DOCUMENT RESUME

ED 080 592 TM 003 109

AUTHOR Hambleton, Ronald K.
TITLE A Review of Testing and Decision-Making Procedures for Selected Individualized Instructional Programs.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE [72]
GRANT OEG-0-72-0711
NOTE 45p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Bibliographic Citations; *Criterion Referenced Tests; *Individualized Programs; *Instructional Programs; Models; *Performance Tests; *Testing; Tests

ABSTRACT

Results are presented of an investigation made to (1) provide a description of the testing models that are currently being used in selected individualized instructional programs, (2) compare three programs along the component parts of the testing model, namely, selection of a program of study, criterion-referenced testing on the unit objectives, assignment of instructional modes, and final year-end assessment, and (3) briefly outline several promising lines of research in connection with the testing methods and decision procedures for individualized instructional programs. The three programs selected for study were: Individually Prescribed Instruction, Program for Learning in Accordance with Needs, and Mastery Learning. An introduction, which includes a brief history, the content areas covered, and an indication of the extent of implementation, is provided for each instructional model. In addition, a description of each instructional paradigm and details on the testing model are provided. An attempt is made to pinpoint the decision points in each model, spelling out the consequences of the various possible actions in relation to each of the "possible true states of nature." A lengthy list of references is included. (DB)

A Review of Testing and Decision-Making Procedures for Selected Individualized Instructional Programs[1]

Ronald K. Hambleton
University of Massachusetts

I. Introduction

1.1 Background

While the idea of developing instructional programs in our schools to meet individual student needs is not a new theme in American education (see, for example, Washburne, 1922; and Wilhelms, 1962), it has only been in the last decade that such programs have been implemented on any large-scale basis in the schools.

The basic argument in favor of individualizing instruction comes from a multitude of research studies that suggest that students differ in interests, motivation, learning rate, goals, and capacity for learning among other things; and, therefore, grouped-based instruction on a common curriculum is inappropriate to meet their educational needs. That change in our schools is necessary is obvious when one notes that schools provide successful learning experiences for only about one-third of our students

[1]The research reported herein was performed pursuant to Grant No. OEG-0-72-0711 with the Office of Education, U.S. Department of Health, Education, and Welfare, Melvin R. Novick, Principal Investigator. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

[2]The author would like to acknowledge the insightful comments and constructive criticisms of Melvin R. Novick of The American College Testing Program on earlier drafts of the manuscript. In addition, Richard Ferguson and Roy Williams provided many useful suggestions.

(Block, 1971). On the basis of Project TALENT data Flanagan, et al. (1964) reported that our current instructional programs are inadequate to handle the large individual differences in any age or grade group. In addition, schools generally fail to help the student develop a sense of responsibility for his educational, personal, and social development or to make realistic educational decisions and choices about his future.

This trend toward individualization of instruction in education has resulted in the development of a diverse collection of attractive alternative models (see, for example, Gibbons, 1970; and Heathers, 1972) that, according to their supporters, offer new approaches to student learning which can provide almost all students with rewarding school experiences. These include: Individually Prescribed Instruction (IPI) (Glaser, 1968, 1970), Program for Learning in Accordance with Needs (PLAN) (Flanagan, 1967, 1969), Computer-Assisted Instruction (CAI) (Suppes, 1966; Atkinson, 1968; Atkinson and Wilson, 1969), Individualized Mathematics Curriculum Project (De Vault, Kriewall, Buchanan, and Quilling, 1969), and Mastery Learning (Carroll, 1963, 1970; Bloom, 1968; and Block, 1971). All of the models, as well as many others, represent significant steps forward in improving learning by individualizing instruction. They strive to actively involve the student in the learning process, allow students in the same class to be at different points in the curriculum, and permit the teacher to give more individual attention.

In matters pertaining to these models, for example, the construction of instructional materials (Popham, 1969; Smith, 1969), curriculum design (Wittrock and Wiley, 1970), and computer management (Baker, 1971; Cooley and Glaser, 1969), there is a substantial body of knowledge. It is perhaps surprising to note then that the amount of information currently available on the testing methods and decision procedures for these programs is quite limited. It is this component that, in principle, facilitates the efficient movement of students through the instructional program.

One reason for a lack of information is that measurement requirements within the context of many of the new programs require new kinds of tests. These are the criterion-referenced tests which are constructed and interpreted in ways quite different from the norm-referenced tests which are more familiar to most practitioners in the field (Popham and Husek, 1969; Glaser and Nitko, 1971; Hambleton and Novick, 1973).

Since one of the major purposes of individualized programs is to maximize the opportunity for all students to learn, it follows that tests used to monitor student progress should be keyed to the instruction. Further, they should provide information that can be used to measure progress along an absolute ability continuum. Norm-referenced tests are constructed specifically to facilitate making comparisons among students; hence, they are not very well suited for making most of the decisions required in individualized instructional programs.

1.2 Criterion-Referenced Testing and Measurement

Much of the discussion in the area of criterion-referenced testing and measurement (for example, see Block, 1971; Ebel, 1971; Glaser and Nitko, 1971; and Hambleton and Novick, 1973) stems from different understandings as to the basic purpose of testing in the instructional models described in the previous section. It would seem that in most cases the pertinent question is whether or not the individual has attained some prescribed degree of competence on an instructional performance task. Questions of precise achievement levels and comparisons among individuals on these levels seem to be largely irrelevant. In many of the new instructional models, tests are used to determine on which instructional objectives an examinee has met the acceptable performance level standard set by the model designer. This test information is usually used immediately to evaluate the student's mastery of the instructional objectives covered in the test, so as to appropriately locate him for his next instruction (Glaser and Nitko, 1971). Tests especially designed for this particular purpose have come to be known as criterion-referenced tests. Criterion-referenced tests are specifically designed to meet the measurement needs of the new instructional models. In contrast, the better known norm-referenced tests are principally designed to produce test scores suitable for ranking individuals on the ability measured by the test. A very flexible definition of a criterion-referenced test has been proposed by Glaser and Nitko (1971): "...[a test] that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specified performance standards." According to Glaser and Nitko (1971), "The performance standards are usually specified by defining some domain of tasks that the student should perform. Representative samples of tasks from this domain are organized into a test. Measurements are taken and are used to make a statement about the performance of each individual relative to that domain." Distinctions between norm-referenced tests and criterion-referenced tests have been presented by Glaser (1963), Glaser and Nitko (1971), Livingston (1972), Popham and Husek (1969), Ebel (1971), Block (1971), Hambleton and Gorth (1971), and Hieronymous (1972).

Hambleton and Novick (1973) have discussed the evaluation of criterion-referenced tests in practical situations. In their formulation, reliability takes the form of an index indicating the consistency of decision making across parallel forms of the criterion-referenced test or across repeated measurements. Validity takes the same form except, of course, that a new test serves as criterion. Both reliability and validity concepts are reformulated in straightforward decision-theoretic terms. However, at this stage of the development of a theory of criterion-referenced measurement, the establishment of cut-off scores is primarily a value judgment. [Further clarification is provided by Hambleton and Novick (1973) and Block (1972).]

1.3 Instructional Models Under Consideration

The major concern in this paper is with instructional models that include a specification of curricula in terms of behavioral objectives, detailed diagnosis of the entering competencies of students, the availability of multiple instructional resources, individual pacing and sequencing of material, as well as the careful monitoring of student progress.

In the programs under consideration, Computer-Managed Instruction (CMI) is an optimal feature. Under CMI the goal is for the computer to service classroom terminals which assist the classroom teacher in assessing a student's strengths and weaknesses, and to prescribe instructional sequences (Cooley and Glaser, 1969). Project PLAN and CAI are implemented in a CMI mode whereas IPI and Mastery Learning are not.

In summary, the goals of individualized instructional programs developed along the general lines of the specifications above are to enable students to work through the units of instruction at a pace reasonable for them, to develop self-direction and self-initiation, to encourage self-evaluation as well as motivation for learning, and to demonstrate mastery in a variety of skills.

Cronbach (1967) reported on three major patterns of dealing with individual differences which provide a framework for the models considered in this paper. Patterns of dealing with individual differences in the school can be described in terms of the extent to which educational goals and instructional methods are varied. In one pattern, the educational goals and instructional methods are relatively fixed and inflexible. Individual differences are handled mainly by dropping students from the program when they begin to encounter difficulty. In a second pattern, goals are selected for students on the basis of interest and potential. They are then channeled into one fixed program or another. Individual differences are handled by providing multiple optional programs. The models we describe in this paper fit into a third pattern where goals and instructional resources are individualized for the purpose of maximizing learning.

1.4 Purposes of the Investigation

The success of individualization depends to a considerable extent on how effectively teachers and students make decisions as to the mastery of specific instructional objectives, the development of individual prescriptions, the selection of instructional resources, etc. However, various writers including Baker (1971) and Glaser and Nitko (1971) have commented rather critically on existing testing techniques and procedures. Relevant background for improving such a situation would certainly include a review of the testing models of some of the more commonly used individualized instructional programs. Such a review would assist in defining the kinds of decisions that are made, and the information on which the decisions are based. This should provide a basis for developing testing methods and decision procedures specifically designed for use within the context of these models. (Although it would be ideal to develop a general measurement model to cover all the instructional models, we are not prepared in this paper to advance such a model.)

The first purpose of the investigation was to provide a description of the testing models that are currently being used in selected individualized instructional programs. Three programs were selected for study: Individually Prescribed Instruction, Program for Learning in Accordance with Needs, and Mastery Learning. [These models as well as others are also discussed by Baker (1971); however, he was concerned with their computer-based instructional management systems which is of only secondary interest in this paper.] These programs were selected in this study because they are among the best known, and because there is a substantial amount of information available on each. In the following sections, an introduction is provided for each instructional model. The introduction includes a brief history, the content areas covered, and an indication of the extent of implementation. Also, a description of each instructional paradigm and details on the testing model is provided. An attempt is made to pinpoint the decision points in each model, spelling out the consequences of the various possible actions in relation to each of the "possible true states of nature."

The discussion of the models is based on descriptions found in books, papers, and reports; on-sight visits; and meetings with many of the developers. It should be noted however that programs are often implemented by teachers quite differently than they are reported in the literature. Also, it should be remembered that these programs are constantly changing; hence, it is possible that certain features of the models are not exactly as they are described here. In particular, it is our impression that PLAN is being implemented in a way quite different from what has been written about it. This is because Westinghouse Learning Corporation has now taken over the development and implementation components.

A second purpose was to compare the three programs along the component parts of the testing model; namely, selection of a program of study, criterion-referenced testing on the unit objectives, assignment of instructional modes, and final year-end assessment.

A final purpose was to briefly outline several promising lines of research in connection with the testing methods and decision procedures for individualized instructional programs.

II. Individually Prescribed Instruction (IPI)

2.1 Background

The Learning Research and Development Center (LRDC) at the University of Pittsburgh initiated the Individually Prescribed Instruction Project during the early 1960's at the Oakleaf School in cooperation with the Baldwin-Whitehall Public School District near Pittsburgh. Major contributors to the project over the years include Robert Glaser, John Bolvin, C. M. Lindvall, and Richard Cox. Initial activities concentrated on producing instructional materials and training materials. More recently, research and evaluation activities have assumed an increasingly important role in Center activities.

As of 1972 the IPI program was being implemented in over 250 schools around the country. Distribution of materials and other information on the program is managed by Research for Better Schools, Inc., a United States Office of Education Regional Laboratory located in Philadelphia. At present, instructional materials are available in elementary mathematics, reading, science, handwriting, and spelling.

2.2 Description of the Instructional Paradigm

While we will discuss the instructional paradigm and the corresponding test model in the context of the IPI mathematics program, the procedures, techniques, etc., described, are in no way limited to that content area. In fact, it should be noted that the mathematics program as implemented is probably somewhat different from what we describe here, since the LRDC is constantly refining and improving the program (Lindvall, personal communication). Fortunately, for our purposes the basic structure of the program remains as described.

It is instructive first of all to describe the structure of the mathematics curriculum. Cooley and Glaser (1969) report that the mathematics curriculum consists of 430 specified instructional objectives. These objectives are grouped into 88 units. (In the 1972 version of the program there were 359 objectives organized into 71 units.) Each unit is an instructional entity which the student works through at any one time. There are 5 objectives per unit, on the average, the range being 1 to 14. A collection of units covering different subject areas in mathematics comprises a level; the levels may be thought of as roughly comparable to school grades. For illustrative purposes, Table 2.2.1 presents the number of objectives for each unit in the IPI mathematics curriculum.

The teacher is faced with the problem of locating for each student, that point in the curriculum where he can most profitably begin instruction. Also, he is responsible for the continuous diagnosis of pupil demonstrating proficiency in each skill prescribed in his particular instructional sequence as he moves along.

At the beginning of each school year the teacher places the student within the curriculum; that is, he identifies the units in each content area for which instruction is required. After completing the gross placement, a single unit is selected as the starting point for instruction, and a diagnostic instrument administered to assess the student's competencies on objectives within the unit. The outcome of the unit test is information appropriate for prescribing instruction on each objective in the unit. In addition it is also necessary to select the particular set of resources for the student. In theory, resources that match the individual's "learning style" are selected. Within each unit, there are short tests to monitor the student's progress. Finally, upon completion of initial in-

Table 2.2.1[1]

Number of Objectives for Each Unit in the
IPI Mathematics Curriculum

Content Area	Levels							
	A	B	C	D	E	F	G	H
Numeration	12	10	8	8	8	3	8	4
Place Value		3	5	10	7	5	2	1
Addition	3	10	5	8	6	2	3	2
Subtraction			4	6	3	1	3	1
Multiplication				8	11	10	6	3
Division				7	7	9	5	5
Combination of Processes			6	5	7	4	5	6
Fractions	3	2	4	6	6	14	5	2
Money		4	4	6	4	1		
Time		3	2	7	9	5	3	1
Systems of Measurement		4	3	5	7	3	2	
Geometry		2	2	3	9	10	7	9
Special Topics			1	3	3	5	4	5

[1]Reproduced, by permission, from Lindvall, Cox, and Bolvin (1970).

struction in each unit, assessment and diagnostic testing takes place. In the next section, we review the tests and the mechanisms for making these decisions. Suffice to say here that it has been found that teachers differ in the extent to which they follow prescri c..ion-making rules (Lindvall, Cox, and Bolvin, 1970).

2.3 Details of the Testing Model

Various reports over the last couple of years have dealt with the testing model and its development (Lindvall, Cox, and Bolvin, 1970; Glaser and Nitko, 1971; Cox and Boston, 1967). A flow chart of the testing model is presented in Figure 2.3.1. To monitor a student through the program the following tests are used: placement tests, unit pretests, unit post-tests, and curriculum-embedded tests. All of the tests are criterion-referenced with performance on the tests compared to performance standards for decision-making.

How sophisticated is the decision-making process utilizing the scores from the various tests? According to Glaser (1968):

> At the present stage of our knowledge, the decision rules for going from measures of student performance to instructional prescriptions may not be very complex, but little is known about the amount of complexity required, although, the individual monitoring of student performance provides us with a good data base to study this process.

Promising developments in the last couple of years include increased knowledge about constructing and evaluating criterion-referenced tests. Al... the research on branched testing strategies (Ferguson, 1969, 1971) has much potential for improving the efficiency of the testing model. This second point will be discussed in greater detail in a later section.

Placement Te..s

When a new student enters the program, it is necessary to place the student at the appropriate level of instruction in each of the content areas.

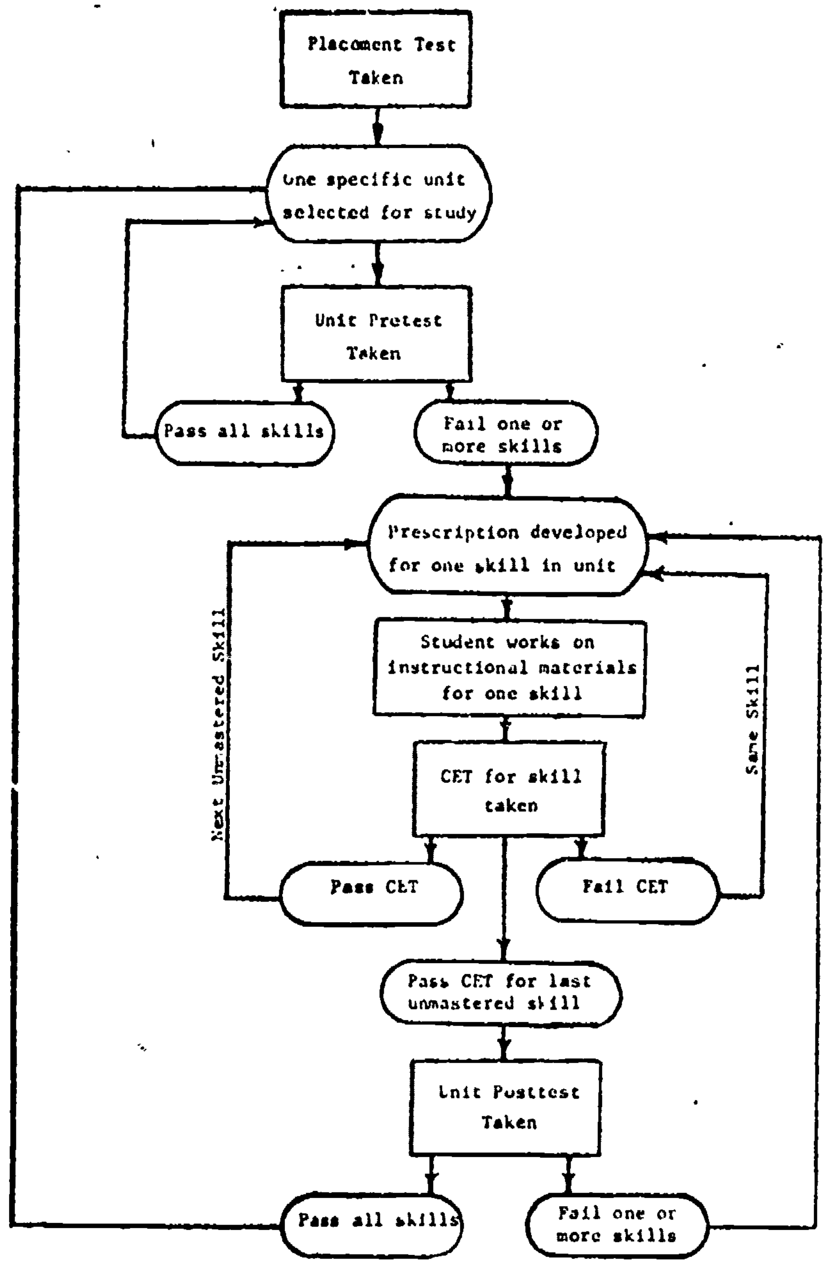

Figure 2.3.1. Flowchart of steps in monitoring student progress in the IPI program. (Reproduced, by permission, from Lindvall and Cox, 1969.)

[Glaser and Nitko (1971) called this stage-one placement testing.] Typically, this is done by administering a placement test which covers all of the subject areas at a particular level (see Table 2.2.1). Factors affecting the selection of a level for placement testing of a student include student age, past performance, and teacher judgment. Generally, the placement test covers the most difficult or most characteristic objectives within each area. Placement tests are administered until a unit profile identifying a student's competencies within each area is complete. At present, the somewhat arbitrary 80-85% proficiency level is used for most tests in the IPI system.

Scores for a student on items measuring objectives in each unit and area in the placement test are used to define an individual program for him. The standard procedure is to assign instruction on units in which placement test performance on items measuring a few representative objectives in the units is between 20% and 80%. If the score is less than 20% for a given unit, the unit test in the area at the next lowest level is administered and the same criterion is applied. If he passes the unit test, he receives instruction in the unit in the next level. In the case where a student has a score of 80% or over, he is tested on the unit in the area at the next highest level. [Further information is provided by Lindvall, Cox, and Bolvin (1970), Weisgerber (1971) and Cox and Boston (1967.]

For example, suppose a student were to achieve scores on level E of 60%, 90%, 60%, 60%, 30%, 30%, 25%, 90%, 50%, 10%, 0%, 30%, 30% in the thirteen areas indicated in Table 2.2.1. It is likely that he would be prescribed instruction at level E in the areas of numeration, addition, subtraction, multiplication, division, combination of processes, money, geometry, and special topics. He would receive the level F placement tests in place value and fractions. If, for example, he scores 60% and 10% respectively, he would receive instruction at level F in place value and probably level E in fractions. He would also be administered the level D placement tests in the areas of time and systems of measurement. If, for example, his scores were 0% and 40%, he would receive a still lower placement test in the area of time and would be prescribed instruction at level D in systems of measurement. If he scores 85% on the level C placement test in the area of time, he would be assigned to level D for instruction.

In order to acquire some information on the average length of the tests, the level E placement tests of the 1972 edition of the IPI program were selected and examined. Analysis revealed that on the average there are 12 items measuring the objectives in each area (with a range of from six to 20).

In summary, we note that the placement test has the following characteristics. provides a gross level of achievement for any student in the curriculum, and provides information for proper placement of students in the curriculum.

Unit Pretests and Posttests

Having received an initial prescription of units, a student proceeds by taking a pretest for a unit at the lowest level of mastery on his profile. [Glaser and Nitko (1971) call this stage-two placement testing.] A unit pretest includes one or more items to measure each objective in the unit. A review of the unit pretests and posttests in level E revealed that the approximate number of items on a test is 37 (the range is from 21 to 64) and the average number of items measuring each objective is six (the range is from four to seven). Lindvall and Cox (1969) report that the length of a pretest is determined by the number of objectives in the instructional unit and by the number of items used to test each objective. No fixed number of items to measure each objective is used because of the diverse nature of the objectives. For example, they note that, "an objective like--the pupil can solve simple addition problems involving all number combinations--will require more items than would an objective like--the pupil must select which of three triangles is equilateral--."

A student is prescribed instruction in each objective in the unit for which he fails to achieve an 85% mastery level.[1] In the case where the student demonstrates mastery of each objective, he is moved on to the next unit in his profile, where he again takes a pretest.

The unit posttests are simply alternate forms of the unit pretests and are administered to students as they complete instruction on the unit. A student receives a mastery score for each objective in the unit. He is required to repeat instruction on any objective where he fails to achieve an 85% mastery score. He is directed to the next unit in his profile if he demonstrates mastery on each objective covered in the unit posttest. Those who repeat instruction on one or m re of the objectives must take the unit posttest again before moving on in their program.

In summary, pretests and posttests are available for each unit of instruction. The proper pretest is administered on the basis of student's curriculum profile, and learning tasks for each skill are assigned (or not assigned) on the basis of a student's performance on items measuring the skill.

Compared with students in many other types of mathematics programs, it is clear that the student in the IPI program spends more of his time taking tests. However, to some extent this can be justified on the grounds that testing is an integral part of the learning process in the IPI model. Nevertheless, there seems to be good reason for researching techniques to reduce testing time.

---

[1]A mastery score on each objective for a student is calculated as the percentage of items on the test that measure the objective that the student answers correctly.

Hsu and Carlson (1972) point out several problems associated with the current version of the unit pretests and posttests. The existing system requires that every objective be tested; hence, the time a student spends taking tests is considerable. Also, because of management and scoring problems, feedback to the student on his results is not immediate. Further, students are occasionally required to take the same posttest on a second occasion. This raises a question about practice effect.

One very promising way to reduce the testing time with the correlated result of producing better instructional decisions is suggested in the branched testing work of Ferguson (1969, 1971). Ferguson showed that by using a tailored testing strategy, a computer terminal to monitor the selection of test items, and information on the hierarchical structure of the items, he was able to significantly reduce unit testing time without any loss in decision-making accuracy. A comprehensive review of the work in branched testing is out of place here; suffice to say here that major contributions to the area include Ferguson (1969, 1971), and Lord (1970). A review of some of the work in the area is provided by Bock and Wood (1971).

Curriculum-Embedded Tests

As the student proceeds through a unit of instruction, his progress must be monitored. This is done by curriculum-embedded tests (CET). As used in the mathematics IPI program, a CET is primarily a measure of performance on one specific objective. There are usually several test items to measure the objective. A review of the CETs in level E of the program revealed that there are on the average about three items measuring the primary objective covered in the CET. The range is from two to five. If a student receives a score of 85%, he is permitted to move on to the next prescribed objective. Otherwise, he is sent back for additional work and then he takes an alternate form of the CET when he is ready.

A secondary purpose of the CET is to pretest, in a rough way, the next objective in the learning sequence. (Objectives in a unit are arranged into a learning sequence.) Students may pretest out of the next skill in the sequence by achieving 85% or higher on the short test which makes up the second part of the CET and on part one of the CET for that skill. It would appear from a review of level E tests that there are about two items measuring the secondary objective. In cases where a student does not need instruction on the next skill, he can skip part two of the CET and move on to the part two of the CET that tests the next skill he needs for his program. This additional pretesting of an objective in the CET gives students a chance to demonstrate mastery of new skills not specifically covered in the instruction to that point and to eliminate that instruction from his program.

Student Diagnosis

Once the student has been assigned to a unit of instruction and the objectives for which he needs instruction have been identified by the unit pretest, there still remains the problem of deciding which of several instructional methods is "optimal" for him. That is, of the available instructional methods for a particular instructional unit, in which of them would a student with a known background in the program and specific goals, interests, and aptitudes stand the "best" chance of learning the material? Glaser and Nitko (1971) call this a diagnostic decision.

III. Program for Learning in Accordance with Needs (PLAN)

3.1 Background

Project PLAN is a major ungraded, computer supported individualized instruction program in education developed by the American Institutes for Research over the last seven years. (For background, see Weisgerber, 1971.) The project was initiated by John Flanagan to handle many of the shortcomings of our educational system as revealed by Project TALENT (Flanagan, et al., 1964).

The PLAN program is currently being used in over 70 schools with more than 35,000 students in grades one through twelve. Instructional materials are available in four areas: social studies, language arts, mathematics, and science. Westinghouse Learning Corporation is now responsible for the monitoring and marketing of Project PLAN materials. They also operate the computer installation necessary for the proper functioning of Project PLAN in a school.

Unfortunately, the implementation of the model in 1972-73 involves far fewer features than was originally described by the proponents of the program a few years ago. Nevertheless, we will describe the more elaborate version of the program in this paper.

3.2 Instructional Paradigm

The basic unit of instruction in PLAN, called a module, is an instructional package made up of about five behavioral objectives. It normally takes a student about two weeks to complete a module of instruction. Also, there are many objectives classified at the higher levels of Bloom's (1956) taxonomy that do not fit nicely into the regular modules. These are named module-set objectives, and examples include concept development and problem-solving skills. They are worked into the regular modules and progress is measured by PLAN achievement tests administered periodically throughout the program. According to Rhetts (1970) there are more than 1100 modules in PLAN. For each module, there are several different teacher-learning units (TLU) assigned individually on the basis of aptitudes, interests, learning style, etc. All modules in the secondary school curricula are coded as to whether, 1) they are part of a state or local requirement, 2) essential for a given educational or occupational area, 3) highly desirable for that area, 4) essential for minimum functioning as a citizen, 5) highly desirable for all citizens to know, or 6) would make the student a particularly well informed citizen.

TLU's are coded according to: 1) reading difficulty, 2) degree to which it requires teacher supervision, 3) its media richness, 4) degree to which it requires social involvement and/or group learning activities, 5) the amount of reading involved, and 6) variety of activities in the module. There are, on the average, two TLU's for each module. Along the lines of Dunn (1970), we will describe the most complex version of the program--the version currently being used in the secondary school.

At the beginning of each year, a program of study is prepared for each student. This includes a list of modules, suggested TLU's, and a recommended sequence in the four content areas. To really provide individualized instruction, it is necessary to know about student needs, goals, abilities, and interests <u>and</u> to use the information in developing a program of study (POS) for him. As part of the PLAN system then, the following information is collected:

1. parent and student educational goals
2. parent and student vocational aspirations
3. student level of achievement and vocational interests
4. student abilities (such as reading comprehension and arithmetic reasoning)
5. past performance of student in program
6. student's learning style.

A variety of questionnaires and testing instruments have been developed to collect the above information.

Abilities are measured each year with the Developed Abilities Performance Test (DAPT). This test consists of 18 scales (see, for example, Jung, 1970) such as those to measure arithmetic reasoning, reading comprehension, abstract reasoning, mechanical comprehension, and ingenuity.

On the basis of the above information, a program is developed and the student is monitored through it by continuous module posttesting and PLAN achievement testing. Let us look now at the testing phase of the program in more detail.

3.3 Testing Model Details

Within a PLAN school, there exists a multitude of decisions to make on each student. These include development of a program of study, periodic assessment of module-set objectives, performance on the modules of instruction, assignment of TLU's, and monitoring yearly, important skills. The major decision points are shown in Figure 3.3.1. Unfortunately, there is little available information on how these decisions are made.

Development of a Program of Study

On the basis of DAPT scores which are matched to Talent data of people

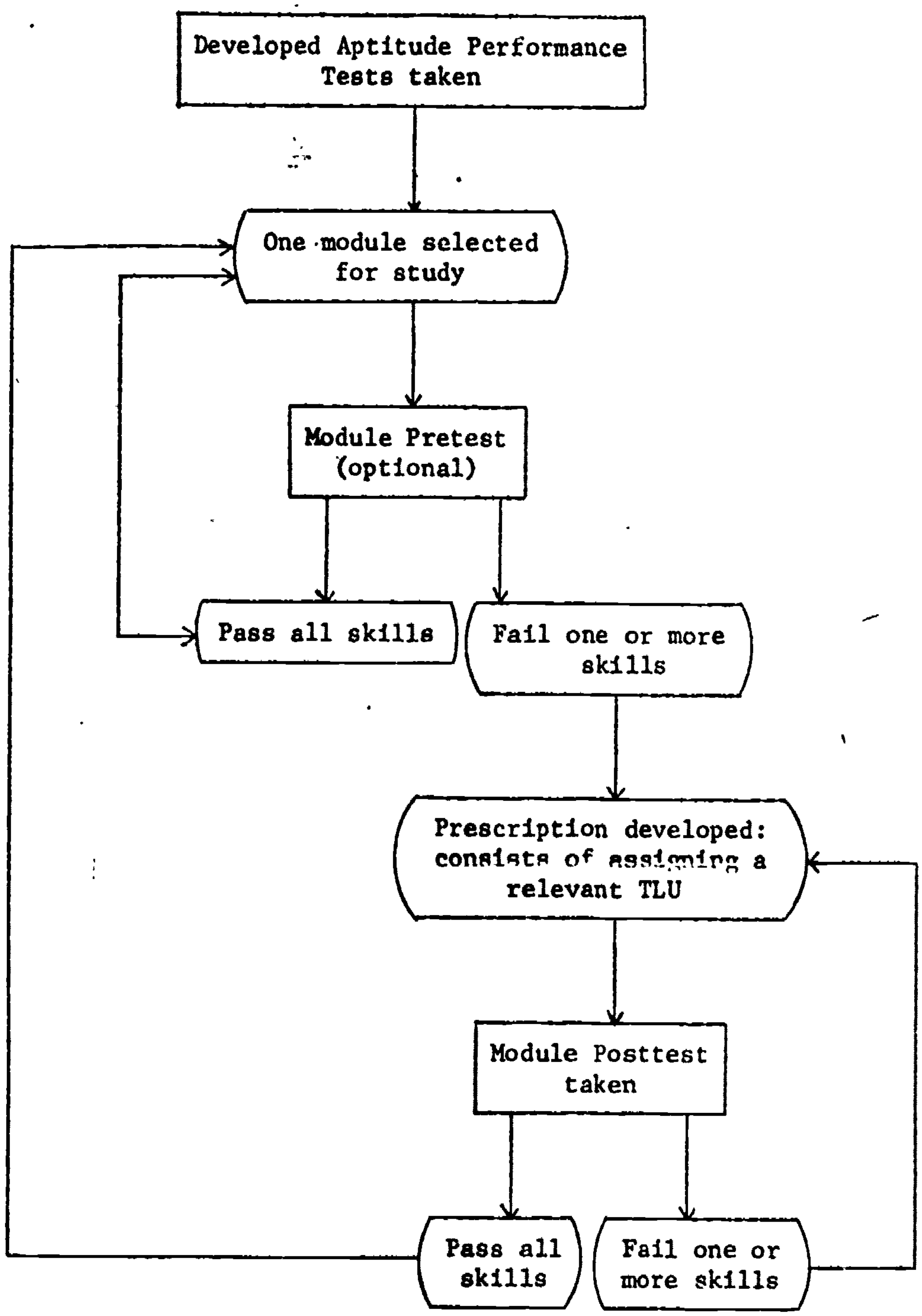

Figure 3.3.1 Flowchart of steps in monitoring student progress in Project PLAN.

in different occupations, the students and parents select a long range goal [(LRG) (one of 12 families of occupations)]. Information on the long range goal along with parent and student information described in the last section is used to develop a program of study. The DAPT is also used in the determination of the number of modules a student will study in a year. Jung (1970) reports that on the basis of weights derived from regression analyses, a quota is identified for each PLAN student in each subject area. Modules are then assigned to him on the basis of his LRG group membership until this quota is filled.

Developed Aptitude Performance Tests

These tests are given at the beginning of each school year. Information on the length, kinds of test items, reliability and validity does not appear to have been published. Also, we do not know whether a different version of the test is used in each year, or whether the same version is used for several years. Regardless, unless comparability of the score scales for the different versions has been carefully done, we doubt whether the change scores (for individuals or groups) on each variable from year to year have very much meaning.

PLAN Achievement Tests

Mastery of the module-set objectives is measured at specific points in the curriculum using PLAN achievement tests. However, we are also unclear on the make-up of the PLAN achievement tests. Apparently, they are measured at "specified points" in the curriculum and the format of these tests is sometimes something other than the paper and pencil variety.

Module Tests

When the student feels he has mastered the materials covered in a module, he can take a criterion-referenced module posttest which has on it several items measuring each objective in the module. The items are presented usually in a selection format to facilitate computer scoring. On the basis of his performance, the computer using built-in decision rules makes one of four decisions. If he answers all items correctly, he is given a "complete" on the module and the computer print out tells him where to go next. If he makes a "few" errors, he is given a result of "Student Review". The computer specifies his performance on each objective and indicates the ones he should review before beginning his next module.

Students who miss a large number of items on the test but still score high enough to pass, receive a result of "Teacher Certify". He is instructed by the teacher on which objectives to review and/or restudy. He is not given his next module until, in the judgment of the teacher, he has mastered all of the objectives. An alternative is to have the student repeat the module posttest. The fourth possibility is student failure to pass the test. In this situation, he is instructed to restudy the module with the same TLU or another. In the case where he misses the test again, the teacher intervenes and takes some appropriate action to clear up the problem.

Assignment to Instructional Modes

The basic problem was described in a discussion of the IPI program, i.e., what particular instructional mode (or in this case, TLU), should the student take to study the module so as to maximize his changes of learning the material. Dunn (1970) notes, "that the computer, from a complex set of decision rules, matches the student with specific TLU's". We wonder what those rules would be, particularly since there is no theory of instruction to guide in developing optional assignment rules. To this point in time educational psycholgists have only been able to find a handful of interactions between background variables and instructional method. A partial answer is provided by Weisgerber and Rahmlow (1971). They noted that teacher learning units are based upon different assumed learning styles of students and are guided by a philosophy of education (Flanagan, 1970) and a theory of learning (Gagné, 1965).

IV. Mastery Learning

4.1 Background

The mastery learning concept was introduced to American Schools in the the 1920's with the work of Washburne (1925). However, because technology was not developed to the point that the program could operate efficiently, interest in the concept steadily diminished until it was revived in the form of programmed instruction in the late 1950's. (Programmed instruction was an attempt to provide students with instructional materials that would allow them to move at their own pace and receive constant feedback on their level of mastery.) The work by Carroll (1963, 1970) and Bloom (1968) and Bloom's students (Block, 1971; Airasian, 1971 and others) was instrumental in bringing mastery learning to the attention of instructional designers and researchers.

Since Bloom's paper in 1968, a great deal of research has been conducted; and the results suggest that the mastery learning model "can be easily and inexpensively implemented at all levels of education and in subjects ranging from arithmetic to philosophy to physics (Block, 1970). The model has been used now with more than 20,000 students.

4.2 Instructional Paradigm

This model is quite different from IPI and PLAN in that it attempts to individualize instruction with a group-based instructional environment. The curriculum is organized into units of instruction made up of a collection of behavioral objectives, and for each unit one or more criterion-referenced tests is used to measure mastery. Individualization is handled via supplemental materials, feedback, and corrective techniques applied to students who do poorly on the posttests.

Mayo (1970) in describing the mastery learning model notes that:

1. Students are made aware of course and unit expectations, so that they view learning as a cooperative rather than as a competitive venture.
2. Standards of mastery are set in advance for the students, and grading is in terms of absolute performance rather than relative performance.
3. Short diagnostic tests are used at the end of each instructional unit.
4. Additional learning is prescribed for those who do not demonstrate unit mastery.
5. Additional time for learning is prescribed to students who seem to need it.

The mastery learning model is less impressive in scope than PLAN, and the requirements for an effective testing plan are less stringent than with IPI or PLAN. Features of mastery learning appear to be that it is easily implementable, does not require the use of a computer, and is appropriate for almost any content area. Also if mastery learning is carried out properly, previous research suggests that students will achieve higher scores and have more interest in school and a better attitude toward school. Unlike the other two models, with mastery learning much of the work has been on research related to the correctness of the model of school learning. An extensive number of content areas have been studied.

It should be noted that there are many variations on the basic mastery model as originally proposed by Bloom (1968). Some of them are summarized by Block (1971), and an example would be the work of Kim (1971).

4.3 Test Model Details

Block (1971) notes that, "To individualize instruction within the context of ordinary group-based instruction, mastery learning relies heavily on the constant flow of feedback information to teacher and learner." It does not seem however that there is as much testing in mastery learning as in IPI or PLAN. A flow chart of the testing component is shown in Fugure 4.3.1.

The mastery learning testing model as described by Airasian (1971) represents a special case of the IPI testing program. There is no placement testing, and unit pretesting and curriculum-embedded testing are not emphasized. Unit posttesting and final assessment represent the two major kinds of testing in the program. In the spirit of Scriven (1967), these two areas are known as formative and summative tests. It should be noted, however, that formative tests or unit posttests, as they are called in IPI, are not used for grading. They are used for diagnosing learning difficulties only.

Formative Tests

A formative test is designed to cover the objectives over a short unit of instruction in the mastery learning program. It is used to determine whether or not a student has mastered the material and to serve as a basis for prescribing supplemental work in areas where the student is weak (Airasian, 1971). Implementers of the mastery learning model have set the passing standard anywhere from 75% to 100%. There is no set number of items or format suggested to measure each objective; however, there is a suggestion that instructional decisions are made on the basis of responses to individual items.

The formative tests in mastery learning represent the key to individualizing instruction since it is on the basis of these scores that

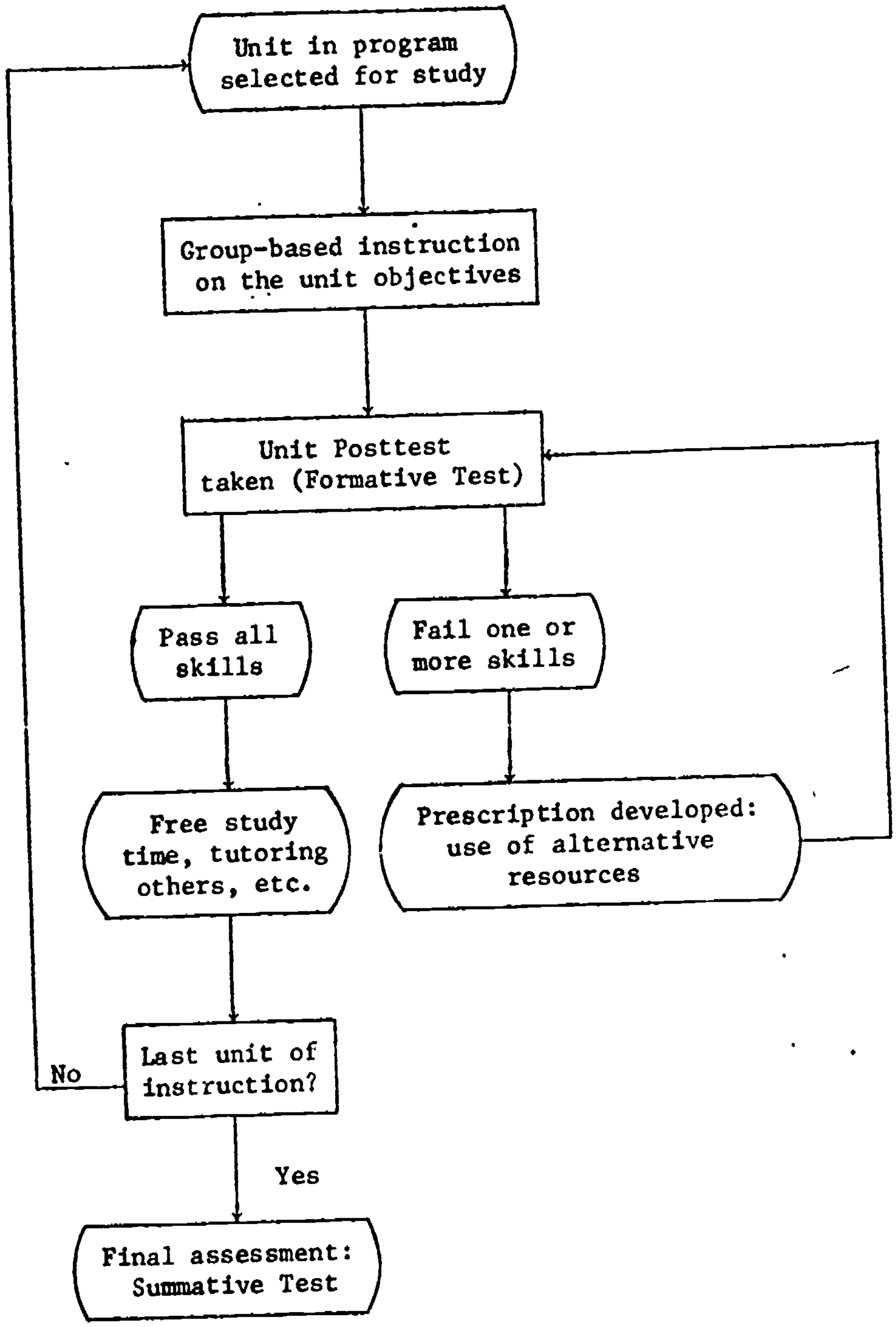

Figure 4.3.1 Flowchart of steps in monitoring student progress in a typical version of a mastery learning model.

individualization of instruction can take place. Units are kept small so that unit testing takes place frequently to increase the effectiveness of the individualization of instruction component of the program.

Summative Tests

The primary purpose of the summative test in the mastery learning model is to grade students on the basis of their achievement of course objectives. The items in the test are keyed to objectives and representative of the pool of course objectives. A criterion-referenced interpretation of the scores is recommended. It is proposed that cutting points be located on the ability continuum and grades should be assigned on the basis of a student's position on the continuum and not relative to other students in the course. A norm-referenced interpretation of the scores is also possible.

Final Comments

Mastery learning is probably the least different from traditional instruction since the principal instruction is always grouped based and final grades are assigned. (However, it is expected that because of various features built into the program that the final assessment testing will not be as threatening a situation for the student as it is in more traditional programs.) Differences with traditional instructional models include features such as individual pacing, and the big difference is the use of frequency tests on small units of instruction to diagnose learning problems. Important features are the feedback/correcting-review techniques. It would appear, however, that there is little in the way of sophistication concerning the testing model. For example, there appears to be no guidelines for determining the optimum number of items to measure each objective on a unit posttest. An exception is the excellent work of

Block (1970) in investigating, among other things, the problem of setting cutting scores on criterion-referenced tests to separate students into two groups--masters and non-masters. His results suggest that setting cutting scores high (95%) may be best for cognitive learning but in the long run positive attitudes and interest in the subject are less likely to develop. With a reduction in the cutting score to 85% there was a reduction in cognitive learning, but selected affective outcomes were maximized.

V. A Comparison of the Testing Models

5.1 Introduction

In the three previous sections we have highlighted the basic testing and decision-making features in three individualized instructional programs--IPI, PLAN, and Mastery Learning. IPI, PLAN, and ML are really generic terms to represent three classes of individualized instructional programs. [Incidentally, these represent only a small portion of the possible classes described by Gibbons (1970), although the classes we selected for study are among the most common and ones that require, generally, more testing.] Within any particular class there is still considerable variation among the various programs caused by local needs, teacher preferences, and methods of implementation. We shall discuss general features since they remain the same from program to program.

Within all three models, instruction is self-paced although mastery learning is somewhat more structured since the initial instruction on a unit is group-paced. With each of the models, the content is organized into units or modules. Generally, in IPI and ML the student is expected to demonstrate mastery on all the units before completing the program of study although by his performance on unit pretests, it is possible for him to avoid instruction on any of the units. (One variation that does come up is the availability of "enrichment materials" which are an optional part of the curriculum.) In PLAN, at any grade level there are far more units than any student could or would ever want to master. Thus, it is first of all necessary to define a content domain of study for each student.

In the remainder of the section, we shall limit discussion to testing and decision-making issues. In order to develop a framework for the discussion, we have chosen to focus on the following issues:

1) selection of a program of study;
2) criterion-referenced testing on the unit objectives;
3) assignment of instructional modes;
4) final year-end assessment.

These represent the extent of the decision paradigms within the three models. The importance and sophistication used in handling each component varies from one model to another.

5.2 A Compendium of Decision Paradigms

Selection of a Program of Study

A program of study is that collection of units which a curriculum designer deems necessary for the appropriate education of the student.

All three models are designed for utilization with a curriculum defined in terms of behavioral objectives arranged into blocks, units, or modules around a common topic or theme. Generally in IPI and ML, students are expected to demonstrate mastery in all of the available program objectives. The starting assumption is that there exists a body of knowledge that the student needs to be able to demonstrate mastery in. This defines the program of study for the student. However, on the basis of high pretest results students can avoid instruction of selected units of instruction.

In PLAN, each student receives a unique program of study. The more advanced the students the more varied their programs of study become.

For reasons described above, selecting a program of study for a student in IPI or Mastery Learning is relatively easy. The decisions to be made reduce, basically, to determining whether students have mastered particular objectives. They will receive instruction only on objectives they have not mastered. In IPI, placement tests are used to determine the level of instruction in each area for the students. Here the problem of giving the student credit for units he has not mastered (a false positive error) seems to be somewhat more serious than mistakenly assigning him to instruction he does not need (a false-negative error). This follows since a student has a second chance to demonstrate mastery of the objectives in a unit through the unit pretest if he is mistakenly assigned instruction on it. To be made exempt from instruction on a unit he has not mastered, particularly if it is an important unit, will plague him in his future studies.

In theory at least in the PLAN program, developing a program of study is a complex affair. Done once a year it requires a wealth of information described in section 3.3 to develop the program. The danger of locating a student in the wrong program because of misjudgment on the part of the parents, teachers, or the student or because of a "less than 100% prediction system" are great; however, this is the same risk we take with selection of program in a traditional school. This is particularly serious in the high school where there is more choice than in the elementary school programs. However, the flexibility of the PLAN program makes switching from one program to another easier.

Criterion-Referenced Testing on the Unit Objectives

There are three kinds of testing appropriate here: unit pretesting, unit posttesting, and curriculum-embedded testing. All three kinds of testing are used in IPI and PLAN although unit pretesting is not stressed in PLAN. The possibility existed for all three kinds of testing in Mastery Learning; however, unit pretesting is not emphasized and a student can avoid the curriculum-embedded testing by passing the unit posttest and thus avoid the remedial instructional materials. It is possible that curriculum-embedded tests are not available in the remedial materials either.

Let us briefly look now at the losses involved in making different kinds of decisions. It should be recalled that the unit tests (or module tests) measure performance on each objective or skill with several items. On the unit pretests, a student receiving credit for non-mastered objectives will likely be "caught" on the administration of the posttest and correct instruction can be assigned. However, to the extent that these objectives are a prerequisite to others in the unit there is a potential problem. (Perhaps, this is a place where Bayesian statistics might be helpful in producing an improved profile of scores across objectives measured by the unit pretest. This would undoubtedly improve the overall decision-making accuracy. Likewise this strategy could be used on the unit posttests.)

To assign instruction on the basis of pretest score results to objectives on what a student has already mastered will prove to be frustrating to him; however, it should be noted that the majority of errors of this type occur because students are close to the cutting score.

Receiving credit for non-mastered objectives on the posttest to the extent that the objectives are prerequisites to others in future units will interfere with the rate of learning at that point. This error seems to be less serious in terms of program efficiency if the objectives are terminal. Failing to receive credit for mastered objectives would seem to be less serious since the student could move through the remedial materials quickly and retake the test.

Since any decisions on the basis of curriculum-embedded test score results affect the student for only a limited amount of time and there exist checks on any decisions with the unit (or module) posttest, there is little concern for developing more appropriate testing decision guidelines at this level.

Assignment of Instructional Modes

An integral component of nearly every individualized instruction program is the feature whereby there exist several alternate instructional modes for the various units on instruction that can be assigned in some optimal way to students. In theory anyway, with IPI and PLAN, past performance and background aptitude variables are used to assist the students in selecting the "best" mode of instruction. With Mastery Learning, this feature can be operationalized following the group-based instruction and the unit posttests. It is at this point that decisions on the proper corrective feedback techniques to use need to be made.

Investigators of the possible interactions between instructional methods and aptitudes are conducting what has been termed aptitude-treatment interaction research (Cronbach, 1967). Disappointing is the fact that while nearly all developers of individualized programs include this feature of utilizing ATI results in assigning instruction, there are few real demonstrations of significant interactions between aptitudes and instructional modes (Bracht, 1970; Cronbach and Snow, 1969). Authors such as Glaser (1972) have attempted to explain these results and suggest some new directions. However, it would appear that we are far from a "theory of instruction" to guide the instructional decision-making in assignment of "optimal" instructional modes to students.

The benefits (assuming equal treatment costs) of the ATI classification scheme for improving the quality of instruction depend directly on the differences among the slopes of the regression lines for predicting criterion scores with different aptitude variables in the different instructional modes. The bigger the difference in slopes the greater is the potential benefit to the student for assigning one instructional mode or another. However in looking at the overall benefits and losses of such a system it would seem that the appropriate baseline for comparative purposes would need to be data derived from a traditional instructional program.

Final Year-End Assessment

The particular feature seems to be handled in much the same way in IPI and PLAN. Information is reported on the number and nature of units that a student has mastered. Little or no information is provided by the school to students and parents that could be used for norm-referenced assessment. In the mastery learning model, a score is reported to measure achievement on the year-long activities. Both norm-referenced and criterion-referenced interpretations are possible.

VI. Some Directions for Further Research

6.1 Concluding Remarks

A review of IPI, PLAN, and mastery learning programs as well as many other objective-based curriculum programs not reported in this paper reveals that there are many important questions remaining to be answered in regards to individual assessment models. In this concluding section a few of the more important problem areas are discussed.

In order to develop an instructional model that is sensitive to individual needs, abilities, interests, and goals in a way that will allow the student to maximize his learning, we need a theory of instruction. A theory of instruction should set down rules on the most efficient way of achieving knowledge (Bruner, 1964). This theory would provide guidelines on how to prescribe instruction to increase learning. One paper that addresses the problem is Groen and Atkinson (1966). Current reports on the related topic of aptitude-treatment interactions are by Cronbach and Gleser (1965), Cronbach and Snow, (1969), Bracht (1970), and Glaser (1972).

In making decisions on the basis of criterion-referenced test scores one assumes a good match between items and the behavioral objectives they are intended to measure. To the extent that test items do not accurately measure the objectives, any decisions based on test performance will be inaccurate. To date a satisfactory methodology for item validation does not exist although several useful papers provide partial solutions (Dahl, 1971; Rovinelli and Hambleton, 1973).

A theory of criterion-referenced tests and measurements is also needed to guide the users of the tests in the context of programs described here. This theory should probably be based on a threshold loss function rather than a squared-error loss function as has been done in classical test theory (Lord and Novick, 1968; Hambleton and Novick, 1973). This theory would include reliability, validity, test scoring and item analysis procedures for criterion-referenced tests. It would also provide guidelines and techniques for setting cutting scores and allocating testing time.

Another problem which has to be reckoned with for criterion-referenced tests is an instance of the bandwidth-fidelity issue (Cronbach and Gleser, 1965). When the total testing time is fixed and there is interest in measuring many competencies, one may be faced with the problem of whether to obtain very precise information about a small number of skills or less precise information about many more skills. Time allocation algorithms (analytical procedures for deciding how many items on a test should measure each objective) of a rather different kind than those presented by Woodbury and Novick (1968), and Jackson and Novick (1970) will be required. The problem of how to determine the number of items to measure each skill so as to maximize the percentage of correct decisions or some similar measure of overall decision-making accuracy on the basis of test results has yet to be resolved.

Estimation of mastery is a problem that is encountered frequently in in the objective-based program. Bayesian methods have been suggested (Hambleton and Novick, 1973) but there has been no empirical demonstrations of their usefulness in this context nor are guidelines for the use of Bayesian methods available at the present time. Prior information for a Bayesian solution might be mastery scores on other skills covered on the test or on performance on skills measured previously. (In the case of posttesting, pretest information could be used as the prior.) Also, just as data from other examinees can improve the precision of estimation of achievement in a norm-referenced testing situation for an individual (Lord and Novick, 1968), so perhaps the same can be done with criterion-referenced measurement problems.

Within many objective-based programs the strategy of branched testing would seem to be an appropriate technique, at least in situations where the objectives in a content area can be arranged into hierarchical sequences. Some of the practical problems have been resolved in the Pittsburgh IPI Program so that the technique can now be used on a limited basis. Nevertheless, many problems remain before adoption should or can proceed with other programs. For example, it would be necessary to develop a non-automated modified version of branched testing for schools without computers. Also, we need to know much more about setting starting places, step sizes, stopping rules, etc., before we can effectively use branched testing in an instructional setting.

References

Airasian, P. W. The role of evaluation in mastery learning. In J. H. Block, (Ed.), Mastery Learning: Theory and Practice. New York: Holt, Rinehart, and Winston, 1971.

Atkinson, R. C. Computer-based instruction in initial reading. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Atkinson, R. C., & Wilson, H. A. (Eds.) Computer-assisted Instruction: A Book of Readings. New York: Academic Press, 1969.

Baker, F. B. Computer-based instructional management systems: A first look. Review of Educational Research, 1971, 41, 51-70.

Block, J. H. The effects of various levels of performance on selected cognitive, affective, and time variables. Unpublished doctoral dissertation, University of Chicago, 1970.

Block, J. H. (Ed.) Mastery Learning: Theory and Practice. New York: Holt, Rinehart, and Winston, 1971.

Block, J. H. Student evaluation: Toward the setting of mastery performance standards. Paper presented at the annual meeting of AERA, Chicago, 1972.

Bloom, B. S. (Ed.) Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: D. McKay, 1956.

Bloom, B. S. Learning for mastery. Evaluation Comment, 1968, 1(2).

Bloom, B. S. Mastery learning. In J. H. Block (Ed.), Mastery Learning: Theory and Practice. New York: Holt, Rinehart, and Winston, 1971.

Bock, R. D., & Wood, R. Test theory. Annual Review of Psychology, 1972, 20, 193-224.

Bracht, G. H. Experimental factors related to aptitude-treatment interaction. Review of Educational Research, 1970, 40, 627-646.

Bruner, J. S. Some theorems on instruction stated with reference to mathematics. In E. R. Hilgard (Ed.), Theories of Learning and Theories of Instruction. (63rd National Society for the Study of Education Yearbook, Part I), Chicago: University of Chicago Press, 1964, 306-335.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Carroll, J. B. Problems of measurement related to the concept of learning for mastery. Educational Horizons, 1970, 48, 71-80.

Cooley, W. W., & Glaser, R. The computer and individualized instruction. Science, 1969, 166, 574-582.

Cox, R. C., & Boston, M. E. Diagnosis of Pupil Achievement in the Individually Prescribed Instruction Project. (Working Paper 15.) Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1967.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. M. Gagné (Ed.), Learning and Individual Differences. Columbus, Ohio: Charles E. Merrill, 1967.

Cronbach, L. J., & Gleser, G. C. Psychological Tests and Personnel Decisions. (2nd ed.) Urbana, Ill.: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. Individual Differences in Learning Ability as a Function of Instructional Variables. (USOE Final Report, Contract No. OEC 4-6-061269-1217.) Stanford, California: Stanford University, School of Education, 1969.

Dahl, T. Toward an evaluative methodology for criterion-referenced measures: Objective-item congruence. Paper presented at the annual meeting of the California Educational Research Association, San Diego, 1971.

De Vault, M. V., Kriewall, T. E., Buchanan, A. E., & Quilling, M. R. Teacher's manual: Computer Management for Individualized Instruction in Mathematics and Reading. Madison, Wisconsin: University of Wisconsin, Research and Development Center for Cognitive Learning, 1969.

Dunn, J. A. The PLAN approach to curriculum definition. Education, 1970, 90, 221-226.

Ebel, R. L. Criterion-referenced measurements: Limitations. School Review, 1971, 69, 282-288.

Ferguson, R. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. Unpublished doctoral dissertation, University of Pittsburgh, 1969.

Ferguson, R. L. Computer Assistance for Individualizing Measurement. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1971.

Flanagan, J. C. Functional education for the seventies. Phi Delta Kappan, 1967, 49, 27-32.

Flanagan, J. C. Program for learning in accordance with needs. Psychology in the Schools, 1969, 6, 133-136.

Flanagan, J. C. Individualizing education. Education, 1970, 90, 191-206.

Flanagan, J. C., Dailey, J. T., Davis, F. B., Goldberg, I., Heyman, C. A., Jr., Orr, D. B., & Shaycraft, M. F. The American High School Student. (USOE Final Report, Cooperative Research Project No. 635.) Pittsburgh: University of Pittsburgh, 1964.

Gagné, R. M. The Conditions of Learning. New York: Holt, Rinehart, and Winston, 1965.

Gibbons, M. What is individualized instruction? Interchange, 1970, 1, 28-52.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Glaser, R. Adapting the elementary school curriculum to individual performance. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Glaser, R. Evaluation of instruction and changing educational models. In Wittrock, M. C. and Wiley, D. E. (Eds.), The Evaluation of Instruction. New York: Holt, Rinehart, and Winston, 1970.

Glaser, R. Individuals and learning: The new aptitudes. Educational Researcher, 1972, 6, 5-13.

Glaser, R., & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), Educational Measurement. Washington: American Council on Education, 1971, 625-670.

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. Psychological Bulletin, 1966, 66, 209-320.

Hambleton, R. K., & Gorth, W. P. Criterion-referenced testing: Issues and applications. Center for Educational Research Technical Report No. 13. Amherst, Mass.: University of Massachusetts, Amherst, 1971.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, in press.

Heathers, G. Overview of innovations in organization for learning. Interchange, 1972, 3, 47-68.

Hieronymous, A. N. Today's testing: What do we know how to do? In Proceedings of the 1971 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1972.

Hsu, T. C., & Carlson, M. Oakleaf School Project: Computer-assisted Achievement Testing. (A Research Proposal.) Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1972.

Jackson, P. H., & Novick, M. R. Maximizing the validity of a unit-weight composite as a function of relative component lengths with a fixed total testing time. Psychometrika, 1970, 35, 333-347.

Jung, S. M. Guidance testing and the individualized program of studies. Education, 1970, 90, 224-231.

Kim, H. The mastery learning project in the middle schools. In J. H. Block (Ed.), Mastery Learning: Theory and Practice. New York: Holt, Rinehart, and Winston, 1971. (Abstract)

Lindvall, C. M., & Cox, R. The role of evaluation in programs for individualized instruction. In National Society for the Study of Education, Educational Evaluation: New Roles, New Means, Sixth-Eighth Yearbook, Part II, 1969, 156-188.

Lindvall, C. M., Cox, R. C., & Bolvin, J. O. Evaluation as a tool in curriculum development: The IPI evaluation program. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNally, 1970.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Lord, T. M. Some test theory for tailored testing. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing and guidance. New York: Harper and Row, 1970.

Lord, F. M., & Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Mayo, S. T. Mastery learning and mastery testing. Measurement in Education, 1970, 1(3).

Popham, W. P. Curriculum materials. Review of Educational Research, 1969, 39, 319-338.

Popham, W. J., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Rhetts, J. E. The impact of student learning style on curriculum assignment and performance in the PLAN program of individualized instruction. Education, 1970, 90, 248-251.

Rovinelli, R., & Hambleton, R. K. Some Procedures for the Validation of Criterion-Referenced Test Items. (Final Report) Albany, N.Y.: Bureau of S[illegible] and Cultural Research, New York State Education Department, 1973.

Scriven, M. The methodology of evaluation. In R. Stake (Ed.), Perspectives of Curriculum Evaluation. Chicago: Rand McNally, 1967.

Smith, H. A. Curriculum development and instructional materials. Review of Educational Research, 1969, 39, 397-414.

Suppes, P. The uses of computers in education. Scientific American, 1966, 215, 206-221.

Washburne, C.W. Educational measurements as a key to individualizing instruction and promotions. Journal of Educational Research, 1922, 5, 195-206.

Weisgerber, R.A. (Ed.) Developmental Efforts in Individualizing Learning. New York: Peacock, 1971.

Weisgerber, R.A., & Rahmlow, H.F. Strategies for learning in PLAN. In R.A. Weisgerber, (Ed.), Developmental Efforts in Individualized Learning. Itasca, Ill.: Peacock, 1971.

Wilhelms, F.T. The curriculum and individual differences. In National Society for the Study of Education, Individualizing Instruction, Sixty-First Yearbook, Part I, 1962, 72-82.

Wittrock, M.C., & Wiley, D.E. (Eds.) The Evaluation of Instruction. New York: Holt, Rinehart, and Winston, 1970.

Woodbury, M.A., & Novick, M.R. Maximizing the validity of a test battery as a function of relative test lengths for a fixed total testing time. Journal of Mathematical Psychology, 1968, 5, 242-259.