiliary spaces for each program

(d) General Services

1). Coat rooms, other closets

2). Lavatories

3). Mechanical and electrical equipment spaces

3. Training and demonstration areas

(a) Demonstration amphitheater

1). Demonstration area

2). Observation area

3). Control room

(b) System (materials and equipment) examination area

(c) Techniques laboratory

1). Laboratory operations space

2). Training laboratories

4. Services Center

(a) Apparatus room

(b) Media maintenance room

(c) Media distribution space

(d) Media services office

(e) Experience recording section

1). Recording studio

ERIC

2). Control and videotape recording room

3). Preparation room

(f) Production editing section

1). Production studio

2). Graphics studio

3). Dark room

4). Editing room

In the preparation of budgets special attention should be given to materials, equipment and miscellaneous items of administrative concern such as:

A. Materials

1. To produce instructional materials (film, tapes, lettering guides, etc.)

2. Office supplies

3. Dissemination materials - brochures

4. Instrumentation (electronic and mechanical supplies)

B. Equipment

1. Typewriters

2. Calculators

3. Computers

4. Recorders, dictation

5. Television

6. Projection and camera

7. Reproduction

C. Miscellaneous items

1. Communication (telephone, postage, publicity)

ERIC

2. Travel

3. Consultants

4. Services

5. Payroll and payroll benefits

Having spent a good part of my life as a research administrator let me leave you researchers who have ideas of building an environment for your colleagues with these parting words. Don't do it. - The joys and rewards of research are the real pay-off for the person who has invested a major part of his life in acquiring research competence. It seems to me that research administration should become a specialty for those men who have high aptitude for administration and a desire for these kinds of activities.

A good course in the philosophy of science, some close associations with good researchers, and ample experience in administrative procedures should constitute the curriculum for a research administrator. This program, to my knowledge, does not exist at present. One is badly needed. Research administrators are urgently needed - as are good researchers. We shouldn't take the latter and convert them into reluctant administrators. Rather, we should take eager administrators and train them in the requirements of research.

ERIC Full Text Provided by ERIC

Section IX: Proposal Writing

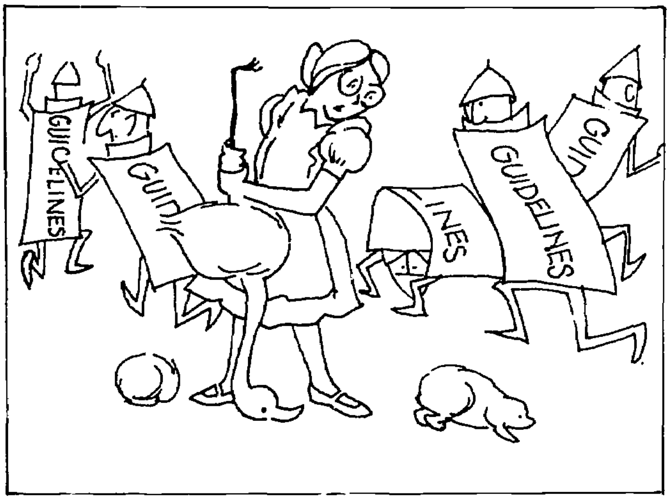
Editorial Foreword

The chapter on proposal writing by Catherine Kielsmeier and the editor was originally developed under a grant from the Cooperative Research and Development Program of the Bureau of Research. However, the present version of the chapter is now a component in a more complete instructional system on proposal writing. Development of the revised system of instruction was supported by the Division of Continuing Education, Oregon State System of Higher Education. It is through their cooperation that we are including the present version of this chapter in this manual. Information on the more complete system of instruction on proposal writing may be obtained from the Division of Continuing Education.

The chapter is directed primarily to two groups of readers, the person who has not written a proposal and the one who has not written any successful proposals. The chapter, thankfully relatively brief, attempts to examine each major aspect of a proposal from two perspectives: that of the author and that of the reviewer. It presents the criteria often used to judge each component of a proposal and then identifies the typical flaws found by reviewers.

The chapter has been forced on a number of audiences and appears to be pretty easy reading. The authors have not been concerned with grantmanship, or the kinds of political negotiation that may be involved in obtaining contracts. Their concern is solely with constructing a logical and communicative proposal.

IX



"Alice soon came to the conclusion that it was a very difficult game indeed."

In one sense the chapter presumes and builds upon the skills identified in all the previous chapters. The authors have attempted to make the chapter useful for both the proposer of research studies and the proposers of the developmental activities. Throughout the chapter you will find separate sections devoted to each respectively.

The writers have been accused of lacking the appropriate grim and somber approach to this important topic. They plead guilty. They have had fun writing it, and fun presenting it.

IX

Proposal Writing

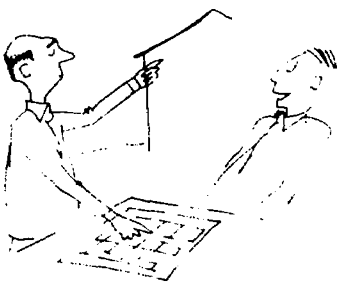
Orientation

This manual contains a short text and several appendices. In the text we have attempted to present an over-all view of proposal writing in a brief and readable form. You can probably read the text in less than an hour. Each component of the manual is expanded in the workbook. The manual will tell you what should be done; in the workbook activities you will develop skill in doing.

The appendices in the manual may be useful for permanent references.

The acid test of whether you have attained the competencies to do clear and effective planning in your own special area, is the writing of an adequate proposal. Some competencies are prerequisite. If you don't know the territory, or if you can't explicitly identify your objectives and describe how you'll feel if they're attained, then resign yourself that any proposal you conceive will be abortive.

But a healthy proposal needs more than mastery of such skills. The "more" consists primarily of certain communicative skills.

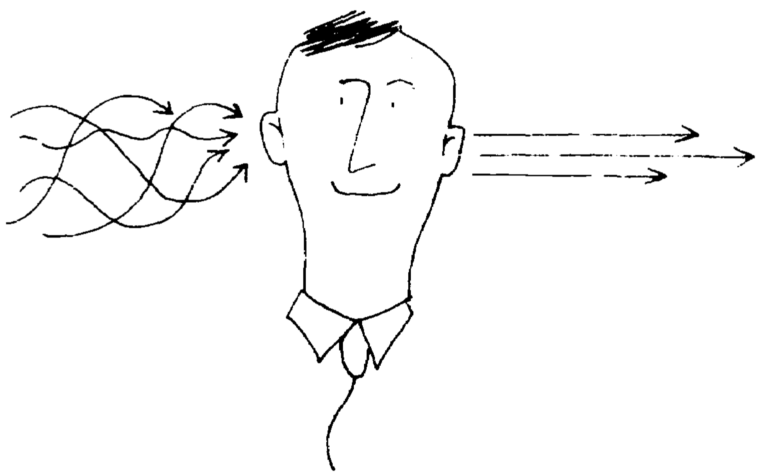


The architect may have created a great design; the blueprint specifications may be accurate, clear, and detailed; the contractors capable and realistic in their estimates, etc., but some-

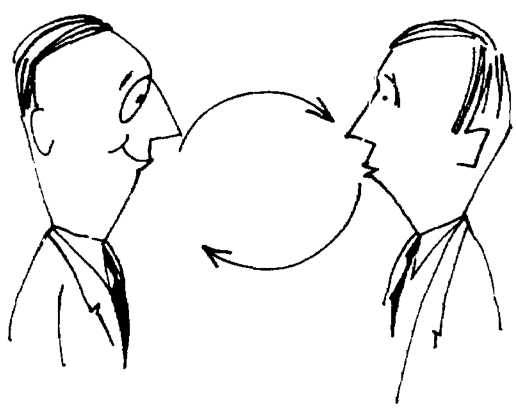
body has to <u>pay</u> for the building costs. The analogous task is to convince the bank that this is a needed and worthwhile investment.

Few developers of projects can, out of pocket, fund the kind of project they want to do. The decision to fund or not will be made by others; and, in most instances, the decision will be based solely on the proposal. However, viewing the proposal as a mercenary media should not obscure its other functions. The proposal can also serve:

1. to clarify ideas and detect limitations in your own thinking,



2. to communicate to peers and thus instigate useful comments and jolly criticisms,



3. as a first-draft report of the completed study.

A sound proposal requires little more than a change of tense, a filling-in of data actually collected, and amplified discussion to become a complete report. Most of us think complete proposals are good training devices for graduate students but not for ourselves. They entail too much hard work.

A proposal, being a detailed plan for a research study, is addressed to three general questions:

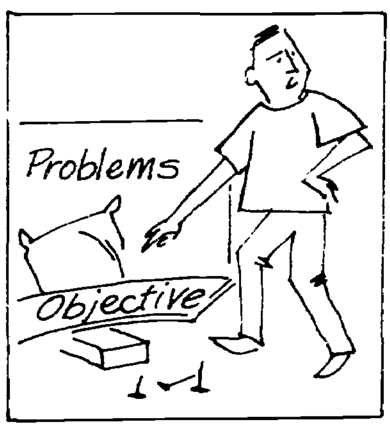
1. Why does this study need to be done?
2. How are you going to conduct the study?
3. What will be the effects of the study when completed?

Proposal formats represent an analytic outline of the issues involved in answering these three questions. Required formats vary from agency to agency, some demanding an exhausting degree of specificity, others relatively informal. In the following discussion, the outline contains those components generally required in any proposal.

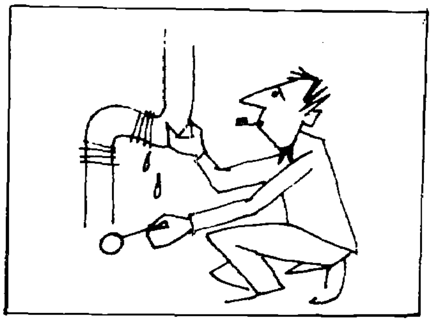
<u>Objectives of This Unit</u>

1. To identify the major components of a proposal, their function and criteria.

Desired Behavior: Recall an ordered set of proposal components with criteria for each.

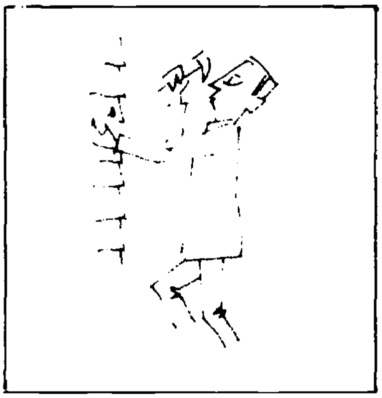


2. To detect and prescribe a remedy for certain typical weaknesses in proposals.



Desired Behavior: Identify weaknesses in examples and indicate change required to approach criteria.

3. To construct a sound proposal outline in a problem area of your interest.



Desired Behavior: Write a good proposal, get it funded, and let us know.

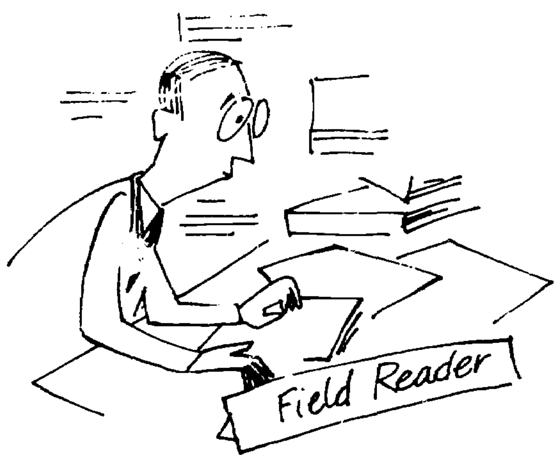
The Target Audience

Most proposals are sent by the funding agency to a board of reviewers. Each of the reviewers independently analyzes the proposal and makes a recommendation. Usually, he reports reasons for his decision, and within the "recommended for funding" category may rate the proposal for priority or desirability. Decisions of reviewers are followed by the funding agency.

Since the reviewer plays such a crucial role, what are some relevant characteristics of the typical reviewer?

1. He is not a specialist in your particular problem area. He will depend upon you to acquaint him in a succinct fashion with the importance of the study and the issues involved.

2. He is extremely busy. He needs to find the meat of your proposal quickly. Obscure and ponderous writing are high on his blacklist.
3. He knows methodology. Weaknesses and omissions will be readily detected, e.g., you can't fool him by skirting or ignoring a control problem.



4. He has an extensive knowledge of other proposals (good and bad) and of completed projects. There is competition for most research funds, and the quality of the competition is increasing. Your proposal will have to rank high among those in his in-basket.

Examine the outline in Figure 1. It illustrates the sequence of topics, and over-all function of each, needed for a typical proposal. As you read, think of ideas of your own and relate them to the components of this outline.

The major components of the proposal will be discussed in the order of the outline. Each discussion will amplify the function of that component for the reviewer: will present some objectives or criteria that the component should attain; and will list some common weaknesses. An itemized count of proposal weaknesses, based upon a study of many proposals, is contained in the appendix.

The Introduction

Statement of the problem. A terse, clear statement of the problem is a good opener. The object of this section is to present the nature and importance of the problem to the reviewer. The section consists of two parts: the major part, a description of a need; and the minor part, a brief previewer. The proposed study will relate to the need. The need for the study may rest upon its probable contribution to knowledge or its action upon society-- sometimes both.

ERIC

Topic	Function in Terms of General Questions Answered
Introduction	
Statement of the problem	What needs to be done and why?
Review of literature	What has been done that is relevant?
Objectives	What are the specific testable goals?
Procedures	
Design	What is the structural plan? What control will it afford?
Sampling	What population will be sampled? What size sample and how drawn?
Measurement and Data	What will be measured? How?
Analysis and Evaluation	How will data from the measures be examined?
Time Schedule	How much time will be needed to complete each portion of the study?
Product and Use	What will be the end-product of the study? What contribution could it make? How?
Personnel and Facilities	Who will do the study? What is their relevant competence?
Budget	What will each part of the study cost?

Figure 1. The Major Components of the Proposal

ERIC

If this section fails to convince the reviewer that the study is worth doing, then the entire proposal fails.

a. Criteria for judging the significance of the problem include:

(1) Generalizability - either in terms of theory, or to other populations, or to a range of practical problems, or a demonstration to be emulated, etc. The fantasy of funding agencies lightly turns toward eternal monuments to world-wide educational progress flowing from each small contract. Soberly, they want to see that you affect something more than your own corner.

(2) Theoretical contribution

(a) testing, expanding, or qualifying previous theories.

(b) adding to, creating, or otherwise contributing to new theory.

(3) Empirical contribution - (similar to (2) above, except relating to observation)

(4) Practical contribution - relating to a critical social need and its solution.

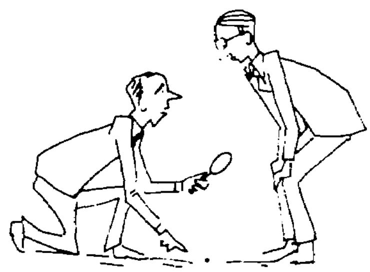
(5) Methodological contribution - creating, refining, or extending an instrument or technique.

(6) Innovativeness - this criterion is part of (2), (3), (4), and (5) but is periodically emphasized by many agencies.

(7) Panic points - "do it now or it will be too late." Either the opportunity to gather the data will be gone, or the problem will have grown unmanageable, or costs will skyrocket, etc.

b. The common weaknesses in statements of the problem are:

(1) The problem does not strike the reviewer as significant. Probably, the description of the need has failed to grab him vitally. Some problem areas are of great significance, but the aspects stressed in the proposal are among the trivial variables involved. How can you determine the most relevant variables of a problem? One technique is to list them by such classes as: learner characteristics, teacher characterisitcs, instructional methods, etc.

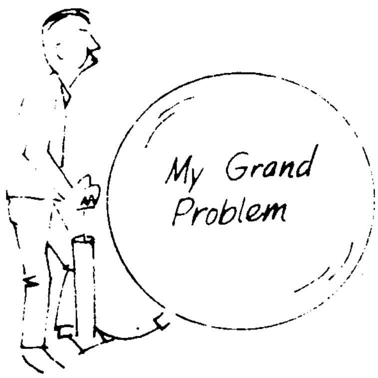


Then compare them singly and in combination. Which seem to hit the heart of the problem? Previous knowledge of both theory and practice are yardsticks, and the criteria mentioned above.

Clarity of expression is crucial here. Try out your presentation on colleagues. Not just old pals, but colleagues who, although relatively bright, stubbornly follow views divergent from your own sensitive outlook.

ERIC

(2) A second error in presenting the problem is inflation. A small grant contract will not erase the woes of education, nor will all the funds in the agency. Reviewers know this well. You can demonstrate your awareness by limiting the problem to treatable or testable scope. If this appears to involve a conflict between generality of the study and manageable limits, cut it down; then think about implications.



The slant of the proposed study may be indicated in a few sentences. Inclusion of an overview of the research direction, and definitions of any esoteric terms or abbreviations to be used subsequently, will warm the heart of the reviewer.

The following two paragraphs are from the problem section of a research proposal.

"Personality evaluation has been a problem of concern to psychologists for many years. The science of psychology has grown through our knowledge of the ways in which man reacts to his environment. This growth, along with man's quest to explain natural phenomena has led some men to make attempts to find some ways to describe human reactions and to make predictions, even if tentative in scope, of how certain persons will react when faced with specific situations.

Knowledge of human behavior has been gained through the many investigations made throughout the history of the science. This knowledge has taken numerous forms, primarily as a result of the divergent frames of reference used by the investigators in their attempts to explain human behavior."

Even the Confederate dollar was never this inflated.

9.1

Related literature. This section should establish the base from which your study moves. Summarize the pertinent research or practices, evaluate them, and demonstrate how your project relates to them--particularly, how it contributes to progress in its specific field. This is the logical first step in proposal preparation. Often, a thorough search will uncover a new perspective to the problem or a refinement of your original approach. Typically, the search means digging through professional journals in the library. Some hints about library and other sources are given in Appendix B.

Functions of this section include:

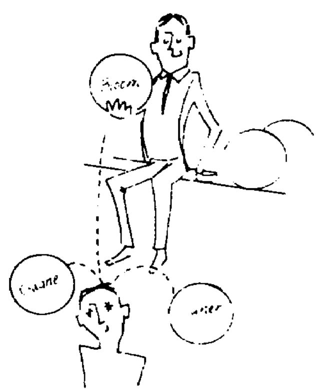
a. Providing the non-specialist reviewer with a succinct overview of the particular area.

b. Offering some evidence of your scholarly skills.

c. Revealing that you are aware of recent developments.

d. Delineating how your study springs beyond previous work.

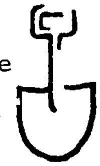
A lengthy list of titles or names is of no use to the reviewer. He has encountered fluent name droppers before. Instead, focus upon a few select references, really relevant to your study, accompanied by a critical analysis of their methodological limitations. This will illustrate some of your research skills as well as your knowledge of the area.



Two kinds of assertions to be avoided as the black death are:

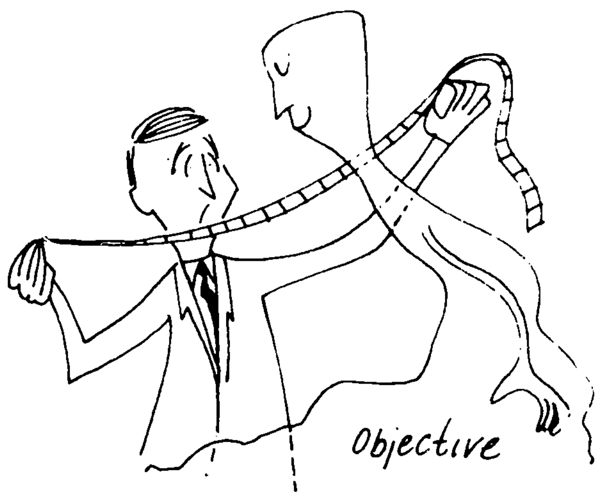
"No relevant work has been done in the area."

or "So much has been done and written in this area that a summary is impossible."



Reviewers feel that the authors of such statements should never have been allowed out of graduate school or never admitted. 9.2

Objectives. If the proposal is primarily directed toward an extension of knowledge, this section will be focused upon questions or hypotheses. A more applied or practical direction may shift the focus to products. The reviewer will be concerned that the objectives are achievable.



One essential requirement is that they be _testable_. If no possible outcome of the study can refute the objective or hypothesis, it is not testable. The procedures must bear on the truth or falsity of stated hypotheses. This is the main reason behind our reiterated emphasis upon operational definitions and behavioral objectives. They constrain concepts and goals to face the evidence.

Many important project topics involve covert processes or categories that are not directly observable, e.g., love, intelligence,

thinking. It is certainly legitimate to construct parallel lists of objectives, a theoretical and an observable set, the latter derived from, or the implied effects of, the former. In a research proposal the observable are a must.

In areas where theory or prior observations are minimal, questions are an appropriate form. As knowledge increases, hypotheses become more appropriate. Questions remain the typical objectives of survey studies: How many? What is the relationship between? However, the more specific the question the more it reflects the preparation of the investigation.

Disregard the nonsense frequently written about stating hypotheses in a "null" or no-difference form. If you have hypotheses, state them the way you expect the results to go. The null is an alternative (or set of alternatives) which is logically indispensible in analyzing your data. It belongs in the results section, not here.

The objectives should flow from your statement of the problem. You have stated a need. Now, in what way are you contributing to filling that need? The reviewer will then look to see if you have set up procedures to attain the objectives. In the methods or procedures section there must be a step-by-step account of what will be done to approach each objective.

Common flaws in this section include:

a. Pie-in-the-sky objectives. Vague global goals impervious to the study's outcomes.

b. Sandbag objectives, not referred to or dealt with by the proposed procedures.

c. Hidden objectives, contained somewhere in the text of the proposal. An astute and diligent reviewer can find them <u>if</u> he takes the time.

Judge the following statements of objectives.

(1) The purpose of the investigation is to show that the unobserved and unmeasured behavior of high school students is different from their behavior when it is observed and measured.

(2) The object of the survey is to find out what proportion of Eskimo graduates of Neeknak High School are earning less than $1,000.00 annually.

(3) The broad hypothesis is that a procedure can be followed which will lead to the initial formulation, revision,

and final development of a broadly conceived theory of education based upon sociological and other relevant research findings.

(1) appears untestable; (3) is vague: (2), while it may not carry weighty import, is clear, succinct, and measurable.

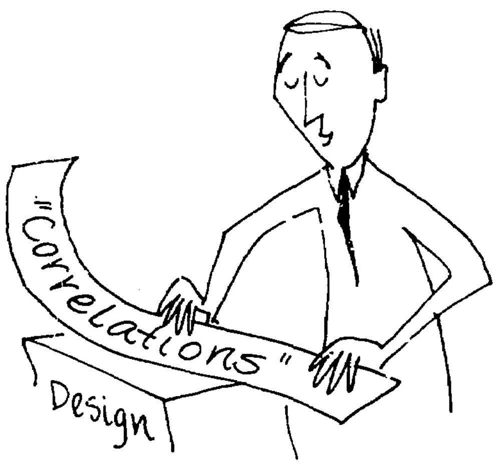
9.3
9.4

Procedures

This section presents a detailed explanation of what you are going to <u>do</u>. It, above all other sections, is cast in behavioral terms and operational definitions: a concrete description of the project. If the reviewer gets this far, he will devote more attention to your procedures than any other section for two reasons: one, he can easily discover from this down-to-earth section just what the study may accomplish; two, procedures are a spawning bed for numerous species of error.

If possible, a short introductory paragraph indicating the overall kind of design and the sequence of steps you will take is helpful.

Research project design. If you are using a design with a conventional name, so label it and then explain how this situational plan

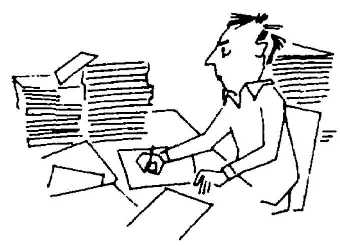


ERIC

handles the independent variables and provides for control. The term "design" in research literature refers to the structure of an experiment rather than to a non-experimental study. You are probably almost painfully aware that the previous materials are strongly slanted toward experimental studies. Evidence from an experiment is more highly regarded than evidence derived from other approaches.

However, not all funding sources are looking for, or require, experimental approaches. A discussion of the relative merits of non-experimental techniques may be found in Kerlinger (1964). If your study is not experimental (i.e., correlational, case study, survey, etc.), indicate its category and give special emphasis to control measures that you will use.

Check back to your statement of objectives. Lay out the steps to answer each question or to test each hypothesis. A summary chart with questions or hypotheses in one column and the procedures relevant to each in an adjoining column may help.

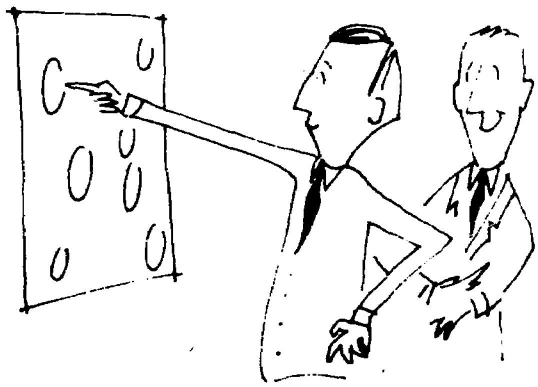


The design description must account for any variable that might influence the results of the study. Indicate how you will control it, e.g., build it into the design and measure its effect, block it off, incorporate its measures into your tools of analysis, randomize, etc. A useful discussion of the kinds and relative merits of control measures may be found in McGuigan (1960).

No design affords perfect control. Some factors will remain as possible sources of contamination. It is the <u>degree</u> to which these

ERIC

are strained out that marks the better design. Reviewers tend to be more favorably impressed if you demonstrate your awareness of the weaknesses of your study. At least itemize the loopholes and indicate why you were unable to control for them. Sometimes you can point



to cost, i.e., the enormous number of requisite control groups, time and effort required, etc. The reviewer usually has broad experience with such limitations. But, he will prefer that <u>you</u> point out the weak spots and why they remain.

As a reviewer, imagine how you would comment on this proposal design:

> A Psychologist proposes to test the hypothesis that early toilet training (Head Start) leads to a type of personality noted by compulsive cleanliness; conversely, late toilet training leads to sloppiness. Previous studies have shown that middle-class children receive toilet training earlier than do lower class children. Accordingly, he asserts he will select two groups, one of middle-class and one of lower-class children. He will give both groups a finger-painting task and compare the amount of smearing and how many times they wash the paint off.

Developmental project design. If you are proposing a developmental project, you may not be concerned with the sophisticated plan of observation referred to as an experimental design. However, if you strive for more and more thorough evaluation of your project, your plan will require something akin to these designs.

Your object in this section is to show the reviewer how you are going to attain the objectives that you have stated in the previous section.

It is appropriate to first present some summary of the general overall steps you propose, then break each large step or phase into its sub-parts. By doing this, you allow the reviewer to maintain a perspective across the entire project.

Two emphases should guide you: clarity, and relation of procedures to objectives. Your plan must describe your steps in straightforward language, so that any reader can tell both what you propose to do and the sequence of your proposed steps. But you don't need to feel locked into a fixed and unalterable sequence. If you wish to have options in your plan, provide for these and state them explicitly . Also, spell out just what information you will use to reach each decision and how that information will be attained.

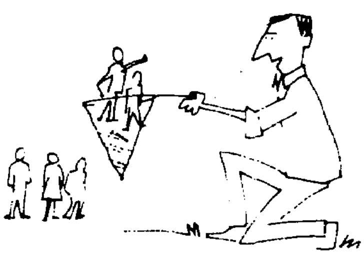
Continually cross-tie the steps of your plan to your objectives as closely as you are able. Preferably, describe the

contribution of each major phase of your project--the objective or set of objectives that this phase bears upon. Then explain how, in turn, the smaller steps within each large component will contribute to the objective or objectives of that component.

It is often helpful to the reviewer if you summarize this in a chart, perhaps listing objectives on the left side and relevant procedures as they relate to each objective, on the right side.

The typical weaknesses of the procedures section of developmental proposals is that the reviewer can't discern the means-end relationship between procedure and objectives. Too often this is due to such vagueness in the description of both sections that the objectives would encompass any procedures, and the procedures would fit almost any objectives. 9.5

<u>Sampling</u>. Because you wish to generalize any findings beyond the particular subjects in your study and/or beyond those particular



days and places when the investigation occurred, most studies are perceived as operating with a sample and generalizing their conclusions to a larger population.

This target population must be adequately described. The reviewer should be able to tell readily who is not included in the population. To a large extent your choice of sample determines the population to which the study applies. Logically, the sequence would be to define

ERIC Full Text Provided by ERIC

the population and then specify how a random sample, or some modification of a random sample, would be selected. In instructional projects you are often constrained to study the students available, so that you must construct the relevant features of the population from the characteristics of your sample. List those subject characteristics which previous knowledge indicates may influence the behavior you are measuring. Omit characteristics which have no effect on the measure, e.g., color of eyes (within the normal range of colors) rarely affects measures of learning.

In theory, in the proposal, and actually, you will probably define the population first--in terms of relevant characteristics--that is, characteristics probably having measurable effects on the dependent variable. Then, show how you will select a sample. The reviewer will be concerned whether your method of sampling tends to systematically bias the measures you take. Point out your efforts at randomization or stratification. If you have to use intact classes, schools, or other inclusive groups, be sure you have clearly stated this.

Random sampling procedures are highly rated by reviewers and for most purposes yield logically defensible results at relatively low cost. Meticulous reviewers tend to look for three kinds of randomization procedures (in order of priority):

(1) Selecting the subjects from the population.

(2) Assigning the subjects to groups, e.g., in a formal group study placing them in Group I, II, III, or IV.

(3) The, and this is a separate step, assigning of the treatments called for in your design to the groups.

Generally, your sampling concerns relate to the people whose behavior you are studying. Occasionally you may have an interest in sampling from a variety of treatments. Get some design consultant assistance on this. It does offer power in the conclusions attained.

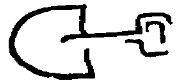
Sample size needs to be specified for each group in the study. Three factors should be considered in determining the number of subjects involved.

1. The degree of effect that interests you. If you desire or demand that your instruction or treatment makes a noticeable effect on a small group (say 20 students), use small groups. If it makes only a tiny effect, you can meet all statistical criteria but you will need larger groups.

2. The cost involved. Larger numbers, if not easily available, may exponentially inflate your budget.

3. Statistical considerations. Previous knowledge of any consistent effect can be translated into the sample size required to reach the desired statistical conclusion, but only when the effect is known to be in the appropriate direction and its extent as well as an estimate of variability has been measured. This is one by-product of pilot studies. Almost any statistics book or graduate student in a statistics course can show you the translation procedure.

If possible, plan to use equal numbers in each group. If you can't assume equality, get some consultant help (statistical) to

point out how you will deal with unequal numbers. Reviewers like to see that you have worried about this.  9.6

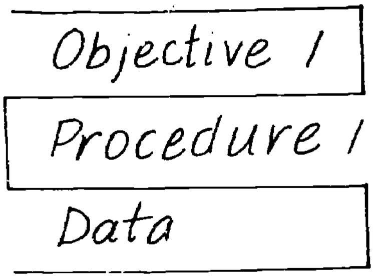
<u>Measurement and Data</u>. If you're going to discover how closely you have attained (or not attained) your objectives, then you must collect some data, however gross it may be.

The kind of data or measures you plan to collect will flow from your statement of objectives. Again, to the degree you have been clear and specific in stating these objectives, the measures to be used will be most clearly indicated. Every major objective and sub-objective should be paralleled by description of the measures you will take to indicate whether that objective is being achieved. It may be that you will find acceptable, conventional and standardized measures that are suitable for your purposes. But, probably not. In a developmental project don't depress yourself by worrying whether your measures are highly sophisticated or meet close psychometric standards of daintiness and purity.

As a rough guide, ask yourself what measures would indicate to a reasonable outside observer whether or not you were obtaining each objective. Enough data in the form of low-grade ore may illuminate your project. This is probably better than waiting in the cold for an unattainable measure.

To communicate your data plan to the reviewer, it may be useful to summarize the points or steps in your project at which you will collect data, and then the kinds of measures you will take. Particularly helpful, to the writer himself, is an outline indicating not

only the relationship between objectives and procedures, but between <u>data</u> and <u>objectives</u>. Possibly a four-column chart could



be constructed in which the left column contains objectives; the second column procedures; the third data to be collected; the fourth, a short statement of how the data collected will relate to the attainment of objectives.

If you feel baffled by this problem, or if you are determined to do a more thorough job than your own resources allow, consider calling in a consultant. However, consultants who are both sophisticated and practical are rare.

<u>Analysis and evaluation</u>. This component describes how you will examine the data to obtain evidence bearing on your objectives. The appropriate examination tools are determined in part by the objectives, in part by the design, and in part by the class of data. (1) Name the analysis tool, (2) show why it is appropriate, and (3) indicate how the product of the analysis will bear on your objectives.

If you have a proposal constructed to this point, and are still a bit cloudy about the appropriate analysis, seek help. In a small institution you can usually find some staff member in mathematics, psychology, economics, or even education who can and will help for the price of coffee and a sympathetic ear for his troubles with the administration. If possible, seek out one who has been teaching statistics regularly and recently. If you can't find local assistance, call the nearest Regional Office of the USOE. Addresses are in Section X. They will refer you to the nearest specialist or provide direct assistance.

This is the place to set out the null hypotheses. A useful technique is to place the research hypothesis of interest in one column; the alternative null in another; and in a third column, the kind of result that will indicate rejection of the null to be the logical decision.

Frequently, you will be unable to cite the appropriate tools because unknown, as yet, dimensions of the forthcoming data will determine what tool to use. Spell out the contingencies as you see them. Show alternative plans of analysis. Reviewers wax warm and beneficent when they perceive you have anticipated problems and have planned coping strategy.

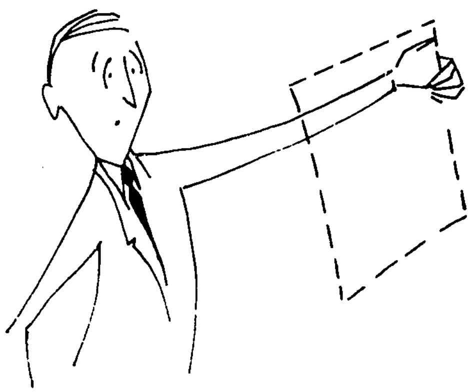
If you are writing a developmental proposal, a section on evaluation will probably be required. However, you may not receive much help from either guidelines or the funding source as to how you are to evaluate your project. The sharpening emphasis on

ERIC

evaluation is a national phenomena, stemming from the gradual awareness that practically nothing is known as to the outcomes (particularly learning outcomes) of most funded projects. This, of course, is not true of experimental or research projects.

Assuming that a reviewer is examining your proposal for the evaluation section, there are three typical weaknesses he will be ready to spot:

1) The Next-to-Nothing Plan

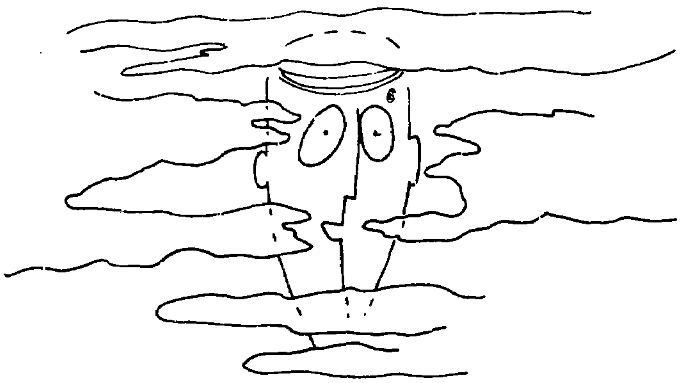


That is, nothing more than a planned autobiographical case report, e.g., here is what I liked about the project; here is what I didn't like, usually administered in the presence of the project administrator and his staff.

2) The Golden Smear Plan

Usually more elaborate than above, but consisting mainly of inferences, opinions (often those of the project administration), and so vaguely describing any other measures, that favorable ones can be gleaned, unfavorable ones can be discarded, without doing violence to the Smear Plan.

3) The Scholarly Fog Plan



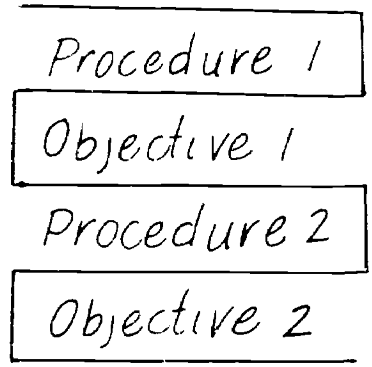
The writer talks learnedly of evaluation parameters but in highly abstract terms. Presents a complicated model of evaluation, but no concrete or detailed translation into terms of the project itself.

The plain truth is that the evaluation of developmental projects has not yet attained a moderately high plane of art or science. Experienced reviewers know that most projects are inadequately evaluated. This means that the level your proposal must reflect, to meet competition, is not high. A set of simple, relevant, straightforward measures will do, amply. Don't worry about obtaining a standardized test for <u>every</u> measure. There aren't enough to go around. A home-made measure that is really appropriate to your objectives will do as a starter.

One way to develop an evaluation section is to return to your objectives. Indicate the measures you'll take to determine the progress made toward each objective, and compile that set of measures as your plan of evaluation. The more clearly, concretely and behaviorally you are able to translate your objectives, the easier this task becomes.

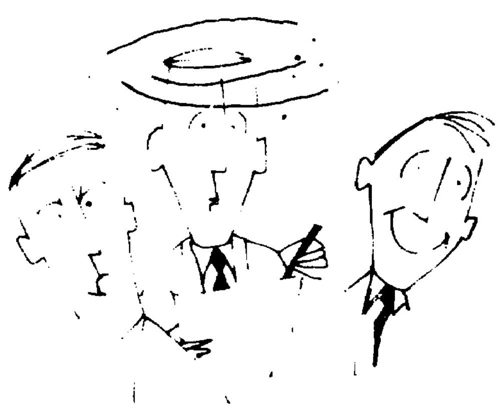
An additional way of developing this section is to review your own procedures to determine just how you will be able to demonstrate what was done on the project. The proposal says what you <u>intend</u> to do. What you <u>actually</u> do may approximate this, or may be at consi-

derable variance. Set forth some simple record scheme, so that what actually goes on can be recorded. If the project was changed, and most projects are changed from the proposal to some degree, then, describe how the decision to change will be made. Any evaluation



plan that includes recording procedures can provide for decision points at which certain classes of information can be revised.

A third, and additional, possibility is to include, besides measures of the immediate <u>objectives</u> of the project, some measures of how the project affects the people and institutions most closely in contact with project processes and outcomes.

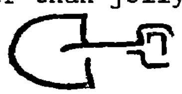


For example, what are their feelings about the project, their opinions, what has it accomplished in terms of their perspectives?

I suppose the logical evaluation sequence would be 1) procedure, 2) objectives, and 3) impact of the project. However, measures of the <u>objectives</u> are of the highest priority.

If you have examined your objectives, but feel so inadequate about measurement that you can't itemize or describe exactly how each will be measured, perhaps you can at least generally indicate the kind or class of measures you will use; or how you and your staff will develop measures; or how you will search for measures, i.e., use other knowledge or expertise.

It is certainly not inappropriate to indicate that you plan to use a consultant to either assist you with the plan of evaluation, or to develop and implement your plan. Unfortunately, in practice you will find that useful consultants are rarer than jolly statisticians.

 9.7

<u>Time schedule</u>. A realistic time schedule adds weight to your proposal. Graphic or flow chart representations are easiest for the reviewer to follow. A typical graphic presentation is given in Figure 2, and a flow chart is shown in Figure 3. Large scale proposals are often accompanied by a P.E.R.T. chart. This system, Program Evaluation Review Technique, takes a bit of study but has a high payoff in increased accuracy of your planned sequence of actions. A P.E.R.T. chart example is given in Figure 4.

If you are not experienced in the kind of project you propose, allow more time than you anticipate. Copies of materials and instru-

ments will refuse to appear on time, subjects will hide or get busy with life or death urgencies, your assistant will misread the instructions and you'll be searching for a new group of subjects as well, etc., etc. And remember, all funding agencies require a report. This takes secretarial and duplication time.

Usually the funding agency will indicate, in advance, a starting date if the project is approved. Allow for a delay, even in this.

9.d

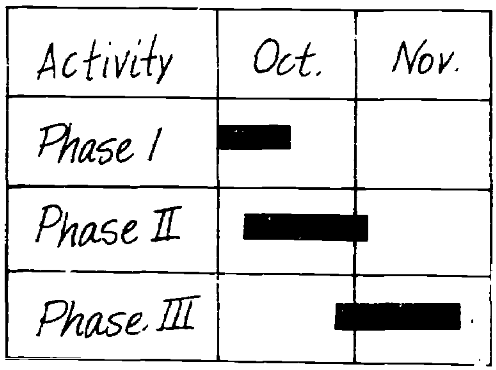


Figure 2. Sample of a Graphic Time Schedule

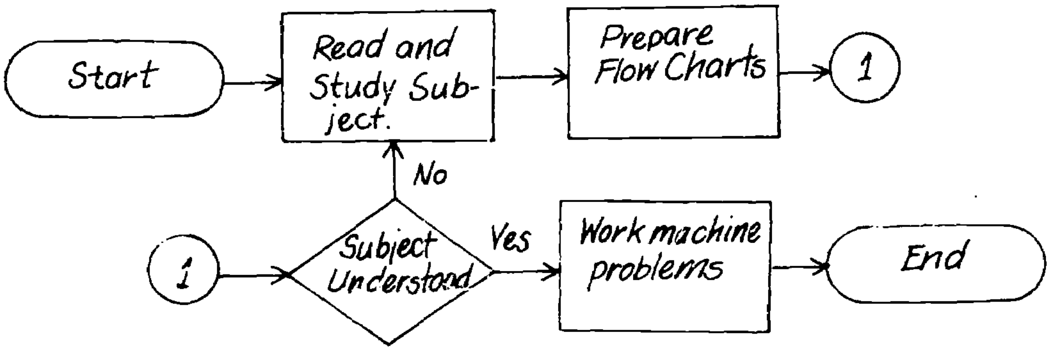


Figure 3. Sample of a Flow Chart

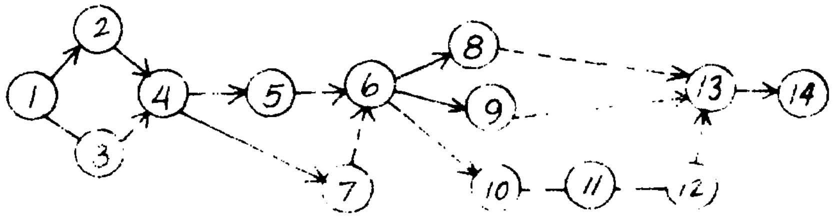
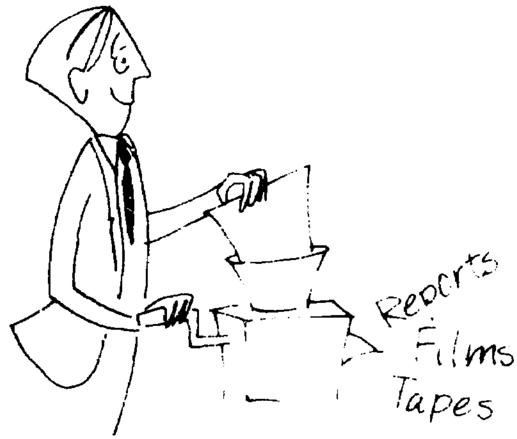


Figure 4. Sample of a PERT Chart

Product and Use

The purpose of this section is to point up what contributions the completed study will make and how. The final report itself may be suitable for dissemination through other channels. A journal article may be a reasonable expectation.



The effects on both the body of knowledge and on working practitioners may be estimated. Suggest how dissemination to the latter may be effected. Often the study will generate by-products such as evaluation instruments, instructional materials, films, etc. These may be highly useful to others.

Perhaps a significant contribution to your own institution is a likely yield. Even the effects on the research team or the individual researcher in terms of professional experience and increased capabilities may be pointed out.

But, keep it brief.

Personnel and Facilities

Name the people who are to work on the study. Briefly indicate experience relevant to the project. If you haven't any, mention that you have read this manual. Evidence of interest in pursuing this

kind of activity is appropriate even if experience is lacking. Don't deliver a eulogy to a bunch of nice guys. Reviewers loathe nice guys.

As a last ditch resource consider acquiring the services of an experienced researcher as a consultant to the project. He will insist on being paid and this will add to the budget. Spell out clearly just how he will serve the project. Retain the decision function and allow him to advise and recommend.

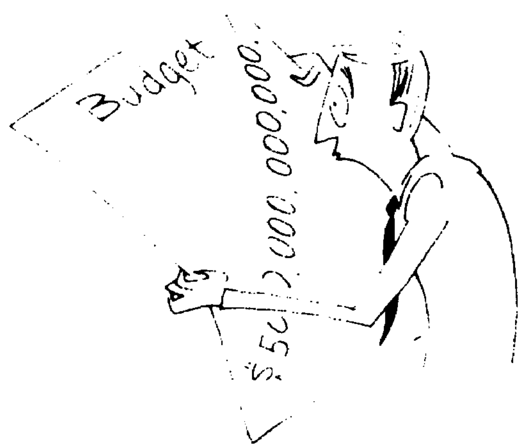
If you have special facilities that will aid or are required by the project, list them. Libraries suitable for graduate work, computers, training programs, secretarial services, and office space may be relevant.


9.9

Budget

The easiest way to get a feel for proper budgets is to peruse some approved budgets for projects granted by the same agency. The agency will have sent you as part of the application form, probably a rather detailed budget schedule. Use some worksheet facsimile and construct a sample budget. Show it to an experienced researcher. Usually, he can quickly scan it and give you some useful suggestions.

Your own fiscal officer may help. The institution may have to contribute costs you hadn't anticipated unless you provide for this. Show each facet of your institution's contribution from staff time, secretarial services, materials, etc.



The usual error of first efforts at budget construction is to underestimate and omit. If your procedure is detailed, the description can be used as a guide. Check every step including preparation of the final report. A sample budget sheet showing the main components is shown in Figure 5.

But, don't be overconcerned with this component. Budgets on approved projects are often changed through negotiation with the funding agency. Consider the proposal budget as a good first draft. Many first-rate projects have to be re-budgeted later.

9.10

Selecting the Funding Agency

Supporting sources are so numerous that this entire manual could not list them. A guide to sources of information about funding agencies is contained in Appendix E of this section. It also contains some guides from the United States Office of Education, your most probable funding source.

I. DIRECT COSTS

A. Personnel

1. Research

2. Support

3. Secretarial/Clerical

4. Consultants

B. Employee Benefits

@____________%

C. Travel

D. Supplies & Materials

1. Project and/or Instructional Materials

2. Office Supplies

E. Communications

F. Services

G. Final Report

H. Equipment and/or Equipment Rental

I. Other Direct Costs

J. Subtotal, Direct Costs

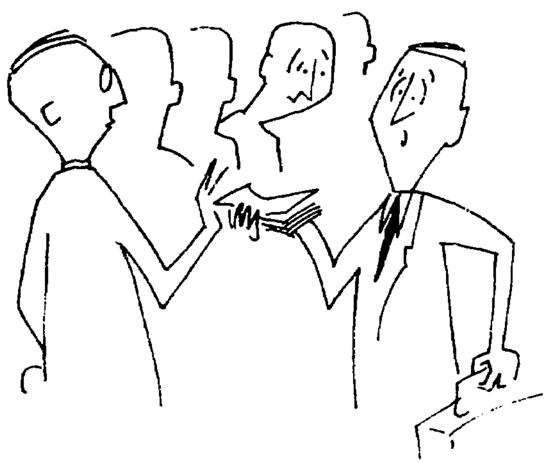
II INDIRECT COSTS

III TOTAL COSTS

Figure 5. Typical budget categories

It might be to your advantage to consider applying to some of the lesser known funding agencies. For your first project, don't overlook your own institution. The president of a small college can often dig up a little seed money for a pilot project.

Agencies will send you their requirements, formats, and usually some indication of their interests. Often a short letter outlining the project in an informal way will help them to decide if it is



worthwhile to submit a detailed proposal. Many agencies encourage such informal letters. You may find, however, that cramming your notions into a two-page letter demands more clarity and brevity than you wish to give.

And, don't overlook the phone. A call to the appropriate official may save weeks. Feel free to call the agency to find out who is in charge of what.

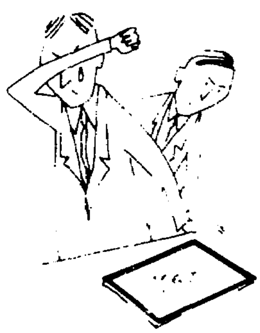
The Research Marketplace

The preference of agencies for different kinds of projects changes over time. I have stressed the orientation that if you are interested in a problem area, work up your ideas, then locate the appropriate funding source. A contrasting approach is followed by some researchers and developers. These social science streetwalkers try to sound out agencies as to the latest priorities, then write up proposals in that area--wherever the demand lies.

Some compromise with reality demands is inevitable. However, the increasing multitude of funding sources lends assurance to the idea that a good proposal will get supported (sooner or later).

What To Do When Your Proposal is Rejected

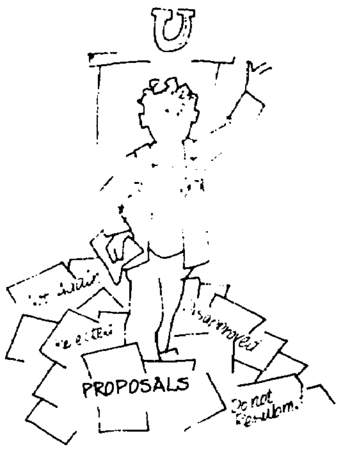
After you have decided to change your specification, written and torn up several suitable rejoinders, etc., find out *why* they turned it down. Most agencies will send you a summary of reviewer comments. Evaluate them. If the critiques are apt, re-do the proposal. Often the agency will indicate if they



think re-submission is suitable. Or if the proposal looks good to you, submit it to another agency. Some good projects have gone through three rejections from different agencies, then been funded and subsequently acclaimed as hallmarks of progress. To gain perspective, talk to any publishing author.

Three complete proposals are contained in the appendices: a research proposal; a fairly large developmental one; and a short, rather informal, developmental proposal. They are fair-to-middling quality (and were funded). You will also find some references related to each component part of most proposals.

Good luck.



Write when you are funded.

Appendix A

Proposal Writing

Crawford, Jack. "Proposal Writing." In _National Research Training Institute Manual_ (C.O.R.D.). Revised. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1966. Chap. IX.

Good

Krathwohl, David R. _How to Prepare a Research Proposal_. (Available from the Syracuse University Bookstore, 303 University Place, Syracuse, New York 13210.)

An excellent checklist approach, especially after proposal is completed.

ERIC

Appendix B

Introduction

1. <u>Statement of the Problem</u>

<u>Formulating the Research Problem</u>. Inglewood, California.: Southwest Regional Laboratory for Educational Research and Development, 1967.

An easy to read, nontechnical presentation to aid the beginning researcher to define his problem practically so that his study will be useful for present and future educational divisions.

2. <u>Review of the Literature</u> - <u>ERIC</u>

The first item to remember when beginning a search through the literature is that the reference librarian is a major source of information. The reference librarian can save countless hours, if only the researcher will present the librarian with information, specific if possible, on what is being sought. The reference librarian can quickly point to the most pertinent information. Crucial books may be kept at the reference desk, so don't overlook asking for suggestions about books which may be held behind the desk. For example, the book <u>How to Locate Educational Information and Data</u> is a reference source which is usually kept at the librarian's desk.

The card catalog itself is a major resource. Most large libraries have extensive subjects headings, in red, for subjects pertinent to research. If "Educational Research" is not located under that heading, try "Research, Educational". Books which contain summaries of educational research may be located through the card catalog. Also, large libraries have special librarians who are authorities in their field, i.e., social science, education, etc. This specialist can be called upon if the reference librarian is not aware of some of the resources.

Indexes: Major sources of information for educational researchers are:

(1) <u>Psychology Abstracts</u> (indexes of authors and subjects at the end of the volume; also separate volumes containing indexes to authors and subjects, cumulative for the past several years, are now available);

(2) _Education Index_, classified by author, title and subject (remember that in most indexes _author_ and _subject_ are usually more accurately included than are titles):

(3) _Sociology Abstracts_ (have separate indexes, of author and subject);

(4) _Dissertation Abstracts_ (has a cumulative index for the year);

(5) There are technical indexes such as _Operations Research Management Science_, _U.S. Government Research and Development_ reports (prior to 1963 indexing is terrible, almost impossible to find anything except under _subject_);

(6) _Business and Periodical Index_, which contains references to articles dealing with business education, for example.

(7) _Books in Print_, Subject Index; and,

(8) _American Library Association_, _Standard Catalog_.

(9) E.R.I.C. Become acquainted with the E.R.I.C. System. This Educational Research Information Center, established by the U.S.O.E. Bureau of Research, appears to be developing into the most useful and comprehensive source.

E.R.I.C. is a national information network of decentralized information centers for acquiring abstracting, indexing, storing, retrieving, and disseminating the most significant and timely educational research reports and program descriptions.

The Educational Research Information Center provides two major classes of service. Its periodic and non-periodic publications announce the acquisition of new material and the collections already in storage. Its document reproduction service provides either hardcopy microfiche reproductions of stored material. The cost is not high. For further information about E.R.I.C. write to E.R.I.C., U.S. Office of Education, 400 Maryland Avenue, S.W. Washington, D.C. 20202.

Downs, Robert B. *How to do Library Research*. Urbana, Ill.: University of Illinois Press, 1966.

An excellent description of library services, research techniques, and lists of sources in a broad range of topics.

3. Writing Hypotheses and Objectives

Mager, Robert F. *Preparing Instructional Objectives*. Palo Alto, California: Fearon Publishers, 1962.

A humorous, easily read approach to objectives.

Appendix C

Procedures

1. Research Design

Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand Mc Nally, 1963.

A non-beginner discussion of several research designs and their validity.

Cox, D. R. Planning of Experiments. London: John Wiley & Sons, Inc., 1958.

A presentation of experimentation design from the statistical viewpoint with a good chapter on randomization. It is for the learner a step above the beginner and not specifically written for educators.

McGuigan, F. J. Experimental Psychology: A Methodological Approach. 2nd Ed. New Jersey: Prentice-Hall, Inc., 1968.

A readable exposition with plenty of examples of a few basic research designs.

2. Sampling

Cox, D.R. Planning of Experiments. London: John Wiley & Sons, Inc., 1958.

A presentation of experimentation design from the statistical viewpoint with a good chapter on randomization. It is for the learner a step above the beginner and not specifically written for educators.

Grobman, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point, Ch. 3. AERA Monograph Series on Curriculum Evaluation, No. 2. Chicago: Rand McNally and Co., 1968.

An excellent presentation on both a conceptual and applied level on evaluation with clear, easy to understand sections on sampling, scheduling and budgeting.

ERIC

3. Measurement and Measurement Instruments

Anastasi, Anne. Psychological Testing (3rd Ed.). New York: The Macmillan Co.

An in-depth study of testing and measurement with greater detail about some of the standard tests in use and should be read after Tyler and Stodola-Stordahl.

Bradley, Jack I. and McClelland, James N. Basic Statistical Concepts. Chicago: Scott, Foresman and Co., 1963.

A self instructional text in basic statistics which should provide a better grasp of other references in measurement.

Buros, O.K. Tests in Print. Highland Park, N.J.: Gryphon Press, 1961.

Reviews and provides data on published tests.

Buros, O.K. (Ed.). The Third, Fourth, Fifth, and Sixth Mental Measurements Yearbooks. Highland Park, N.J.: Gryphon Press, 1965-1968.

A massive collection of data and reviews of tests in use.

Helmstadter, G. C. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1964.

A more theoretical treatment of measurement with discussion of application.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt Rinehart and Winston, 1965.

It makes a dull subject interesting. Chapters 1-3 describes the concept of probability related to statistics, set theory, and variance.

Oppenheim, A. N. Questionnaire Design and Attitude Measurement. New York: Basic Books, Inc., 1966

The practical aspects of questionnaire design.

Shaw, Marvin E., and Wright, Jack M.. Scales for the Measurement of Attitudes. New York: McGraw Hill, 1967.

Chapter one describes attitude measurement, and the rest of the book discusses the theoretical aspects of attitude measurement, surveys the literature and presents and classifies examples of tests.

Stanley, Julian C. Measurement in Today's Schools, 4th Ed. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964.

A basic textbook for measurement and evaluation concepts and little background is required.

Stodola, Quentin and Stordahl, Kalmer. Basic Educational Tests and Measurement. Chicago: Science Research Assoc., 1967.

Covers theory and application of tests and measurement but greater depth than Tyler (see below).

Tyler, Leona E. Tests and Measurements. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

The theory and application of testing and measurement, also a description of selected samples of standardized tests.

Webb, E. J., Campbell, D. T., Schwartz, R. D. and Suchest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

An easily read presentation which requires no great depth of background.

4. Data Collection

Grobman, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point. AERA Monograph Series on Curriculum Evaluation, No. 2. Chicago: Rand McNally and Co., 1968.

An excellent presentation on both a conceptual and applied level on evaluation with clear, easy to understand sections on sampling, scheduling and budgeting.

5. Data Analysis

Bradley, Jack I. and McClelland, James N. Basic Statistical Concepts. Chicago: Scott, Foresman and Co., 1963.

A self instructional text in basic statistics which should provide a better grasp of other references in measurement.

Bruning, James L. and Kintz, B. L. Computational Handbook of Statistics. Chicago: Scott Foresman and Co., 1968.

A "cookbook" approach which presents the most widely used tests in an easy to follow, step-by step manner.

Edwards, Allen L. Statistical Methods For the Behavioral Sciences. New York: Holt, Rinehart & Winston, 1954.

A good introductory text.

McCollough, Celeste, and Van Atta, Loche. Statistical Concepts, A Program for Self-Instruction. New York: McGraw-Hill, 1963.

A cute easy book.

Popham, W. James. Educational Statistics: Use and Interpretation. New York: Harper & Row, 1967.

A presentation of educational statistics which side steps mathematical exposition whenever possible, relying on verbal or graphic explanations.

Siegel, Sidney. Non-Parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

A good presentation of non-parametric statistics.

6. Evaluation

Grobma, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point. Ch. 2,4,5. AERA Monograph Series on Curriculum Evaluation, No. 2. Chicago: Rand McNally and Co., 1968.

An excellent presentation on both a conceptual and applied level on evaluation with clear, easy to understand sections on sampling, scheduling and budgeting.

ERIC

Tyler, Ralph W. (Ed.) Educational Evaluation: New Roles, New Means. 68th Yearbook of the NSSE, Part II. Chicago: The National Society for the Study of Education, 1969.

An anthology of evaluation covering a broad range of applications.

Tyler, Ralph W., Gagne, Robert, and Scriven, Michael. Perspectives of Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally and Co., 1967.

Chapters 1, 2 and 4 are applicable to areas beyond curriculum evaluation and provide a good conceptual background for evaluation studies.

Wilhelms, Fred T. (Ed.) Evaluation as Feedback and Guide. Washinton, D.C.: Association for Supervision and Curriculum Development, NEA, 1967.

A broad non-technical discussion of evaluation which ranges from "what is it?" to "How do we get to where we want to be?"

7. Project Time Scheduling

Cook, Desmond L. Program Evaluation and Review Technique. Washinton, D.C.: U.S. Dept. HEW - Office of Education, 1966.

A monograph describing "PERT" an effective management system for research projects.

Evarts, Harry F. Introduction to P.E.R.T. Boston: Allyn and Bacon, 1964.

An elementary and fundamental presentation of the PERT management technique.

Appendix D

Budgets

Grobman, Hulda. Evaluation Activities of Curriculum Projects: A Starting Point. Ch. 5. AERA Monograph Series on Curriculum Evaluation, No. 2. Chicago: Rand McNally and Co., 1968.

An excellent presentation on both conceptual and applied level on evaluation with clear, easy to understand sections on sampling, scheduling and budgeting.

Appendix E

Where To Obtain Information On Funding

A. Sources and Procedures

1. Guide to Support Programs for Education. Education Service Press, Visual Products, 3M Company, St., Paul, Minnesota. 1967 (2nd ed.) Price: $12.00.

   This is a worthwhile investment for the novice. Deals with principal federal laws aiding education, aid from business and foundation sources, and provides helpful hints generally on how and where to obtain support for research, development, or training programs.

2. College and University Reporter (Topical Law Reports). Commercial Clearinghouse, Inc., 420 Lexington Avenue, New York. Price: $500 yearly subscription, includes weekly bulletin on recent developments in Washington, D.C. in legislation relating to education and research.

   This is an expensive reference source, but the services provided to the subscriber are well worth the investment. Two large loose-leaf books are included, containing detailed and very much up-to-date information on all major developments in the field of education. The volumes are revised each week when the company (CC) mails supplementary loose-leaf pages to all subscribers. Included, too, are weekly bulletins dealing with recent developments in Washington, D.C., and a copy of each law or pending law in both the House and Senate pertaining to education. This is a must for larger research organization.

3. Programs and Services: U.S. Dept. of Health, Education and Welfare. Superintendent of Documents, U. S. Printing Office, Washington, D.C., 20402, Price: $2.00.

   This is a useful and comprehensive guide to federal programs administered through the Department of Health, Education and Welfare. In light of the coverage and relative cost, this is highly recommended as a valuable reference source.

4. Office of Education Support for Research and Related Activities. Dept. of Health, Education and Welfare, Washington, D.C. Free on request.

   This 22-page manual provides summary information on patterns of support and application procedures through the U. S. Office of Education. It is a helpful reference guide for anyone involved in education research.

5. Grant Data Quarterly. (First four issues in 1967) Academic Media, Inc., 10835 Santa Monica Boulevard, Los Angeles, California 90025. Price: Single subscription $35.00 (10% off on 2 or more).

   The first four issues present detailed information on government support programs, business and professional organization support programs, and foundation support programs. This quarterly would be valuable as a reference source for college libraries, or progressive departments contemplating a substantial volume of research and development activities.

6. Catalogue of Federal Assistance Programs. Information Center, Office of Economic Opportunity. Executive Office of the President, Washington, D.C. 20506.

   First-rate description of all federal domestic agencies.

B. Partial List of Fund Sources for Educational Research

   1. Cooperative Research Program (H.E.W.)

      References: See College and University Reporter (C.C., Inc.), at 1651; 8911 and 8401. See Programs and Services of the U. S. Office of Education, H.E.W., pp. 223-224, and 231-232.

      a. The Cooperative Research Programs[1] administered by the U. S. Office of Education include support for both basic and applied research, demonstration project, and curriculum improvement projects. Funds were also authorized for the creation of research and development activities.

      b. Small contracts. Perhaps of greatest significance to the researcher just beginning a career in educational or behavioral research is the small contracts program. This program is intended to provide support for small-scale research or development projects which require less than $10,000 in federal assistance.

---

[1] Authorized by the Cooperative Research Act of 1954, with extension through the Elementary and Secondary Education Act of 1965.

The program "supports experimental research, surveys, demonstrations and curriculum studies," and also "assists in making exploratory studies designed to determine the feasibility of more extensive research on specific problems." (College and Univ. Reporter, 1651.)

Proposals are submitted through the Bureau of Research, U. S. Office of Education, (H.E.W.), through the regional offices. There are no specific deadlines for small contract proposals, and the program is designed so that proposals are processed with a minimum of delay. The reader should refer to the Office of Education pamphlet "Support for Research and Related Activities" for details on proposal format.

c. See references above for discussion of large-scale project or program support under the Cooperative Research Act.

2. Research in Educational Media (H.E.W.)

References: See College and University Reporter, (C.C., Inc.), at 1653-4, 8671 and 10,001. Also, for PL 89-209, see 9651.

a. Title VII of the National Defense Education Act of 1958 (PL 85-864 as amended) provides support for research and experimentation in more effective utilization of education media.[2] This includes television, radio, motion pciture films, slides, tapes, programed instructional devices, and other media designed to supplement instruction. Part A of Title VII provides support for research, while Part B includes authorization for dissemination of media (or information on application of new media).

b. This program, also administered by the Bureau of Research, provides for several types of grants--both large and small, and has wide applicability across subject-matter areas. It is possible to obtain support of $10,000 or less for projects designed to improve or evaluate media applications within a department or specific course (i.e., minimum generalizability). Small grant proposals are submitted to the U. S. Office of Education, Regional Offices.

---

[2] The National Foundation on Arts and Humanities Act of 1965 (PL 89-209) provides authorization for research on media in arts and humanities instruction.

3. National Science Foundation (N.S.F.)

References: See College and University Reporter, (C.C., Inc.), at 4568-73; and 9001-51.

a. The Social Science Division of the National Science Foundation provides support for basic research, for research in the anthropological sciences, economic sciences, sociological sciences, and research in the history and philosophy of science. Authorization is through the National Science Foundation Act of 1950 (PL 507).

b. Specific programs among those alluded to above require different guidelines, and each program includes individual deadlines for submitting proposals. Be sure to note deadlines in planning a proposal. You should count on several months of thought, writing, and criticism from colleagues in preparing a research proposal. This is extremely important.[3]

c. One recent emphasis with N.S.F. has been on science education, and there are now three divisions with the foundation which are responsible for the various science programs in education: Division of Pre-college Education, Division of Undergraduate Education, and the Division of Graduate Education. Each division is responsible for a variety of science education programs, many of which overlap those of the U. S. Office of Education.

4. Arts and Humanities Endowment Funds

References: College and University Reporter, (C.C., Inc.), at 1720, 1722, 1724, 9651.

a. Research support is available for both broad and specific studies, with emphasis upon American history and literature.

b. The Humanities Endowment cooperates with the U. S. Office of Education in sponsoring research on teaching of the humanities at the pre-college level.

c. Send for pamphlet: "National Endowment for the Humanities--Initial Programs," September, 1966, for additional information, or refer to source listed above.

---

[3] N. S. F. proposal guidelines may be found in College and University Reporter, beginning at 5401.

5. National Institute of Mental Health, (N.I.M.H.), Divison of the Public Health Service, H.E.W.

   *References: See College and University Reporter (C.C., Inc.), at 4125-8.

   a. The N.I.M.H. supports basic and clinical research relating to the etiology, diagnosis, treatment, and prevention of mental illness. Grants are also available for instructional research and teaching.

   b. The N.I.M.H. further provides funds in support of research on mental retardation. Support is available on the treatment, care, management, and training of mentally retarded in light of biological, psychological, or socio-cultural factors involved.

      *Future emphasis will be on research on child rearing practices, teaching programs, parent and child therapy techniques, and interdisciplinary approaches to treatment and rehabilitation. (4126)

6. Child Health and Human Development

   References: See College and University Reporter, (C.C., Inc.), at 4111.

   a. The National Institute of Child Health and Human Development, a division of the National Institute of Health (N.I.H.) of the Public Health Service (Dept. of Health, Education and Welfare), provides support for research and training related to "maternal health, prenatal care, child health, and human development." N.I.H. emphasizes research in four areas: reproduction, growth and development, mental retardation, and aging.

   b. Specific guidelines and notification of deadlines for proposals should be requested from the National Institute of Child Health, Public Health Service.

7. Other sources**

   a. Office of Naval Research (D.O.D.)

   b. Department of the Army (D.O.D.)

   c. Advanced Research Project Agency (D.O.D.)

d. Agency for International Development

**See Guide to Support Programs for Education, 3M Company, St. Paul, Minnesota, 1967.

**See also pp. 16 and 17 of the U.S.O.E. pamphlet "Support for Research and Related Activities" for programs administered by the Bureau of Research

ERIC Full Text Provided by ERIC

Appendix F

Shortcomings Found In Study-Section Review of 605 Disapproved Research Grant Applications, April-May, 1959

No.	Shortcoming	%
	Class I: Problem (58 per cent)	
1	The problem is of insufficient importance or is unlikely to produce any new or useful information.	33.1
2	The proposed research is based on a hypothesis that rests on insufficient evidence, is doubtful or is unsound.	8.9
3	The problem is more complex than the investigator appears to realize.	8.1
4	The problem has only local significance, or is one of production, or control, or otherwise fails to fall sufficiently clearly within the general field of health-related research.	4.8
5	The problem is scientifically premature and warrants, at most, only a pilot study.	3.1
6	The research as proposed is over-involved, with too many elements under simultaneous investigation.	3.0
7	The description of the nature of the research and of its significance leaves the proposal nebulous and diffuse and without clear research aim.	2.6
	Class II: Approach (73 per cent)	
8	The proposed tests, or methods, or scientific procedures are unsuited to the stated objective.	34.7
9	The description of the approach is too nebulous, diffuse, and lacking in clarity to permit adequate evaluation.	28.8
10	The over-all design of the study has not been carefully thought out.	14.7
11	The statistical aspects of the approach have not been given sufficient consideration.	8.1
12	The approach lacks scientific imagination.	7.4
13	Controls are either inadequately conceived or inadequately described.	6.8
14	The material the investigator proposed to use is unsuited to the objectives of the study or is difficult to obtain.	3.8
15	The number of observations is unsuitable.	2.5
16	The equipment contemplated is outmoded or otherwise unsuitable.	1.0
	Class III: Man (55 per cent)	
17	The investigator does not have adequate experience or training or both, for this research.	32.6
18	The investigator appears to be unfamiliar with recent pertinent literature or methods, or both.	13.7
19	The investigator's previously published work in this field does not inspire confidence.	12.6

Appendix F - Shortcomings Found (cont.)

20	The investigator proposed to rely too heavily on insufficiently experienced associates.	5.0
21	The investigator is spreading himself too thin; he will be more productive if he concentrates on fewer projects.	3.8
22	The investigator needs more liaison with colleagues in this field, or in collateral fields.	1.7
	Class IV: Other (16 per cent)	
23	The requirements for equipment or personnel, or both, are unrealistic.	10.1
24	It appears that other responsbilities would prevent devotion of sufficient time and attention to this research.	3.0
25	The institutional setting is unfavorable.	2.3
26	Research grants to the investigator, now in force, are adequate in scope and amount to cover the proposed research.	1.5

Source: Allen, Earnest M. (NIH) "Why Are Research Grant Applications Disapproved?" Science, Nov. 25, 1960, p. 1533.

ERIC Full Text Provided by ERIC

Appendix G

Example of an Experimental Research Proposal

I. Project Title

Directed Discovery vs. Programed Instruction: A Test of a Theoretical Position Involving Educational Technology.

II. Problem

Increasingly more importance is being placed on the process of learning by "directed discovery." Representative of this increasing emphasis is the newer techniques and materials presently under development by the UICSM (University of Illinois Committee on School Mathematics) directed by Professor Max Beberman. By requiring the learners to discover for themselves new relationships in mathematics, it has been demonstrated that high school students are capable of learning advanced concepts more effectively than otherwise. The act of discovery itself is considered to be the primary reason for the success of the new mathematics curriculum (Beberman, 1955; Hendricks, 1961).

Paradoxically, increasingly more importance is also being placed on the process of learning by programed instruction ("teaching machines"). "Discovery" by the student is minimized in programed instruction which characteristically presents the material in such small steps that the learner does not have to search far to find an answer. This is especially true when prompts and cues are added. Yet, the research evidence to date indicates that programed instruction is at least as effective as more conventional methods of instruction and may be far more efficient in terms of teaching time.

The fact that both the "discovery" and the "programed" instructional methods are presently in the forefront of attention serves to revive an older, unresolved research problem in a new context, and with greater potential for solution. The problem is to determine which process of learning is superior, a highly directed (formal) learning which places the learner in a position of complete dependence on the teacher, or non directed (informal) learning in which the learner must rely almost completely on his own cognitive capabilities. Advocates of the nondirected, informal (hereafter called "discovery") process claim a number of advantages, most of which are included in a recent article by Bruner (1961). Bruner hypothesized that learning by discovery benefits the learner in four ways: (1) it increases the learner's ability to learn related material, (2) fosters an interest in the activity itself rather than in the rewards which may follow from the learning, (3) develops ability to approach problems in a way that will more likely lead to a solution, and (4) tends to make the material that is learned more readily accessible in memory-that is, easier to retrieve or reconstruct.

Research evidence is not entirely supporting of the claims of Beberman and Bruner. The more recent evidence suggests that learning by discovery does not necessarily benefit the learner directly in terms of retention and transfer, but it does foster interest in the task (Kersh, 1958). When interest is generated, the learner tends to continue the learning process autonomously beyond the formal learning period. As the results of his added experience, the learner then raises his level of achievement, remembers what he learns longer and transfers it more effectively. However, sufficient interest to continue learning evidently does not appear unless the learner expends intensive effort without help for an extended period (Kersh, 1961).

In other words, the evidence referred to above suggests that learning by discovery is superior to highly directed, formalized learning only in terms of increasing student motivation to pursue the learning task. Learning with direction is far more efficient in that the student learns more during a given period of time and, when the direction is pertinent, comes to understand what he learns more completely. Obviously, programed instruction may be allied with highly directed, formalized teaching methods.

The explanation for the elusive drive generated by independent discovery is not evident, but several have been offered, including Zeigarnik effect of resumption of incomplete tasks (Kersh, 1961). Neither of the latter is entirely adequate because they simply describe the conditions under which the motivation may be expected to appear. Actually, the motivating effect may appear even when the tasks are completed. Another explanation, and the one which is of primary concern in this present proposal, is that the motivating effect is learned through a process of operant conditioning. By this theory, the learner, who is forced to discover the solutions to problems without help, engages in a kind of behavior often described as "searching." The searching behavior is reinforced by the teacher who is monitoring the learning process, and by the learner's own successful progress, towards a solution This explanation would support the claims that Bruner makes regarding the increase in the learner's ability to learn in related areas, to solve similar problems, and to continue the searching behavior beyond the formal learning period whether or not there are any extrinsic rewards involved. The theory also fits well with the research evidence which suggests that the motivating power of learning by discovery does not appear unless the learner engages in such searching behavior over a relatively long period of time.

If the above explanation is true, then it should not be necessary to rely completely on the less efficient, time-consuming discovery techniques. In short, it should be possible to employ the techniques of programed learning to achieve the same desirable benefits which may accrue from learning by discovery. Many of the techniques of programed learning are based on the operant conditioning model and, through the use of appropriate educational media, are adaptable to any teaching objective which may be translated into behavioral terms.

The purpose of this present project is to conduct an experiment to test the theory that a teaching (machine) program which is specifically designed to develop and foster the kind of searching behavior that characterizes the discovery process, will reproduce equally as well the desirable benefits to the learner which are usually attributed to discovery techniques alone.

If the experimental evidence substantiates the theory, it will warrant further study. The development and eventual application in the classroom of efficient and effective techniques for using the new media associated with programed learning will no doubt follow.

III. Objectives

The present project has two objectives:

1. To develop a "programed" modification of an existing unit of instruction (The Associative Law) modeled after the UICSM "discovery" method, and adapt it to a group-pacing technique which provides individualized feedback to the learner (see detailed description under Procedure, below).

2. To conduct an experiment with the programed method designed to test the following hyotheses:

   a. Students taught by the programed modification (Objective 1, above) achieve the learning objective more rapidly than students taught by the discovery method.

   b. Students taught by the programed method spontaneously employ the learned material as frequently after the formal learning period as students who are taught by the discovery method.

   c. Within a period of eight weeks following the formal learning period, students taught by the programed method are able to recall the learned material from memory and apply it in the solution of mathematical problems as well as students who are taught by the discovery method.

   d. In a task of new learning following within 24 hours after the formal learning period, students taught by the programed method reveal by their written work and self reports that they employ the same or the equivalent techniques of independent discovery and problem solving as students taught by the discovery method.

IV. Procedures

The project will be conducted in two phases: (1) materials development phase, and (2) the experiment. During the materials development phase, the necessary modifications to the discovery materials will be produced and the techniques developed for using them. During the same phase, the teachers for the experimental classes will be trained in the use of the materials and techniques. The first phase is expected to last approximately eight months. A more detailed description of the programed materials and techniques is given below.

The Experiment

Design. Two experimental groups of elementary school chiildren (grade 5) will be taught the Associative Law in different ways: one group by the Discovery method, and the other by Programed instruction. Two different instructors will be involved. Then the experiment will be replicated with two new groups of children. This time the same two instructors will teach, but they will exchange methods of instruction.

The counter-balanced design is diagramed below:

	Instructional Method	
	Discovery	Programed
I.	Teacher A n = 15	Teacher B n = 15
II.	Teacher B n = 15	Teacher A n = 15

N = 60

Selection of subjects. A total of 60 fifth graders will be selected from ABC Elementary School (and neighboring schools) on the basis of a pre-test of their knowledge of those number concepts and arithmetical operations that are pre-requisite to learning the Associative Law. To better insure that the children will be capable of learning the required concepts, their records of scholastic aptitude and achievement in arithmetic will be examined. Only children within prescribed limits of ability and achievement will be accepted. Thereafter, the group will be sub-divided into four groups of 15 by random processes, and each group will be assigned to one of the two experimental groups.

The actual instruction will take place on the campus of XYZ College. The physical facilities will be the same for each group with the exception of the special equipment used with the Programed Group.

Experimental techniques. The Programed Group will be taught in the specially equipped laboratory classroom. The laboratory is equipped with a 25-student-station TELETEST Communication System (see appendix). This system will permit the classroom instructor to present programed materials to the class and to provide individualized feedback to each student immediately after the student signals that he has made his response. The fact that each student has made a response and the particular response made is recorded in coded form immediately on an IBM (International Business Machine) type card at the instructor's station. The instructor, therefore, has the capability of determining, at any point and time, precisely what proportion of the class is ready to go on, and which individuals are progressing satisfactorily and which need help.

The programed method of instruction, using the apparatus described above, will not be the self-instructional, individually-paced approach which typically characterizes the teaching machine. This present technique uses the programed materials as an integral part of the more conventional approaches to classroom instruction. A teacher is present at all times, and group-pacing techniques are used. The teacher conducts the class by providing information, assigning problems, interacting with individuals, etc., as required. The unique characteristics of the proposed programed method are the following: (1) The teacher will follow a program of questions, problems, presentations, etc., which will employ such techniques as "vanishing" cues, small steps, branching, etc., as may be appropriate. (2) Each student will record his needs and reactions (e.g., "progressing satisfactorily," "need help," "repeat"), or his solutions to problems through the use of the communication system described above. (3) The teacher will receive and collate the students' reactions to questions and problems almost immediately and adjust his program accordingly. (4) The program of problems and tasks will require the students to employ problem solving and discovery techniques with the increasing frequency and on increasingly more difficult tasks. In the beginning the program will be highly directive in "teaching" the students principles for solving mathematical problems and using the inductive method.

The Discovery Group will be taught by techniques and materials modeled after the UICSM course of study. The materials for this present study were developed for the fifth grade level at ABC College.

The formal learning period is expected to last approximately ten hours distributed over two weeks. During the learning period, ea subject will be brought to the same level of achievement as indicated by pre-established criteria. The learning criterion will be in terms of each student's performance on problems requiring the application of the mathematical law to be learned. Since both teaching methods use group-pacing techniques, all students will complete at approximately the same time. However, in the Discovery Group especially, considerable freedom is allowed the individual to learn at his own rate, so every student will be given the opportunity to demonstrate his achievement on a short written examination whenever he and the instructor agree that he may be ready. When an individual demonstrates that he has achieved the required level, he will be excused from further attendance in the experimental classroom.

The entire experiment will be repeated twice with a three-month time interval as outlined in the experimental design, above.

Test of new learning. After each individual student completes the initial program of instruction, and within 24 hours thereafter, he will be taken aside individually and given a new task to learn. The task will be that of discovering a novel rule for adding a series of odd numbers. The rule is usually discovered within a period of 20 minutes. During the new learning period, voice recordings will be taken of the subject's verbal report of their thought processes, and any scratch work will be retained.

Post-test of recall. A post-test consisting of problems similar to the ones used during the learning period will be given to each subject within eight weeks following the formal learning period. One third of each group will be administered the post-test two days after the formal learning period, another third will be given the post-test after two weeks, and the final third after eight weeks. At the same time the post-test is given, each subject will be asked to fill out a questionnaire on his use of the learned material during the intervening period. They will be asked the number of times the rule was used spontaneously, and the purpose and occasion in each case.

Analysis of data. The data will be analyzed by comparing the two experimental groups in terms of time to learn, techniques employed during the new learning task, performance on the post-test, and use of the learned material during the period intervening between the learning period and post-test.

ERIC

Comparisons in rate of learning (Hypothesis a, above) will be in terms of the number of class sessions (or hours) to complete the learning task. Chi square will be used to test the significance of the obtained differences against a theoretical position of equal time to complete.

Comparisons in average frequency of use of the new learning (Hypothesis b) will be based on the questionnaire data in which each student will be asked to estimate the total number of times he used th nformation during the intervening period (together with corroboration information). If the frequency data meet the essential distribution and variance requirements, the mean differences between groups will be tested by the t test; otherwise, a non-parametric statistic such as chi square will be employed.

Hypothesis c, pertaining to memory for the task, will be tested with the post-test data. The tests will be scored on a pass-fail basis, since the concern is with memory for a mathematical law - not with computational accuracy or specific procedures which may contribute to overall test variance. Consequently, the chi square technique may also be employed in this analysis.

The data pertaining to the last hypothesis (d) will be in the form of frequency distributions of the techniques of proble solving and discovery used by the experimental subjects. The differences in the distributions for each group will be tested statistically by chi square, if suitable.

Approximate time schedule

The developmental phase of the project will last approximately eight months. The task of programing involves several try-out and revision cycles, each of which is slow and tedious.

The experiment will require eight months to conduct and two months for analysis and reporting.

Expected end product

The experiment will provide new evidence on the relative effectiveness of the methods of programed learning and discovery particularly as involves the theoretical position outlined above.

Publication plans

The results will be submitted for publication in the Journal of Educational Psychology. A discussion of the methods employed together with their respective theoretical basis will be submitted for publication in the Mathematics Teacher and related periodicals.

V. Personnel

Project Director - Joe Smith, Ph.D. (Educational Psychology, University of California, 1955) present position; Associate Professor of Education and Psychology, XYZ College. Research experience: Consultant in Educational Research, University of DEF, 1952-1954; recipient of three research grants from the Graduate School of the University of HIJ, 1955-1959; Human Factors Scientist, System Development Corporation, KLM, 1959-1960.

Director of Curriculum and Instruction - Bill Jones, M.S., University of Illinois, 1959) Assistant Professor of Mathematics, XYZ College. Related experience: Developed the grade 5 experimental materials and techniques at XYZ College which are modeled after the UICSM course of study and employ the discovery method. Completed a graduate course in mathematics instruction at University of Illinois taught by UICSM personnel. Five years teaching experience in public schools.

Experimental teacher I. Bill Jones, M.S. (above)

Experimentalteacher II. To be selected from the staff of the ABC Elementary School, the laboratory school for XYZ College. He will be instructed in the experimental techniques by Professor Jones.

Research Assistant - Don Green, M.A., Research Instructor, XYZ College. Related experience: Research assistant under the supervision of Dr. Joe Smith, 1960-62. Majored in Mathematics and Education.

VI. Facilities

The resources personnel, office and equipment of XYZ College. This includes modern motion picture production facilities, advanced electronic computer systems, completely equipped laboratory schools, and a distinguished faculty of behavioral scientists.

The resources of the ABC Elementary School, the Laboratory School on the campus of XYZ College.

VII. Duration

Total amount of time required: 18 months.

Beginning: January 1, 19xx

Ending: June 30, 19xx

BUDGET

Category	OE Funds Requested	Funds Contributed
I. Direct Costs		
Personnel		
Project Director (.25 FTE, 18 mos., 1st 6 mo. paid by State, next 12 mos. Fed. Funds)	$2,700	$1,300
Director of Curriculum & instruction (.25 FTE, 6 mos.	850	
Experimental teacher (one only, 1.0 FTE, 1 mo.)	600	
Research Assistant (.50 FTE, 15 mos.)	2,400	1,200
Secretary (.25 FTE, 12 mos.)	425	425
Programming consultant (one only 7 days @$50/day)	350	
Supplies and Materials		
Programming materials (mimeo., 35 mm. film, etc.)	75	
Instructional materials		25
Office supplies		50
Other Direct Costs		
Payroll assessments (6% salaries)	419	175
Depreciation on Teletest System (Total value of equip. $4,500 - Depr. est. for 1 1/2 yrs. for proportion of time equip. used on project)		100

(continued on next page)

Communications		$ 50
Travel (for experimental subjects)	$ 50	
Total direct costs	$7,869	$3,325
Indirect Costs (25% of direct costs)	1,967 -	none

Summary

Office of Education funds requested	$ 9,836
XYZ College contribution	3,325
Total cost of project	$13,161

Other Support

This proposal has not been submitted to any other agency or organization.

The research proposed herein is not an extension of or addition to a project previously supported by the Office of Education.

Estimated cost to Federal Government by fiscal year.

Category	Federal Funds Requested
Fiscal year 19xx	909
Fiscal year 19xx	8,927
Total Cost (all year)	$9,836

BIBLIOGRAPHY

Bruner, J.S. "The Act of Discovery", Howard Educational Review, Winter 161, 31, 21-32.

Beberman, M. An Emerging Program of Secondary School Mathematics. Cambridge: Harvard University Press, 1958.

Hendrix, Gertrude. "Learning by Discovery", The Mathematics Teacher, 1961, 54, 290-99.

Kersh, B.Y. "The Adequacy of 'Meaning' as an Explanation for the Superiority of Learning by Independent Discovery", Journal of Educational Psychology, 1958, 49, 282-292.

Kersh, B.Y. "The Motivating Effect of Directed Discovery", (accepted for publication, Journal of Educational Psychology, 1961).

Appendix H

Example of a Developmental Project Proposal

A COMPUTER-BASED TEST DEVELOPMENT CENTER

I. PROBLEM

Historically, formal assessment procedures have played an important role in public school settings. It would be difficult today to find a public school in the United States that does not use one or more standardized testing procedures during the course of its school year. Further, because of its inherent specificity and recent impact on educational practice, programed instruction and allied innovation have emphasized the need for more effective evaluation procedures. While the precise evolving role of testing is unclear, it is not difficult to demonstrate the value of well-designed tests in improvement of educational endeavors at any level.

Typical achievement testing programs in schools rely upon three sources for their tests: (1) teacher-made tests, (2) published tests, and (3) locally (i.e., school district) developed tests. Teacher-made tests often lack technical quality, are limited in range, and are useful only for small groups of students. Published tests, because they are usually written for national distribution, often lack content validity, are not appropriate for all learners within a school or for a specified homogeneous group of students, and frequently report norms inappropriate for a given school system. Locally constructed tests (as we now know them) while they can overcome many of the shortcomings of the teacher-made and published tests, are simply too costly (in time and money) for broad implementation in most school systems. Construction of good tests, if current procedures are employed, requires trained personnel, and considerable time unavailable to most school districts.

ERIC

The proposed project is based upon the premise that achievement testing will become increasingly important in the educative process. This importance will be evidenced in two ways: (1) achievement tests will be used increasingly for guiding specific learning experiences in individual learners, and (2) school districts will rely more and more upon tests for assessment and revision of total educational programs. Successful use of tests for these purposes is dependent upon the adequacy of tests available. It is unlikely that teacher-constructed tests (as we now know them) will ever attain the level of quality required for such purposes. In the first place, teachers generally do not have the training or experience in test construction, and while it might be possible for school districts to provide necessary training, the investment would not significantly upgrade the product. Secondly, teachers seldom have sufficient time to devote to preparation of tests that will ensure a quality product. Finally, it is probably inappropriate, in light of the myriad of functions which demand teacher time and effort, to expect or require them to build tests of the type that would provide them with desired diagnostic, prognostic, or survey information.

Present procedures for development of published tests do not permit the flexibility and specificity required to facilitate the testing functions now required by most users. Further, it is unlikely that test publishers can be expected, within the near future, to economically provide ready-made instruments for the wide diversity of uses required by teachers.

Locally prepared tests appear to be the most appropriate avenue through which to approach solutions to this problem. It is toward more

ERIC

effective, efficient, and economical development of locally constructed achievement tests that this project is directed. Successfully meeting teachers' testing needs will require test development procedures which (1) minimize the time lag between request for the test and its availability for use; (2) permit complete flexibility of test content, i.e., teachers and/or others can specify precisely the content of the test; (3) develop tests with adequate reliability; and (5) are economical. It is proposed that this task be accomplished through the use of electronic data processing equipment.

A computing system which could accomplish this task would be characterized by flexibility which would permit utilization by a wide variety of persons representing a wide variety of needs. This system should be useful to teachers as well as to individual school districts or to an even larger unit such as that represented by the submitting group. No matter what the group, however, it is envisioned that the test user would have his test generated by submitting (1) a list of objectives covered during instruction, (2) a list of pertinent characteristics that describe the learner group, and (3) an indication of desired test length. With this information the computer would then assemble an appropriate set of items, print these out along with directions for their administration in sufficient quantities for the teacher's use. This would be accomplished through computer retrieval of appropriate items from the large item pool which has been stored according to (1) the behavioral objective for which it is appropriate, (2) the student characteristics for which the item is appropriate, and (3) information regarding the quality of the item.

Basic components of the proposed test development system are, (1) a large pool of items classified according to curricular objectives, student characteristics, and information relative to the quality of the item itself, and (2) a "language" which will permit users to request construction of specific tests for specific learners.

Content validity of the retrieved tests would be ensured. An estimate of the reliability would also be provided. Reliability estimates of this type are possible because of the relationship that exists between item statistics and the reliability of a group of items having a given level of discriminability, all of which could be determined during development of the item pool.

The task of developing appropriate measures of student behaviors in instructional settings is a crucial one. The proposed solution; i.e. utilization of electronic data processing equipment, establishes this process within the greater complex of an evolving system of processes now known as Computer-Assisted-Instruction (CAI).

Three classes of activities may be identified within the general concept of CAI,

<u>Class I Activities</u>. Those in which the computer system's main frame performs the function of controlling the students' learning activities. These types of systems are characterized by learner-computer interaction.

<u>Class II Activities</u>. Those in which the computer system main frame augments or supplements the teachers' instructional capabilities. These types of systems are characterized by teacher-computer interaction.

<u>Class III Activities</u>. Those in which the computer system main frame

augments, supplements, or enhances the handling of administrative control and record data. These types of systems are characterized by administrator-computer interaction.

All three classes of activities are related in varying degrees of directness to the actual teaching function. All three classes of activities are relatively undeveloped. All three classes of activities have the potentiality to significantly improve the process of education in the United States. Initial efforts within all three classes of activities are now underway in the State; e.g., the Title III Computer Instruction Project in City A (Class I), the Title III OTIS project in City B (Class III), the Title III Oregon Compact project in City C (Class III), the Title III Computer-Based Test Development Center (Planning Grant Project preceding this proposal) in City D (Class II), and the City D Public Schools Cybernetics projects in science (Class II). Taken _in toto_ these efforts constitute a basic development effort unequaled anywhere. Developed _in toto_ the total end product far exceeds the sum of individual contributions on each. Steps have already been initiated to effect liaison between these projects which make joint utilization of efforts a reality.

The proposed project assists instruction in ways other than provision of good measuring instruments. In the first place, the project services will permit teachers to utilize time normally spent in test development for the preparation of learning materials and planning of learning activities. More importantly, however, the system will be able to provide direct assistance in the planning of instructional strategies. Such assistance will be possible as a result of the curricular analysis activities

that are correlated with the test item development and storage activities. Curricular analysis will consist of the systematic statement of all instructional objectives in major curricular areas included in elementary and secondary school programs. These instructional objectives will be organized into a behavioral taxonomy which identifies the logical inter-relationship of objectives within a given curricular area. Thus, a teacher might identify a given objective and request from the system a listing of those objectives which logically must precede and follow the given objective. With this type of information the teacher can begin to plan materials and experiences which will lead students toward mastery of each identified objective. In essence, therefore, the ordering of objectives becomes a roadmap for the entire instructional sequence. This type of planning has demonstrated its value in developing instructional sequences. Without assistance, however, it is a long, arduous task, and well beyond the capabilities (time and energy) of most teachers.

Schools within the project area budget approximately $2.00 per pupil per year for testing programs. For well-developed programs this budget provides up to four group measures of mental ability and six group measures of academic achievement for each pupil during his 12 year tenure in the school system. The proposed project will have, upon its completion, the capability to replace the achievement testing portions of such programs and will, in addition, provide significantly more capability in the achievement domain since (1) it may be used by individual teachers and (2) it will incorporate achievement objectives such as attitudes and values not now normally included in existing testing programs.

Once the proposed center is operational, it is anticipated that its service can be financed easily within the limits of testing budgets of the schools. However, during the developmental years, initial costs will exceed the capabilities of the schools, thus the request for Title III funds. This will be described more fully in Section VI of this proposal.

OBJECTIVES

The major objectives of the proposed project are as follows:

1. To identify the specific objectives of instruction in all elementary and secondary school curricular areas for all grade levels,
2. To develop a pool of test items appropriate for assessing the attainment of the instructional objectives identified under objective 1 above,
3. To design and operate the computer system necessary for the implementation of the test development activities,
4. To conduct the necessary in-service training required to ensure optimum utilization of the test development center,
5. To improve the over-all effectiveness of instruction in all curricular areas and grade levels represented in elementary and secondary schools in the participating districts.
6. To conduct the necessary informational and dissemination activities which will make this system a model for similar centers to be located throughout the United States.

Each objective is considered in detail below.

1. Curricular Objectives. Construction of sound measurement devices is dependent upon well-developed and concise statements of the behaviors to be tested. Given these conditions, test items can be readily developed. The core activity of this project is the development of such statements. Successful accomplishment of this objective will enhance the probability

ERIC

of success of the entire project immeasurably. The attempt to develop such statements for the entire curriculum has never been made. It is vitally needed, not only for this project but also for the field of education generally.

2. Item Development. The system is designed to include a core of test items appropriate for the myriad of objectives covered within the elementary and secondary school curricula. While it is impossible at this time to estimate the total number of items in the pool, the Center activities are designed to develop a minimum of 10,000 test items per year.

3. System Development. The system will be designed to permit users to interrogate (i.e., request tests) it in their own language; i.e., it will be a "natural language" system. During the first year the rudimentary system will be designed so that within the first year tests may be requested. Specifically, the system will be operational within the first year. During succeeding years the system will be enhanced to increase its efficiency and scope.

4. In-Service Training. No service is worth developing if its users are not trained in its use. During each of the first three years of operation, approximately 200 teachers will be directly involved with the Center. Key personnel in each participating district will be thoroughly trained in its use. Additionally, within each district, model schools will be established for purposes of demonstration and self-enhancement.

5. Improvement of Instruction. Through curriculum planning, teachers can more effectively meet the needs of their pupils. This system will be available to school personnel for such purposes.

6. <u>Dissemination</u>. The proposed Center is unique in education. The objective here is to make known through all appropriate channels the activities of the Center so that it may become a model for such efforts elsewhere.

II. PROCEDURES

The basic organizational structure of the Center is outlined in Figure 1.

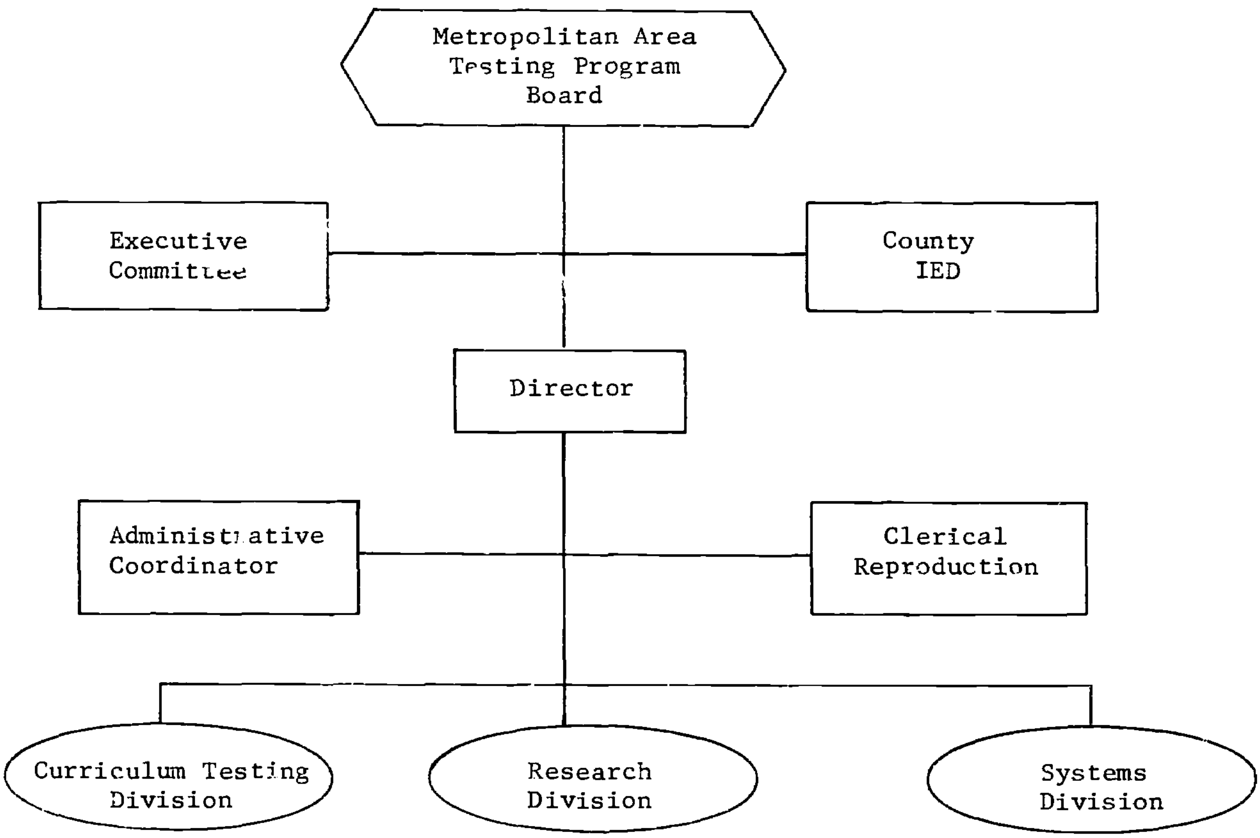


Figure 1. Organizational Chart - Computer-Based Test Development Center.

As depicted, the Center will operate as an activity sponsored by the Metropolitan Area Testing Program Board (MATPB). The Director will be responsible to MATPB through an Executive Committee elected by district representatives to MATPB.

All phases of the Center's operation will be monitored closely to determine the most effective and efficient basis on which to operate. The proposed plan is characterized by several important features:

1. Large numbers of teachers from all participating school districts are engaged in salient Center activities. This sets the stage for full-scale utilization of the Center's services as well as bringing a wide diversity of talent to the project.
2. The central staff is small. This reduces the necessity for large space needs and permits constant staff interaction and cooperation on all phases of the Center's operations.
3. The funding base is diverse and heavily dependent on outside (non-Title III) funds. This keeps the central staff active and creates a necessary urgency for continual searching for more effective ways to complete the job.
4. The functions are phased to permit maximum utilization of staff time, thus reducing costly overstaffing and slack periods.

The various components of the Center's operations are portrayed graphically in Figure 2. As indicated by the broken lines, some components are reserved for latter stages of the Center's operations; however, all will eventually be developed.

Each division of the Center will have its own unique set of objectives, which, when accomplished, will permit attainment of the overall objectives of the Center. Directors of each division in association with the Center Director and Administrative Coordinator will compose the operational directorate of the Center. The activities of each division are described, in turn below.

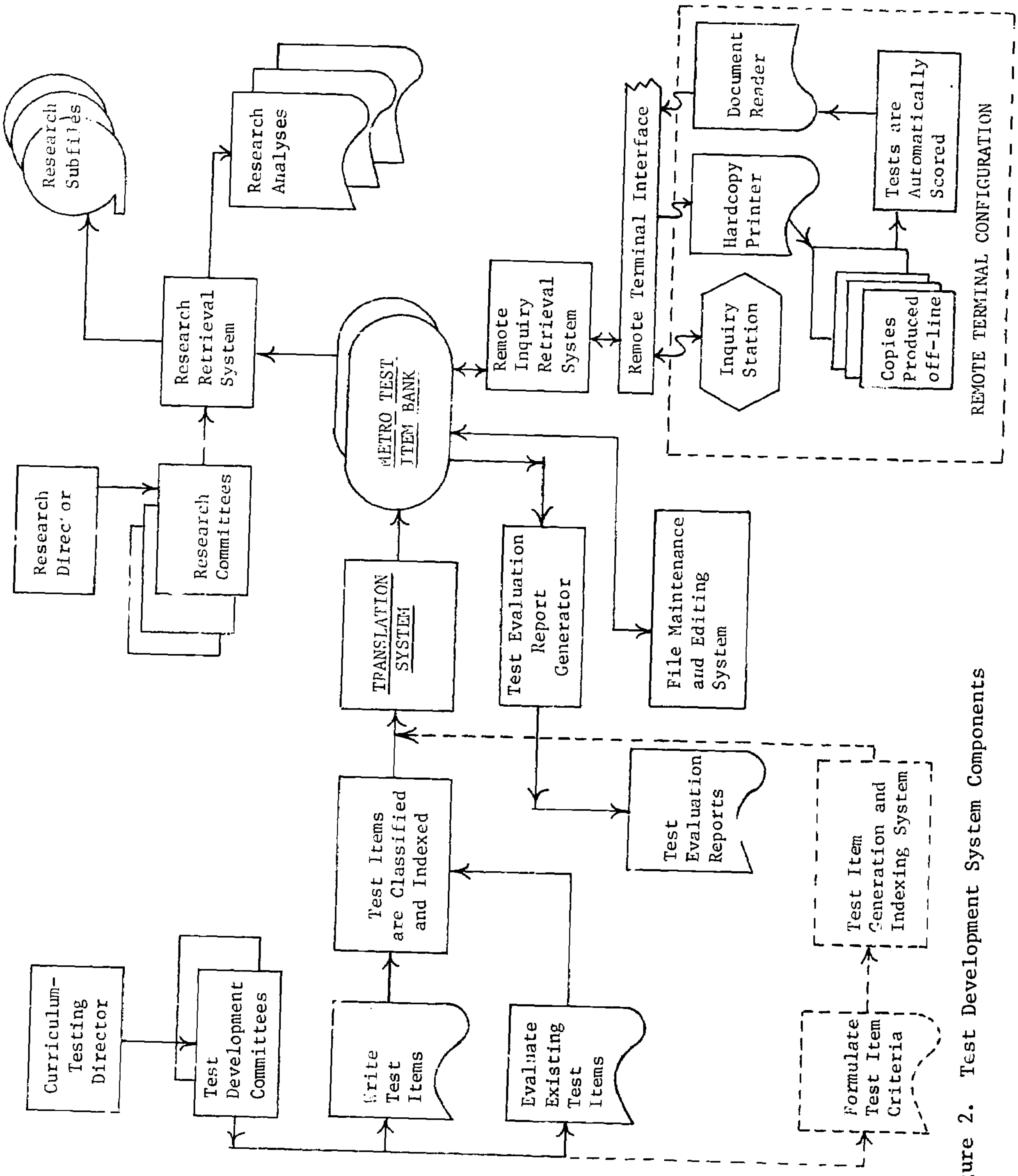


Figure 2. Test Development System Components

1. <u>Curriculum</u> - <u>Testing Division</u>

This Division is responsible for (1) the determination of behavioral objectives for various curricular components, (2) development of test items and/or procedures for assessing those objectives, (3) validation of test items, and (4) organization of objectives into a useful and meaningful behavioral taxonomy.

To accomplish the first two objectives (statement of objectives and item construction) the Director of the Division will organize work sessions for masterful teachers representing participating school districts. These work sessions will cover a period of four working days, and will be conducted continually during the course of the project. Each work session will be attended by 20 teachers, who will work in five-man teams supervised by a specialist in the curricular area being covered.

During August of each year, four work sessions will be conducted; one each week. Additional work sessions will be conducted on Saturdays during September, October, January, February, March, and April of each year.

Each work session will be devoted to the following activities:

1. Orientation to the project (1/2 day),
2. Training in preparing objectives (1 day),
3. Stating objectives (1 1/2 days), and
4. Writing test items (1 1/2 days).

Periodically, groups of items will be administered to various target groups for purposes of item validation. The Division Director and Psychometrist will be responsible for identifying appropriate trial samples and for completing the selection and revision of trial items. Test items thus validated will be placed in the item pool.

The Director will also organize a taxonomical team composed of his supervising specialists and other Center personnel to accomplish the final objective.

2. Research Division

This Division is responsible for (1) identification, design, organization, and completion of basic measurement research required to maintain and extend the functions of the Center, (2) acquisition of funds to conduct this research, and (3) development of basic test development specifications for the Center's systems functions.

Initially research will be conducted to delineate student characteristics which affect test-taking behavior and therefore act as determinants of item form. Additional research topics include, but are not limited to:

1. Estimating test reliability from item statistics,
2. Item discrimination indices for criterion and normative function,
3. Estimation of test norms from item trial with pseudo samples.
4. Differential characteristics of test item which do and do not survive item analysis.
5. Computer generation of test items.

Because of the basic nature of much of this research, funds will be sought from additional sources to support the activity. Initial funding from Title III sources is sought only to provide a sound base and to insure continuity.

The Research Director will consolidate all additional fund-seeking activities and will act as the design consultant for all research propo-

sals initiated in the Center. Additionally, this unit will keep Curriculum-Testing Division personnel updated in current testing approaches so as to insure maximum effectiveness of test items produced. Finally, the Research Division shall be responsible for design and completion of all phases of the evaluation of the Center's activities.

The Director will be assisted by two or more project personnel and will have as a specific objective the development of the Research Division so that it may attract a minimum of $100,000 in research grants each year.

3. <u>Systems Division</u>

This Division is responsible for (1) development of all software components of the test development system, (2) operations of the system, (3) processing of all requests for tests, and (4) completion of such research as is required within the general realm of computing hardware/or software.

The Systems Director will supervise a staff of programmers, machine operators, and key punch operators. During the first year basic software elements will be developed for use on the IBM 360/30 system. During succeeding years the additional elements will be programmed. The activities of this Division will be highly oriented to development during the first two years of operation. This orientation will gradually shift to one of service and updating after the second year.

EMPHASIS

This project is considered to be innovative. As previously described, the general problem is one of improving the measurement capabilities of

various types of school personnel involved in instructional decisions. Improvement in these capabilities results from the immediate availability of well-constructed tests designed for the specific use of the tester. The project brings to bear all the sophistication in test construction that has been developed over the past decade and, through the high-speed capabilities of the computer makes this sophistication-available to the educational practitioner. Additionally, through the creative use of the system's curricular features, practitioners may more effectively plan relatively long instructional sequences.

EVALUATION

Project evaluation will be a continual process directed towards determination of the Center's effectiveness and efficiency. Project efficiency will be determined through evidence gathered relative to specific operational goals. These goals can be firmly identified only upon initiation of the project since experience with this class of functions is not yet available. The goals will be in the form of operational standards; e.g., per cent of requests handled within one day, relationship of staff size to number of requests etc.

Project effectiveness will be determined through answers to the following types of questions.

1. Do test users characteristically identify 90% or more of test items provided as relevant to their requests?

   During the tenure of this project each test user will be asked to judge the relevancy of each item on each developed test. System modification will be accomplished until all tests reveal a minimum level of 90% relevant items.

2. Do the majority of teachers in appropriate subject matter areas in participating schools routinely utilize the test development services of the Center?

   This is a crucial question. An affirmative reply will indicate that teachers (1) do utilize the service and (2) that once they have initiated contact with the Center they continue to use its services. Utilization files will be maintained of all users. Utilization patterns will be analyzed for determination of the answer to this question.

3. Is there evidence of increased school participation in the project? The Center's goal is to provide its services to all schools in the four county area. The project will be effective if increasing members of districts request inclusion in the project.

4. Are Center-produced tests judged to have superior validity when compared to published tests by test users?

   Samples of users will be asked to compare the validity of tests provided to them by the Center with such published tests that would ordinarily be used in place of Center-produced tests.

5. Are test utilization patterns of participating schools superior to patterns in nonparticipating schools?

   Participating and nonparticipating schools will be matched on such variables as size, type of locale, and per-pupil expenditure. Comparisons of test employment will be made to determine types of tests used, frequency of use, and testing purposes.

6. Do teachers in participating schools report increased capacity to provide improved instructional practices?

   Samples of users will be interviewed by neutral interviewers to identify their reactions to the program. Care will be taken to gather candid evidence of the effects of the Center on the teachers' actual work day.

7. Do the curricular objectives covered in the Center significantly exhaust each curricular area?

   When the Center staff is reasonably satisfied with its efforts in a given area, the objectives will be presented to an independent panel of experts in that area for study. The staff's opinion must be verified by the expert panel.

8. Are curricular objectives developed in the Center accepted, in substance, by other groups?

   Evidence to answer this question will be in the form of the extent of requests for the Center-developed lists of objectives coupled with indications of adoption, for various purposes, of the objectives by requesting groups.

9. Are testing programs of participating schools expanded as a result of Center services?

   It is expected that the existence of the Center will permit participating schools to expand their testing activities without additional costs to them. All participants will be surveyed continually to determine the nature and costs of their testing programs.

ERIC

10. Do teachers creatively utilize the total features of the system provided?

    Unique uses of the Center's services will be documented and collated to provide users with suggestions for ways in which teachers and other types of users may utilize the system.

It is realized that each question will require its own unique type of data and collection. The entire evaluation effort will be handled within the Research Division and will employ the techniques of design, data collection, and analysis most appropriate.

DISSEMINATION

Three avenues of dissemination will be instigated: (1) quarterly newsletter and annual report, (2) reports and symposia at professional meetings, and (3) an annual 5 day conference for 25 to 30 participants.

Newsletters and annual reports will be distributed to various local, state, and national agencies. The distribution list will not exceed 250 users. Efforts will be made to report various phases of the Center's activities at such meetings as AERA, NCME, OEA Research Conference, etc.

During July of each year, an invitational conference will be held. Participants will be limited to 30 and will be invited from representative school districts throughout the nation. Participants will be expected to provide for their own expenses.

III. PERSONNEL AND FACILITIES

Qualifications of Professional Personnel

1. Center Director. Dr. Joe. R. Doakes (Ed.D., University of Neb-

braska, Educational Psychology and Measurement). Present position: Associate Research Professor, Somewhere College of the State System of Higher Education (19xx present). Previously, Assistant Professor of education, University of Toledo, 19xx-xx, Instructor in Educational Psychology, University of Nebraska, 19xx-xx, Dr. Doakes has taught in the public schools and has done much research in psychological scaling, predictive measurement, instructional simulation, computer simulation, and communication. All of the above-listed areas were sponsored through grants received from various agencies.

2. Research Director. Dr. John R. Doe, Jr. (Ph.D., Michigan State University, Educational Research Design and Development, 1965). Present position: Assistant Research Professor, Somewhere College, (19xx-present). Principle Investigator for two USOE grants in area of programed instruction. Has presented papers at various material meetings including "Helping Faculty Members Specify Objectives" at AERA in 19xx, conducted workshop on Evaluation in the Elementary Language Arts; worked with measurement and evalution problems at all levels of education.
3. Curriculum - Testing Director. Dr. Bill X. Green (Ed.D., Stanford University, 19xx. Present position: Assistant Research Professor, Somewhere College. Has been an elementary classroom teacher for four years, and college instructor in reading methods courses and supervisor of student teachers for four

years; since 1963 has been active in development of instructional simulation materials for reading methods courses; currently is project director for project supported by USOE grant concerned with the development of low cost simulation instructional materials.

4. Systems Director. The Systems Director is the administrative head of the Systems Group and will make all final decisions concerning that group. Responsibilities include: Report to Project Director, communicate and plan with other Directors, control all systems work, broad direction of operations, supply broad direction of operations, supply broad technical guidelines for technical staff, write job descriptions, recommend personnel hired, request services from other groups, public relations, make assignments of Applications Programmers. Qualifications: To be able to establish working rapport with other personnel, extensive data processing experience, education in data processing, knowledge of behavioral sciences, knowledge of statistics, directed or participated in a major software effort, and management abilities.
5. Senior Development Programmer. Responsibilities: Development of specific technical guide lines and specifications, make all specific programing assignments, see that documentation for operations is developed, review of system performance and requested changes, work with the systems coordinator in selection of personnel, and maintain systems software. Qualifications:

Qualified in assembly language programing, participated in major software development such as a compiler, education in data processing, experience in Natural Language Processing (exposure and understanding of problems), and leadership qualities.

6. Developmental Programmer. Responsible for design of larger segments of system software. Qualifications: B.A. degree, training and experience in system layout and basic programing.
7. Psychometrist. Responsible for directing and supervising a try-out of test items in selected school districts and communicating the results of try-out to the Curriculum and Test Coordinator. Qualifications: A master's degree in testing and/or guidance who has experience in the administration and scoring of tests. A person who is able to interact effectively with teachers and administrators and who has supervisory experience or qualitites.
8. Administrative Coordinator. Responsible for all accounting, purchasing, contract negotiation, salary administration, physical plant, nonprofessional personnel. Qualifications: B.A. degree in Business Administration with accounting training. Experience in supervision, office management, account management.
9. Machine Operator. Responsible for operation of computing and peripheral equipment. Qualifications: Experience and training in machine operations.

<u>Facilities, Equipment, and Materials</u>

The Center will be housed in a small school building (approximately 5,000 square feet) located in East A B C County. The applicant agency

retains sole ownership and will make the facilities available to the Center at no cost.

Arrangements have been initiated to secure computing services from the Nearby Regional Educational Laboratory at no cost to the Center. The majority of office equipment will be obtained from Federal and State surplus materials centers at no cost to the Center. The only equipment for which funds are requested are those types of equipment which experience has revealed are inferior at surplus centers.

ERIC

IV. BUDGET

Proposed Budget Summary No. 1, for Title III, P.L. 89-10 Funds

Name and Address of Applicant: ABC County IED, P.O. Box 9172, City, State

Grant Period beginning: July 1, 19xx and ending June 30, 19xx

	Administration (100)	Fixed Charges (800)	Capital Outlay (1230)	TOTAL
Salaries				
Professional	$23,900			$23,900
None Professional	4,200			4,200
Contract Services	23,170			23,170
Materials and Supplies	2,100			2,100
Travel	4,160			4,160
Equipment			5,210	5,210
Other Expenses	2,000	3,743		5,743
TOTALS	$59,530	3,743	5,210	$68,483

Schedule A

FY 19xx-19xx

Expenditure Account No. 100	FTE	RATE	Title III Funds	Other Funds
1. Salaries				
Center Director	1.00	$14,160	$ 7,080	$ 7,080
Curriculum-Testing Director	.50	12,480	6,240	
Research Director	1.00	11,500	5,750	5,750
Systems Director	.50	14,000	7,000	
Administrative Coordinator	.50	8,000	2,000	2,000
Psychometrist	1.00	10,000	5,000	5,000
Sr. Developmental Programmer	1.00	13,000	6,500	6,500
Research Assistants	1.50	7,200	5,400	5,400
Clerical	3.00	4,200	6,300	6,300
Consulting Teachers (400 days @ $40/day)				16,000
Consulting Specialists (100 days @ $50/day)				5,000
Technical Consultants (20 days @ $100/day)				2,000
Subtotal Salaries			$51,270	$61,030
2. Materials and Supplies				
Consumable Office Supplies (Paper, Stationery, desk sets, envelopes, IBM cards, etc.)			2,100	
Subtotal Supplies and Materials			$ 2,100	
3. Travel				
In-State Travel (40,000 miles @ $.10/mile)			4,000	
In-State Per Diem (Meals and lodging, 10 days @ $16.00/day)			160	
Out-of-State Travel (10 trips for technical consultants)				2,100
Out-of-State Per Diem (30 days @ $21/day)				630
Subtotal Travel			$ 4,160	$ 2,730

4. Other

Telephone Communication and Postage	1,500	
Computing Services (Main frame rental)		3,750
Report Printing and Duplication	500	
Subtotal Other	$ 2,000	$ 3,750
TOTAL Account No. 100	$59,530	$67,510

Expenditure Account No. 800

Employee Benefits (7.3% of salaries except consultants	$ 3,743	$ 2,776
TOTAL Account No. 800	$ 3,743	$ 2,776

Expenditure Account No. 1230

3 Steel Secretarial Desks	$ 300
3 Secretarial Chairs	120
20 4-drawer Steel File Cabinets	1,200
5 IBM Model 224 Dictation Units	2,125
3 IBM Model 212 Transcription Units	1,215
1 Spirit Process Duplicator	250
TOTAL Account No. 1230	$ 5,210

Appendix I

Example of an Informal Developmental Proposal

A PROPOSAL FOR THE DEVELOPMENT OF CULTURALLY RELEVANT LEARNING MATERIALS

Problem

The Localia School District has a relatively high proportion of Indian-American students. It is adjacent to the other school districts which enroll the students from the large Localia Indian Nation. The educational materials presently used with Indian-American children are based on the typical American Middle-class culture, and have little relevance to the Indian child. Although "culturally deprived" in the economic sense of the phrase, these Indian children are not without a culture as the term implies, but in fact have a deep cultural heritage which is in imminent danger of disappearing.

There is a lack of effective educational materials and activities for children which are relevant to and reflect the cultural mores and values of the Indian-American, and which are at a relatively low reading grade level. This lack is apparent across the nation. Provision of such materials would be of value not only locally but would meet an unfilled need in other areas of the United States with Indian students.

Furthermore, there are few materials which present a realistic perspective of Indian values and culture to the student from either typical middle-class or from other minority group backgrounds. The Localia School District has initiated the development of Indian cultural materials through the efforts of John Doe and George Jones, a member of the Localia Nation. Mr. Jones has recorded many of the legends of his people and has made numerous presentations to students accompanied by his original art work.

The present proposal will extend these initial efforts through the use of both a systems approach to the development of the materials and the employment of a variety of media to build self-instructional learning packages.

Objectives

1. To design and develop supplementary curricular materials specifically for the Localia Indian school population which are culture-specific and media oriented. These would be developed and evaluated with the cooperation of a local advisor group. They would be packaged in a variety of formats, such as films, illustrated books with audio tapes, slide tapes, educational games, and would be adaptable to more than one grade level.

2. To develop such materials so that learners achieve stated behavioral outcomes in both cognitive and affective domains.

3. To develop teachers manuals to accompany each of the above sets of culturally relevant materials.

4. To develop such materials, including teachers manuals, using pilot test, revision, field test cycle so that evidence of the learning outcomes of the materials will be provided.

Materials

A typical set of such materials might consist of:

1. A slide-tape presenting an Indian cultural legend. Visuals could include original Indian Art and colored slide photographs of relevant scenes.

2. A set of alternative follow up activities:

   a. A simulation learning game
   b. Materials with which students can construct some objects, scenes of characters of the legend; or, ways of artistic expression of some aspect or impression of the legend
   c. Creative dramatic exercises, both pantomime and oral, stemming from the legend
   d. A set of picture cards arranged as a jigsaw puzzle. When completed they portray a recapitualation of the story.

3. Behavioral objectives of the unit: for example:

   a. Words, concepts, principles, and schema the child will be able to identify
   b. Descriptions the child will be able to give
   c. Solutions, to selected problems, the child will be able to demonstrate
   d. Attitudes and interest the child will demonstrate

All such objectives will be accompanied by relatively simple indicators so that a teacher can assess the degree to which any student has attained the objectives.

4. A brief teachers manual explaining:

   a. The objectives
   b. The nature of the materials
   c. The field test evidence
   d. Directions for using the materials
   e. Suggestions for further supplementary activities and materials

ERIC

Each set will be organized so that alternatives in presentation time and depth available to the teacher, e.g. from thirty minutes to several days.

Nine sets of materials will be developed.

Procedures

1. Briefly assess present prototype materials.

2. Identify and select a local Indian Advisory Group to advise on the authenticity and acceptance by their people of each package.

3. The cycle for each learning package will consist of the following steps:

   a. Formulate pilot materials.
   b. Present pilot materials to Advisory Group.
   c. Pilot test, revise, and field test.
   d. Place revised materials in school district and present workshop to district personnel explaining materials, objectives, evaluation, and how the materials can be used and adapted.

4. Prepare slide tape and written report for dissemination; one copy of each to State Department of Education.

Time Schedule

The total time for this project would be from July 1, 19xx to May 30, 19xx. Planning and development of initial packages would take approximately two months, so that the first materials might be expected to be in Localia schools by September, and subsequent materials would continue to be provided at monthly intervals.

The development and evaluation of these materials will be undertaken by the Local Research Center in cooperation with the Localia School District.

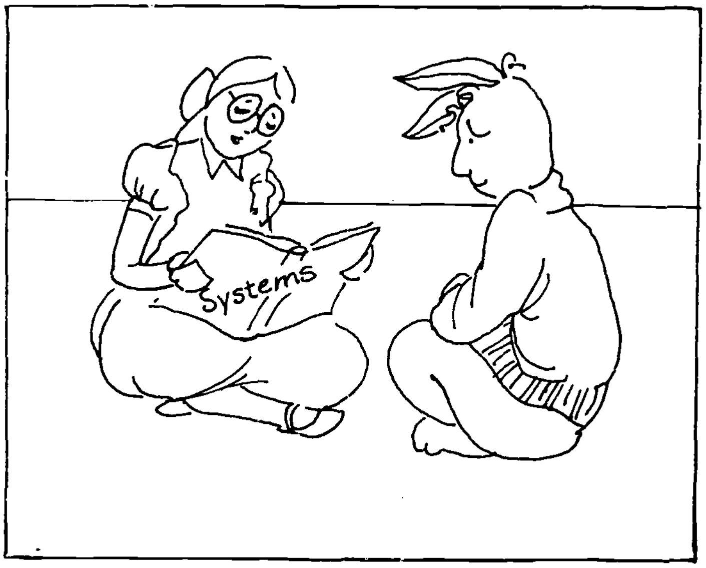
The Localia Research Center will develop and evaluate the materials through the field test stage. The Localia School District will provide classes for the pilot and field test of the materials.

Proper credit for the development of these materials will be given to George Jones and the Localia School District. These materials will be made available, at cost, to other districts in the State.

ERIC

Budget

I.	DIRECT COSTS		
	A. Personnel: Production and Evaluation	$4,789.00	
	B. Employee Benefits @9%	431.00	
	C. Travel and per diem	995.00	
	D. Materials and Rentals	1,420.00	
	E. Communication	150.00	
	F. Data Reduction and Analysis	250.00	
II.	TOTAL DIRECT COSTS		$8,035.00
III.	INDIRECT COST @24.39% OF DIRECT COSTS		1,960.00
IV.	Total Costs		$9,995.00



The dream child moving through a
land of wonders wild and new....
and half-believed it true.

the end

ERIC