DOCUMENT RESUME

ED 047 014 TM 000 389

AUTHOR Pinsky, Paul D.
TITLE Feedback Information in the Comprehensive Monitoring of Educational Achievement.
INSTITUTION Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
PUB DATE Oct 70
NOTE 20p.; Paper presented at the 11th American Meeting of the Institute of Management Sciences, Los Angeles, California, October 1970

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Achievement, Behavioral Objectives, *Computer Oriented Programs, Curriculum Evaluation, Decision Making, *Feedback, Flexible Progression, Individualized Instruction, Information Processing, *Instructional Technology, Item Analysis, Program Design, Retention, Student Evaluation, Systems Approach, *Testing Programs
IDENTIFIERS CAM, *Comprehensive Achievement Monitoring

ABSTRACT

This paper considers the design and analysis of feeding back information about achievement test results to teachers and students in a high school classroom. An introductory section discusses the designing and administering of tests throughout the year in such a fashion as to provide pre-instruction, post-instruction, and retention achievement levels of groups of students in specific areas of the course. Three basic types of information are considered: feedback for individual students, feedback for homogeneous groups of students, and feedback for curriculum evaluation. The design of the formats of the output has evolved during the four years that the Comprehensive Achievement Monitoring (CAM) project has been in operation. Teachers' comments, as well as statistical analysis, has produced feedback information designed so that both teachers and students can dynamically make decisions concerning instructional activities. Data from several years of classroom monitoring in the subject areas of history and mathematics is presented. (Author)

FEEDBACK INFORMATION IN THE COMPREHENSIVE MONITORING OF EDUCATIONAL ACHIEVEMENT [1,2]

Paul D. Pinsky

Department of Operations Research
Stanford University
Stanford, California

[1] A paper presented at the 11th American Meeting of the Institute of Management Sciences, Los Angeles, California, October, 1970.

[2] The research and development reported herein was performed pursuant to a grant from the Charles F. Kettering Foundation to the principal investigator, Dwight W. Allen, and under the project director, William P. Gorth, both of the School of Education, The University of Massachusetts, Amherst, Massachusetts 01002.

1.0 Introduction

This paper discusses the types of information that can be derived from achievement tests in a classroom learning environment. Initially, four classes of information are defined and discussed. Then the details of a simplified testing scheme designed to obtain comprehensive achievement data are presented along with the computerized output from such a testing scheme. Next a brief discussion of the adaption of this simplified design to more complex classroom learning environments and a discussion of actual classroom experience is presented.

The fifth section discusses the role of the time-shared computer in the processing of such data. The last section contains some thoughts on the research and development efforts that are needed to successfully implement on a large scale the testing schemes presented in the paper.

2.0 Types of Information

This section discusses four basic types of information that can be derived from achievement tests.

2.1 Information for dynamic decision making about individuals

In many schools throughout the country, the curriculum (usually stated in behavioral or performance terms) is designed to allow students to progress at their own rate. In most instances this means that a student studies a unit or package of material until he feels that he has mastered the concepts involved and then takes a test (often called a posttest) on this material. Based upon his performance on this test, some decision is made, i.e., he goes on to the next unit in sequence (he "passed"), he goes back and continues studying the same material

(he "failed"), or he goes on to some enrichment area ("partial failure"). The main requirement of this type of information is that the data have a high degree of statistical reliability (see Kriewall [5] and Lord and Novick [6]). Superficially, this decision process seems quite simple. However, attempts to develop an underlying decision structure for controlling the path that students follow in a curriculum have run into serious problems (see Groen and Atkinson [3], Karush and Dear [4], Smallwood [9], and Suppes [10]).

2.2 Diagnostic information

Achievement test data can be used to indicate those areas in the curriculum that individual students have not mastered and suggest instructional activities to aid the student in attaining mastery (if mastery be the student's goal). In today's flexibly scheduled schools, this information will become increasingly important as the students are scheduled more independent study time. In practice, the prescription of instructional activities is usually based upon statistically unreliable data (often being based upon one or two multiple choice test items). A previous paper by the author discusses the above in the context of a resource allocation problem, the resource being the student's testing time (Pinsky [7]).

As an example of the difference between decision making and diagnostic information consider a pretest on a unit or package of material given to a student. A decision is made (based upon the percentage of correct responses) as to whether the student "knows" the material in the unit and should proceed to the next unit. If the student "fails" the

pretest, his responses to individual items are used in a diagnostic fashion to direct his study within the unit.

2.3 Summative group information

The first two types of information are concerned with the individual student. In many instructional environments, information about groups of students is useful as outlined below.

2.3.1. When a lecture is being given to a group of students, decisions need to be made as to what material should be covered in the lecture, i.e., should the teacher review some material, present enrichment topics, proceed with the normal course plan, skip ahead in the course plan, etc.

2.3.2. Most input-output analysis in education (such as planned-programming-budgeting systems) use group data. These types of analysis often require the groups' pre-instruction, immediate post-instruction, and retention achievement levels; a value function on these various achievement levels; and the cost of the instructional processes.

2.3.3. Research into the effectiveness of various instructional processes usually requires summative group achievement data over time as one of the criterion for judging the value of these processes.

2.3.4. Achievement test data about groups is often statistically more reliable than test data about individuals (see Pinsky [7]).

2.4 Curriculum evaluation information

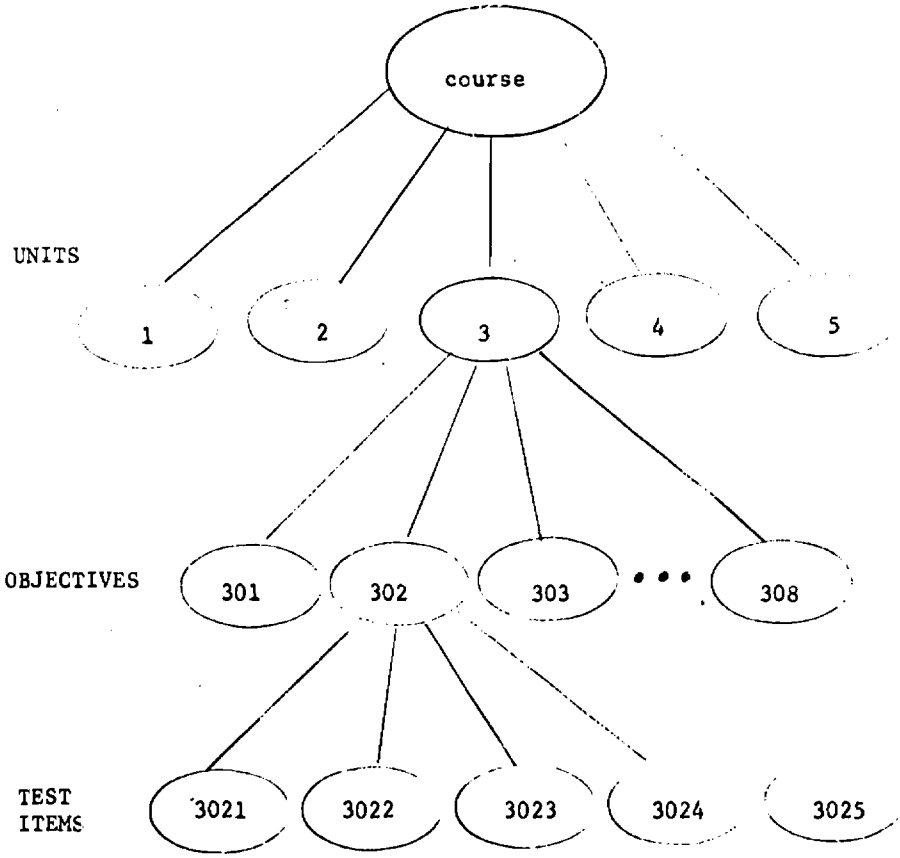
With the growing trend in education toward the use of performance (or behavioral) objectives and test items designed to measure mastery of these objectives, there is a need for information to use in the evaluation

of the objectives and test items. When multiple choice test items are used, the problem of constructing and revising the pool of items can be substantial, especially when the writing is being done at the local level. Information that is useful for curriculum evaluation is the pre-instruction, immediate post-instruction, and retention achievement levels of groups of students on all objectives in the course. These achievement levels are obtained from the performance of students on test items related to each objective.

3.0 A simple comprehensive achievement monitoring design

Consider a one semester course structured as shown in figure 1. This course has 5 units, 8 performance objectives per unit, and 5 test items per objective. The units are numbered 1 through 5; the objectives have a three digit number - the first digit indicates the unit which the objective is in, while the last two digits are the unique objective number within the unit; the test items have a four digit number - the first three digits indicate what objective the item is related to, while the fourth digit is the unique item number within the objective. This course has an enrollment of 210 students. Each student attends one large lecture (with 210 students), two medium group sections (with 30 students), and one lab section (with 10 students) each week. All students cover approximately the same material each week, i.e., the teacher paces the assignments.

In this course one might like to know (in a statistically reliable sense) each student's achievement level on each objective before, immediately after, and well after instruction. However, given



The Structure of a Hypothetical Course

Figure 1.

a fixed amount of testing time per student each week, this is not possible (Pinsky [7]). But the above achievement levels can be measured with relatively good statistical reliability for groups of students (the medium group and lecture sections) by using the following testing scheme. In figure 2 are presented 10 interchangeable test forms with 20 items per form. These forms are comprehensive in the sense that they cover material related to all five units of the course, and are interchangeable in that they are essentially 10 different forms of a 20 item final examination for the course. Note that each form contains 4 items per unit, and that every item on a form is related to a different objective. Every two forms (i.e., 1 and 2, 3 and 4, etc.) contains a test item related to all 40 objectives in the course.

At the beginning of the semester course, each student responds to one of these monitor forms, and each form is responded to by 21 students. Furthermore, three students in each medium group respond to each of the 10 forms. Then two weeks later, the process is repeated, but with each student taking a different test form. For example, if a student took form 1 the first time, then he takes form 2 the next time, etc. This process is repeated every two weeks throughout the semester. For details, rational, and variations of the above design, refer to Gorth [1], [2]. Every testing period, the students' responses are processed by computer and the following output is generated. The examples presented here are the output for the seventh monitoring period.

3.1 Individual student report

Each test the student responds to is processed by the computer

Monitor Form Number

Question Number	1	2	3	4	5	6	7	8	9	10
1	1013	1021	1012	1025	1014	1023	1011	1024	1015	1022
2	1041	1035	1044	1033	1042	1034	1045	1031	1043	1032
3	1053	1063	1054	1064	1051	1061	1055	1065	1052	1062
4	1081	1073	1085	1075	1084	1071	1083	1072	1082	1074
5	2015	2021	2014	2025	2011	2022	2013	2024	2012	2023
6	2043	2035	2044	2033	2042	2031	2045	2034	2041	2032
7	2054	2061	2055	2062	2051	2065	2052	2063	2053	2064
8	2071	2084	2072	2082	2073	2085	2074	2083	2075	2081
9	3024	3011	3023	3012	3025	3014	3021	3015	3022	3013
10	3031	3043	3035	3045	3032	3041	3033	3044	3034	3042
11	3053	3065	3051	3063	3052	3064	3055	3062	3054	3061
12	3083	3071	3085	3075	3081	3072	3082	3074	3084	3073
13	4011	4025	4015	4024	4012	4021	4013	4023	4014	4022
14	4035	4041	4034	4043	4033	4042	4032	4044	4031	4045
15	4064	4054	4063	4055	4062	4051	4065	4053	4061	4052
16	4073	4082	4072	4084	4071	4085	4074	4083	4075	4081
17	5014	5021	5015	5023	5011	5022	5012	5024	5013	5055
18	5031	5045	5035	5044	5034	5043	5032	5042	5033	5041
19	5064	5051	5061	5052	5063	5053	5064	5055	5065	5054
20	5083	5075	5082	5074	5081	5073	5084	5072	5085	5071

test item numbers

10 Interchangeable Test Monitor Forms

Figure 2.

and the results are output to the student as shown in figure 3. The report contains the student's name, the monitoring period, the date (coded by year, month, and day), the test form the student took, the percentage of correct responses on this monitor and all preceeding monitors (past scores) taken by the student. The responses to individual items are reported by the objective to which the item is related (OBJ), the type of response (A) (+ = correct, - = incorrect, 0 = no response), and whether the student should have been able to correctly answer the question based upon his instructional work (INS).

3.2 Teacher summary data

The computerized teacher summary output is shown in figure 4. This output list is for one of the lab sections. For each student the output contains his number, name, the form he took during the latest monitoring period (period 7 in this example), whether the form he took was the correct one (an X indicates an incorrect form), and his past scores represented in percentage of correct responses.

3.3 Group summary data

For any group of students and any group of test items, the computer calculates the total number of responses of each student group to each question group and the percentage of theses responses that are correct. In the example shown in figure 5, there is one student group (medium group section 3) and six question groups (all test items from all units, all test items from unit 1, all test items from unit 2, all test items from unit 3, all test items from unit 4, and all test items from units 5).

( ) AMES, TOM FORM 2
PERIOD 7 - 691208

PERCENTAGE CORRECT IS 65

OBJ	A	INS	OBJ	A	INS	OBJ	A	INS	PAST SCORES PD	PCT
102	+	YES	504	+		0			1	15
103	+	YES	505	-		0			2	30
106	+	YES	507	-		0			3	25
107	+	YES	0			0			4	50
202	-	YES	0			0			5	40
203	-	YES	0			0			6	55
206	0	YES	0			0			7	55
208	+	YES	0			0			8	0
301	+	YES	0			0			9	0
304	+	YES	0			0			10	0
306	+	YES	0			0			11	0
307	+	YES	0			0			12	0
402	+	YES	0			0			13	0
404	+	YES	0			0			14	0
405	+		0			0			15	0
408	0		0			0			16	0
502	0								17	0
									18	0
									19	0
									20	0

( )

OBJ = performance objective
A = answer, + = correct, - = incorrect, 0 = no response
INS = instruction completed on the given performance objective

Individual Student Report

Figure 3.

PAST SCORES - PERCENTAGE CORRECT

			MONITORING PERIOD									
NUMBER	NAME	FORM	1	2	3	4	5	6	7	8	9	10
21080	AMES, TOM	2	15	30	25	50	40	55	65	0	0	0
21120	BOLT, JIM	4	15	25	20	30	25	40	40	0	0	0
21204	DOBBS, SUE	7	30	25	40	60	50	60	70	0	0	0
21402	EVANS, ANN	7X	20	25	30	30	50	60	60	0	0	0
21450	HOPE, JUNE	1	50	50	60	65	60	75	80	0	0	0
21608	PINSKY, ANN	10	40	50	0	40	60	70	65	0	0	0
21710	SMITH, BILL	3	0	0	40	50	60	60	70	0	0	0
21820	TODD, PAUL	9	10	20	25	30	50	70	50	0	0	0
21934	WADE, SUE	0	40	0	0	50	0	50	0	0	0	0
21980	YOUNG, RAY	5	30	30	40	35	50	45	60	0	0	0

FORM = form taken during the latest monitoring period (7 in this example)

0 means no test form taken

X means incorrect test form taken

Teacher Summary Report

Figure 4.

GROUP SUMMARY DATA - PERCENTAGE CORRECT

STUDENT GROUP 3

QUESTION GROUP	MONITORING PERIOD 1	2	3	4	5	6	7	8	9	10
1	20	30	38	50	52	55	62			
2	18	55	80	72	68	60	72			
3	30	35	48	80	82	78	85			
4	32	40	42	50	60	70	64			
5	12	10	12	36	42	50	70			
6	8	10	8	12	10	18	20			

GROUP SUMMARY DATA - NUMBER OF RESPONSES

STUDENT GROUP 3

QUESTION GROUP	MONITORING PERIOD 1	2	3	4	5	6	7	8	9	10
1	600	580	560	580	540	600	580			
2	120	116	112	116	108	120	116			
3	120	116	112	116	108	120	116			
4	120	116	112	116	108	120	116			
5	120	116	112	116	108	120	116			
6	120	116	112	116	108	120	116			

STUDENT GROUP 3 = medium group section 3

QUESTION GROUP 1 = all test items from all units

" " 2 = " " " " unit 1

" " 3 = " " " " " 2

" " 4 = " " " " " 3

" " 5 = " " " " " 4

" " 6 = " " " " " 5

Group Summary Data

Figure 5.

At the end of the course, an item analysis program is run on the computer to provide curriculum evaluation information. A record is kept of the day or days of instruction on each performance objective throughout the semester. This data along with the students' response data (which includes the date of each response) to all monitors throughout the semester are used as input into the item analysis program. Then, for any group of students, the program calculates the total number of responses and the percentage of correct responses to each test item in the course for three time periods - PRE-INSTRUCTION, POST-INSTRUCTION, and RETENTION. These time periods are specified by the user. Then the data for all items related to the same objective is totaled to produce output relative to objectives as shown in figure 6. In this example, PRE-INSTRUCTION is defined as any time before instruction on objective 302, POST-INSTRUCTION is defined as from 0 to 40 days since instruction, and RETENTION is defined as longer than 40 days since instruction.

4.0 General comprehensive achievement monitoring designs

The structure of most courses in schools today is not as simple as the hypothetical course previously discussed. This section is concerned with the adaption of this simple comprehensive monitoring design to actual curriculum. Two such designs are presented in some detail and mention is made of several other courses in which such designs are currently being employed. The author wishes to emphasize his feelings that monitoring designs must be adapted to fit the course structure and not vice versa.

ITEM ANALYSIS - STUDENT GROUP 1

		PRE-INSTRUCTION		POST-INSTRUCTION		RETENTION	
OBJ	302	NUM	PER	NUM	PER	NUM	PER
	ITEM 3021	400	20	310	80	299	70
	ITEM 3022	402	23	312	30	201	28
	ITEM 3023	410	62	307	90	303	90
	ITEM 3024	391	50	310	60	298	80
	ITEM 3025	397	35	311	50	300	82
OBJ	TOTAL	2000	38	1550	62	1501	70

NUM = total number of responses

PER = percentage of correct responses

PRE-INSTRUCTION is defined as before instruction on objective 302

POST-INSTRUCTION is defined as from 0 to 40 days since instruction

RETENTION is defined as longer than 40 days since instruction

STUDENT GROUP 1 = all students in the course

Curriculum Evaluation Data

Figure 6.

4.1 A tracked curriculum

A 9th grade mathematics course in Jackson High School in Portland, Oregon is divided into five tracks - 1. arithmetic, 2. very slow algebra, 3. slow algebra, 4. algebra, 5. advanced algebra. Each track contains 16 units. At the beginning of the year, every student (approximately 300 total) is assigned to a track and starts at unit 1 in that track. The students study material related to their unit (individually and in groups) for two weeks, and then take a unit posttest covering the material in the unit they just studied. These tests are optically scanned by machine and input into a computer which processes the responses and assigns a unit of study for each student for the next two-week period. This unit may be the next one in sequence in the same track, the same unit, or a unit in a different track. In addition, the computer uses the posttests to prescribe review work in the unit just completed for each student.

Every other week the students take comprehensive monitors similar to the ones described in the previous section. The design of these monitors is more complex than those presented in section 3 because all the students are not studying the same material at the same time. These monitors provide diagnostic information (primarily pointing out those areas of the course that the student has previously studied and failed to exhibit mastery of), summative group information, and curriculum revision information.

4.2 A basic skills course

A 9th grade mathematics course in Sequoia High School in Redwood

City, California is divided into 4 blocks. The course (covering basic mathematical skills) is often taken by students to fulfill the mathematics requirement for graduation from high school. The student completes the course by showing competence in each of the 4 blocks successively, each student progressing at his own rate. Every six days the student takes a monitor covering the block of material that he is working on. His answers to the constructed response questions are graded by the teacher and coded by an aide onto machine-readable data cards for input into the computer. A student exhibits competence in a block by answering correctly 90% of the questions related to that block on any monitor, or by answering correctly 80% of the questions related to the block on two successive monitors. The individual student report generated by the computer indicates what questions should have been correctly answered if the student were progressing at the minimal pace needed to complete the course in one year. In addition, the monitors contain test items related to previous blocks, i.e., review questions. Once having shown competence in a block, the student is expected to be able to exhibit mastery on the material several months later. If a student fails to correctly answer a test item from a block that he has shown competence in, the computer generates a recommended review exercise. On an experimental basis, the course performance objectives will be sent to the parent of each student. Then, periodically throughout the year, the computer-generated student reports will be sent to the parents.

The following achievement monitoring projects are being implemented this fall as an outgrowth of the initial CAM developmental work. In the Sequoia Union High School District in Redwood City, California,

courses in arithmetic, algebra, biology, earth science, general science, geography, and physical education will be involved with computer processing of comprehensive monitors. In Hopkins, Minnesota a demonstration evaluation center is being developed under a title III grant. This fall courses in algebra, 3rd and 4th grade mathematics, and social studies will have monitoring data processed by computer. William Gorth of the University of Massachusetts and Robert O'Reilly of the New York State Department of Education are directing similar projects in Massachusetts and New York ([1], [2]). In past years, the CAM project has monitored courses in history, algebra, and trigonometry at Kailua High School in Kailua, Hawaii, and two algebra courses in Portland, Oregon.

5.0 Potential for time shared terminals

The current method used in processing the test data described above is as follows: the students respond to the test items on answer sheets or cards; these sheets or cards are taken to a local data center and keypunched or machine scanned onto punched computer cards; these cards are taken to the computer for batch processing; the printed output is taken back to the school; and finally the results are distributed to the students and teachers. The average time between taking a test and getting the results back is approximately three days. The computer cost to process 250 tests (which includes printing individual student reports, teacher summary reports, and group summary data; and updating the master tape file) is approximately ten dollars when done on an IBM 360/65 at Stanford University. Although no data is available on the cost of getting the student responses to the computer, the author assumes

that it is much higher than ten dollars.

An alternative data collection procedure would be to have the students input their responses on a remote terminal (either by typing the responses directly, or by using a card reader attached to the terminal); have the computer process the student's responses in real time and print out the individual student report; and store the responses on disk or drum. Then, during the evening or night shift when the computer is running in batch mode, the teacher summary reports and summative group data could be processed and printed out on a high speed printer at the computer installation, and the master tape file updated. The output could then be sent to the teacher the following morning.

The advantage of the second procedure is that the decision and diagnostic information is provided immediately, while the more costly processing of summative data and updating of files is done at a more economical time. Title III projects in Hopkins, Minnesota and Redding, California (in the Enterprise Elementary School District) are now in the initial phases of developing such a real-time system. It is hoped that in the near future the author will be able to present some cost estimates concerning alternative configurations for processing instructional management data.

6.0 Conclusion

This paper has discussed the types of information that can be generated by computer processing of achievement tests. Much research and developmental work is needed to evaluate the effectiveness of such information. Some research (Richards [8]) has been done concerning the

the effectiveness of criterion referenced testing by subject area, i.e., mathematics, physical sciences, social sciences, languages, and humanities, but much more is needed. Little research has been done as to the effectiveness of such procedures by grade level, i.e., primary (K-3), intermediate (4-8), secondary (9-12), and college. For example, in the primary levels, written tests are often not feasible, and therefore computerized processing may not be economically attractive.

Detailed cost analysis of various degrees of computer processing of insturctional data is needed. For instance, when a real-time computer terminal is installed in a classroom for instructional data processing purposes, does the teacher devote more time to instruction and less time to clerical work? Does instant feedback of test data for individual students result in higher achievement levels over the year? It is the author's hope that persons with a management science-operations research background will be able to help the educational community obtain quantitative answers to the above types of questions.

References

[1] Gorth, W. P. (organizer), "Designing Instructional Systems with Longitudinal Testing Using Item Sampling", a symposium presented a the Annual Meeting of the American Education Research Association in Minneapolis, Minn., March, 1970.

[2] Gorth, W. P., Schriber, P. E., and O'Reilly, R. P., Training Manual for Comprehensive Achievement Monitoring, The New York State Education Department, Division of Research, Albany, N. Y., June, 1970.

[3] Groen, G. J., and Atkinson, R. C.,"Models for Optimizing the Learning Process", Psychological Bulletin, Vo. 66, #4, October, 1966.

[4] Karush, W., and Dear, R. E., "Optimal Strategy for Item Presentation in a Learning Process", Management Science, Vo. 13, #11, July, 1967.

[5] Kriewall, T. E., "Applications of Information Theory and Acceptance Sampling Principles to the Management of Mathematics Instruction, Part I", Technical Report No. 103, Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, Wisconsin, October, 1969.

[6] Lord, F. M., and Novick, M. R., Statistical Theories of Mental Test Scores, Addison-Wesley, Boston (1968).

[7] Pinsky, P. D., "A Mathematical Model for the Measurement and Control of Classroom Achievement", a paper presented at the Meeting of the Operations Research Society of America, Washington, D. C., April, 1970.

[8] Richards, J. M., Recent Research on the Assessment of Student Accomplishment in College, American Institute for Research, Palo Alto, California, February, 1970.

[9] Smallwood, R. D., Quantitative Methods in Computer-Directed Teaching Systems, Institute in Engineering-Economic Systems, Stanford, California, 1967.

[10] Suppes, P., "Problems of Optimization in Learning a List of Simple Items", Human Judgement and Optimality, edited by Shelly and Bryan, John Wiley and Sons, New York, 1964.