BOSTON UNIVERSITY
SCHOOL OF EDUCATION

THE USE OF INTERACTIVE, COMPUTER-MANAGED INSTRUCTION TO CONTROL THE QUALITY OF SELF-PACED TRAINING

JESSE M. HEINES

S.B., Massachusetts Institute of Technology, 1970
M.S., University of Maine at Orono, 1974

Submitted in partial fulfillment of the
requirements for the degree of

Doctor of Education
in
Educational Media and Technology

April, 1979

READER'S APPROVALS

First Reader: Louis P. Aikman

Dr. LOUIS P. AIKMAN
Prof. of Education
Dept. of Humanistic and Behavioral Studies
School of Education
Boston University

Second Reader: Gaylen B. Kelley

Dr. GAYLEN B. KELLEY
Prof. of Education and Chairman
Dept. of Educational Media and Technology
School of Education
Boston University

Third Reader: Ethan A Scarl

Dr. ETHAN A. SCARL
Assoc. Prof. of Mathematics
College of Liberal Arts, and
Assoc. Prof. of Computer Science
Metropolitan College
Boston University

Fourth Reader: C. Victor Bunderson

Dr. C. VICTOR BUNDERSON
President, WICAT, Inc. and
Adjunct Prof. of Instructional Science
Brigham Young University

DEDICATION

This dissertation is dedicated to Scott Douglas Heines and Russell Gordon Heines in the hope that they will always remember The Little Engine That Could: if you think you can, you can.

ACKNOWLEDGEMENTS

The author wishes first to express his appreciation to the members of his doctoral committee for their guidance and advice. In particular, he thanks Dr. Victor Bunderson for extensive comments on the dissertation proposal and Dr. Louis Aikman for long term guidance during the entire doctoral program.

Second, the author thanks Digital Equipment Corporation for its financial support of his doctoral studies. Thanks are due to Mr. Richard Kirschbaum, Mr. Peter McVay, and Mr. John Bowers for their work in writing The BASIC Primer. The contributions of Mr. Del Lippert, Mr. Wayne Saunders, and Mr. Dennis Brown also deserve acknowledgement for their roles in making this study possible.

A debt of gratitude is owed to Ms. Carlene Washburne for assistance in attending to the details of completing the doctoral program. The author also thanks Ms. Marion Borders for assistance in conducting the doctoral study and in typing the dissertation.

The author also wishes to express his appreciation to his parents, Mr. Philip Heines and Mrs. Dorothy Heines, and to Mr. James Thacker for their constant encouragement during the peaks and valleys of the doctoral program.

Finally, the author wishes to offer an understated "thank you" for the contributions and sacrifices made by his wife, Evelyn Heines. Without her constant assistance, support, and love, this doctoral program could never have been concluded. "Gospodeen Slone lyoobet svoyoo zhenoo."

ABSTRACT

The purpose of the study was to demonstrate the feasibility of using a computer system to control the quality of self-paced training. An interactive, computer-managed instruction (CMI) system was built to complement a self-paced training course on BASIC language programming. This system assessed students' work to determine their statuses in an established module hierarchy and administered pretests and posttests on the modules for which they had met the prerequisites. The CMI system was highly human-engineered so that it could be used without the presence of an instructor. The system's mastery algorithm was based on a sequential probability test ratio intended to reduce test lengths and testing time without sacrificing test reliability.

The complete training package was used for teacher training in two public school districts and one junior college for a period of approximately two months. During this time, the CMI system recorded usage data which was periodically transferred to magnetic tape and mailed to the author for analysis.

Analysis of the specific usage data yielded the following major results:

(1) The system was successful in assuring that students worked through the module tests in accordance with the learning hierarchy prescribed by the course developers.

(2) The system was generally successful in discouraging excessive test taking even though students were free to repeat the computer-administered tests as often as they liked.

(3) The mastery algorithm based on a sequential probability test ratio was able to reduce test lengths significantly on all tests except pretests on which examinees were classified as masters.

(4) Test length reductions achieved by the sequential probability test ratio did not significantly impact the criterion-referenced test reliability.

(5) The system was successful in collecting data for criterion-referenced item analysis, but assessment of this data for evaluating the validity of the items was inconclusive due to the small number of test administrations for some modules and the lack of a qualitative definition of a "good" criterion-referenced item in terms of its item analysis results.

PREFACE

This dissertation concerns a study on one aspect of customer training at Digital Equipment Corporation. Digital is a manufacturer of minicomputer components and systems and is the author's employer. Although this study was restricted to customers of this specific company, the author feels that most other computer manufacturers face or will soon face similar customer training problems.

TABLE OF CONTENTS

	page
Dedication	ii
Acknowledgements	iii
Abstract	v
Preface	vii
List of Tables	xi
List of Figures	xii
Chapter	
1 THE PROBLEM AND ITS SETTING	1
The Evolution of Customer Training at Digital Equipment Corporation	1
Statement of the Problem	2
Research Questions	5
2 REVIEW OF RELATED LITERATURE	8
Availability of Related Literature	8
The Evaluation of Self-Paced Instruction	10
Criterion-Referenced Testing	11
Basic CRT Theory	11
The Dichotomous Outcomes Model	12
Domain-Referenced Testing	13
Decisional Analysis Models	16
CRT Reliability	21
Computer-Assisted Testing	27
Test Printing Systems	28
Test Construction Systems Using Item Banks	31

Chapter page

Algorithmic Approaches to Item Construction 32
Interactive, Branching Tests 34

Computer-Managed Instruction 45

Batch CMI 46
Interactive CMI 49

3 DESIGN AND IMPLEMENTATION OF THE STUDY 51

Description of the Self-Paced Course 51
Description of the CMI System 54
The Mastery Decision Model 61
The Reliability Model 69
Conducting the Study 70

4 RESULTS OF THE STUDY 73

The First Subproblem 73
The Second Subproblem 77
The Third Subproblem 89
The Fourth Subproblem 96
The Fifth Subproblem 102

5 DISCUSSION 111

Summary of Results 111
General Comments 112
Directions for Future Study 113

Appendices

A Directions to Students on Using the CMI System 116

B Test History Data for Wachusett Regional School District and Rhode Island Junior College 120

C Test Length Data for Individual Test Sites 143

D Test Reliability Data for Individual Test Sites 147

References Cited 152

Vita 160

LIST OF TABLES

Table | page

1 Pools for Constructing Items on Electron Configurations 33

2 Sequential Testing Parameters for Pretests and Posttests 65

3 Distribution of Learner's Academic Levels 75

4 Summary Test Time and Related Data for Each Site 90

5 *A Priori* Error Probabilities and *A Posteriori* Test Lengths 95

6 Model Reliability Data 97

7 Test Reliabitlity Data for All Sites Combined on Modules 2 through 16 98

8 Computation of the Kologorov-Smirnov D for the Data in Table 7 102

9 Observed Ranges of Discrimination Indices for Items in the Core Modules 110

LIST OF FIGURES

Figure		page
1	Hively's Domain-Referenced Testing Model	15
2	Ferguson's Method for Determining Proficiency on a Criterion-Referenced Test	20
3	Frequencies of Agreement Between Mastery and Non-Mastery Classifications on Two Sets of Test Data	24
4	Sample Classification Frequencies	26
5	Hansen's Sequential Item Tree Network	36
6	Strategies of Adaptive Testing	42
7	Module Hierarchy for <u>The BASIC Primer</u>	52
8	Relation Between On-line CMI System and Off-line Learning Modules	53
9	Program Components of the CMI System	54
10	First Display Generated by the CMI Program . . .	55
11	Sample Display of a Multiple Choice Item as Presented to a Student	58
12	Sample Display of a Test Item in Review Mode . .	60
13	Graphs of the Pretest and Posttest Decision Rules	68
14	Test History Data for Falmouth High School . . .	78
15	Test Length Data for All Sites Combined on Modules 2 through 16	93
16	Item Analysis Data for All Sites Combined on Module 2	104

Chapter 1

THE PROBLEM AND ITS SETTING

The Evolution of Customer Training at Digital Equipment Corporation

Digital Equipment Corporation sold its first computers in 1961 and offered its first customer training courses in 1962. These courses were conducted in the basement of the company's manufacturing facility in Maynard, Massachusetts. Today, Digital offers over 100 customer training courses that are available in six regional training centers in the United States and eleven centers in other parts of the world.

In the past few years, three factors have heavily influenced the evolution of customer training at Digital:

(1) drastic decreases in the costs of minicomputer systems due to new technologies,

(2) significant increases in the costs of customer training due to inflation, and

(3) elimination of "previous computer experience" as a prerequisite for using many of Digital's newer systems.

Even with the regional training centers, the first two factors often make it financially unfeasible for some customers to send their employees to a Digital facility for training. For example, a company that buys a $10,000 minicomputer system cannot be expected to spend several hundred dollars to cover tuition, travel, and per diem expenses for each employee that they wish to train to use that system. The third factor has created a large class of users who do not have the prerequisite skills for Digital's standard courses, even though their employers may find it financially feasible to send them to a Digital facility for training.

In 1975, Digital began developing self-paced instruction (SPI) courses to deal with these evolutionary factors. These courses are written in a modularized, criterion-referenced format similar to that developed by Robert Mager (1974). Many of these courses use audiovisual media for delivering instruction, and all are designed for use in an on-the-job customer environment without requiring the presence of a Digital instructor. This dissertation deals with one approach to controlling the quality of these self-paced customer courses.

Statement of the Problem

The introduction of self-paced training has created numerous quality control problems. The three most visible problems at Digital are:

(1) it is difficult to control the sequence in which students work through learning modules,

(2) it is difficult to assure that students master the objectives in each module that they study, and

(3) it is difficult to get accurate feedback on the use of the training program.

This research involved the development, implementation, and evaluation of a computer-managed instruction (CMI) system to address these problems. This CMI system runs on a broad range of computer systems manufactured by Digital Equipment Corporation to make it usable in the largest possible number of customer environments. On a functional level, the CMI system:

(1) directs students on which module to study next based on their test results and the defined course structure,

(2) generates, administers, and scores pretests and posttests on each learning module, and

(3) collects data on the use of the training program.

The first subproblem. The first subproblem was to develop a workable CMI system that could be customer-installed and used by computer-naive learners. These qualities were essential for keeping the cost of the system low and for making it a viable tool for introductory self-paced courses. Another aspect of this subproblem was to implement a scheme for gathering system usage data from customer sites for analysis.

<u>The second subproblem</u>. Self-paced courses are usually divided into a series of modules that are arranged in a specific learning hierarchy. Learners can misuse these courses by studying the advanced modules before mastering the prerequisite ones and taking tests over and over without studying in between. The second subproblem was therefore to design the CMI system so that learners were forced to work through the modules in the prescribed hierarchy and so that excessive test taking would be discouraged.

<u>The third subproblem</u>. Learners can also misuse the tests that are included in industrial SPI course packages by hurrying through them. They may skip the tests entirely, read the tests and look at the answers simultaneously, or take the tests once, check their answers, and move on without really trying to understand why they answered certain items incorrectly. Working on the assumption that these behaviors occur because learners like to get through tests as quickly as possible, the third subproblem was to design the CMI system so that its use required a minimum amount of time.

<u>The fourth subproblem</u>. When test lengths are reduced to minimize testing time, reliability is usually sacrificed. The fourth subproblem was therefore to implement an algorithm that maintained reliability with short

tests, and to devise and implement a scheme for assessing the validity of this algorithm.

The fifth subproblem. The fifth subproblem was to devise and implement a scheme for assessing the validity of the CMI system's test items. This was necessary because all of the reliability data assumed that each test item was an equivalent, valid measure of one of the course's objectives.

Research Questions

Five research questions were formulated that relate respectively to each of the five subproblems.

The first question. How well did the entire quality control system operate?

This question was addressed from five points of view:

(1) How easy was it for customers to install the CMI system?

(2) What problems, if any, did learners have in using the CMI system?

(3) How reliable was the software?

(4) What percentage of students who began the course actually completed it?

(5) How easy was it to collect data on the system's usage?

<u>The second question</u>. Was the CMI system effective in controlling the sequence in which learners studied the course modules and in discouraging excessive testing?

This question was addressed by analyzing chronological test history data. This data provided information on the order in which each learner went through the course modules and the elapsed times between each test.

<u>The third question</u>. Did use of the CMI system require a minimum amount of time?

This question was addressed by analyzing records of:

(1) the amount of time that each learner spent on-line,

(2) the number of times that each learner used the CMI system,

(3) the amount of time that each learner spent taking tests,

(4) the number of tests that each learner took, and

(5) the number of items administered on each individual test.

<u>The fourth question</u>. Was reliability maintained when short tests were administered?

The CMI system employed a sequential probability test ratio to reduce the number of items necessary to determine a learner's mastery state on each module. This algorithm terminated testing whenever decisions on student mastery could be made with specified levels of certainty.

To assess the validity of this algorithm, every fifth test was extended to a long test of 30 items, and the mastery decisions made on these extended tests were correlated with those made on the shorter versions.

The fifth question. Were the test items administered by the system valid?

This question was addressed by analyzing criterion-referenced item analysis data. The number of times that each alternative was chosen was tallied for each test item in four categories:

(1) pretests on which mastery decisions were made,

(2) pretests on which non-mastery decisions were made,

(3) posttests on which mastery decisions were made, and

(4) posttests on which non-mastery decisions were made.

These data allowed the computation of pretest/posttest and master/non-master discrimination indices.

Chapter 2

REVIEW OF RELATED LITERATURE

Availability of Related Literature

There are several subject areas that relate to this study. The most prominent of these are:

(1) the evaluation of self-paced instruction (SPI),

(2) criterion-referenced testing (CRT),

(3) computer-assisted testing (CAT),

(4) computer-managed instruction (CMI), and

(5) the application of SPI, CRT, CAT, and CMI to industrial training.

Most of the more recent literature in the first area is subsumed under CRT and CAT, and a considerable amount of literature is available on these areas (Heines, 1975). Literature on CMI is more scarce, possibly because this application is even newer than the first three.

Literature on the fifth area, however, is virtually non-existant. Of the approximately 235 articles published in Training in the 18 month period ending March 1978, only five dealt with self-paced or individualized instruction, including one entitled, "Let's Jump Off the Individualized

Instruction Bandwagon" (Maginn, 1977). (This periodical is widely distributed to professionals in industrial training organizations.) The only time that this periodical discussed the use of computers in training during this period was in a 10-page "Special Report: Computer-Assisted Instruction" (McLagan et al., 1977). Not one of the subarticles in this section described a fully operational system, and the 14 item "reader's starter list" for additional information included such general works as an IBM User's Guide, a book on man-computer dialogues, a book entitled Computing Machinery and Intelligence, and an article on the historical development of CAI. Again, no descriptions of existing, operating systems were referenced.

The above statements should not be construed to indicate that the author believes that such systems do not exist. Rather, they simply are not often discussed in the literature. This lack of literature on industrial applications may exist because industry is just beginning to apply these techniques, because industrial trainers are not as heavily motivated to report their work as academicians are, and/or because much of the information on industrial training is proprietary.

The second problem, the lack of publishing by industrial trainers, may be lessening. The Association for the Development of Computer-based Instructional Systems (ADCIS) has just recently formed a Special Interest Group in

Computer-Based Training. This group publishes a quarterly newsletter which reports on instructional computer applications in business, government, and industry. Several articles in the three newsletters published to date are referenced in this literature review.

The Evaluation of Self-Paced Instruction

The problem of evaluating a large group of students moving through individualized material at different rates became immediately apparent with the introduction of "audio-tutorial" instruction by Postlethwait in 1962. Postlethwait's evaluation model required that students be paced for weekly quiz sessions and that they meet at a specific time to take exams. Butzow and Pare (1972) modified this approach to allow students greater freedom of scheduling. Their model did not require students to work at any specific pace, but simply obliged them to fulfill a criterion-referenced grade contract by the end of one semester (Pare, 1973).

Both of these evaluation models used manual techniques for administering, evaluating, and recording student examinations. Heines (1974) extended the latter model by implementing a computer-managed instruction system to evaluate some aspects of the audio-tutorial program and provide qualitative data on student performance. Most SPI evaluation systems now use some sort of computer processing to handle one or more facets of the evaluation process. Discussion of these systems is therefore included in the sections that follow.

Criterion-Referenced Testing

Basic CRT theory. Criterion-referenced testing (CRT) differs from norm-referenced testing (NRT) in the following way:

> Norm-referenced measures are those which are used to ascertain an individual's performance in relation to the performance of other individuals on the same measuring device. . . Criterion-referenced measures [are used] to ascertain an individual's status with respect to some criterion, i.e., performance standard. (Popham and Husek, 1969)

Glaser (1963) adds that NRT provides "information about the capability of a student compared with the capabilities of other students", while CRT provides "explicit information on what the individual can and cannot do".

Cox (1971) feels that "it is possible for a single test to yield both norm-referenced and criterion-referenced information". This posture appears to oppose that held by Glaser, who feels that the choice of items differentiates test design. Many researchers (Adams, 1974; Cox, 1971; Glaser, 1973; and Popham and Husek, 1969) do agree, however, that traditional item analysis information (difficulty and discrimination indices) and test characteristics (reliability and validity) have different meanings in CRT than they do in NRT. That is, decisions on the value of a given item or the worth of a given test would be different in the two applications.

For example, Cox and Glaser both note that NRT items must discriminate between individuals on a single test. Therefore, items with difficulty levels of 1.00 or discrimination indices of 0.00 are useless in a norm-referenced test. A criterion-referenced test, however, is designed to be "generally difficult for those taking it before training and generally easy after training" (Glaser, 1963). Therefore, items that are useless in a norm-referenced test could be retained in a criterion-referenced test if they are generally answered correctly after training but generally answered incorrectly before, i.e., if they provide pretest/posttest discrimination.

<u>The Dichotomous Outcomes Model</u>. The ideal CRT is one which yields a single, unambiguous answer to the question: "does the learner possess the skill being tested?" This ideal is described by Adams (1974) as the "Dichotomous Outcomes Model" (DOM). In this model, a learner may be either in the mastery state or the non-mastery state, exclusively. On an ideal, valid test item, all learners in the mastery state will always give correct responses, and all

learners in the non-mastery state will always give incorrect responses.

The DOM implicitly demands 100% correct performance, but this goal is unattainable in an imperfect world with imperfect measuring instruments. Meskauskas (1976) states that "considerations of measurement error essentially always result in the adoption of standards that demand less than the model seeks". Adams acknowledges this limitation by remarking that an "error of testing occurs whenever learner performance on an item does not reflect his true competence in the trait in question".

Thus, Adams points out that two types of errors can occur. One type of error occurs when a learner who is in the mastery state gives an incorrect response on a valid item. The other type of error occurs when a learner who is in the non-mastery state gives a correct response on a valid item. The goal of the test designer is to minimize the probabilities of these errors by requiring learners to respond to a large enough number of test items to assure reliability, yet to maximize the cost effectiveness of the testing procedure by keeping the number of items as small as possible.

<u>Domain-referenced testing</u>. An important sibling field to CRT is Domain-Referenced Testing (DRT). Hively (1974a) differentiates the two as follows:

> The world of psychometrics may seem as a contrast between Domain-Referenced Testing and Norm-Referenced Testing. The distinction is essentially the same as the one Robert Glaser made between Norm-Referenced Testing and Criterion-Referenced Testing. But the term "criterion" lends itself to misinterpretation. It carries surplus associations to mastery learning that are best avoided by using the more general term "domain" instead. Most people who talk about Criterion-Referenced Testing assume that the technology of Domain-Referenced Testing exists, but they often do not fully recognize what that would imply.

Hively further clarifies DRT theory with the diagram shown in Figure 1.

[This author feels that the distinction between CRT and DRT is most important when working with the cognitive and affective domains, where the universe of target behaviors can indeed be abstract and infinite. In the psychomotor domain, the universe of target behaviors can usually be more clearly defined and approach a concrete domain, thereby minimizing the distinction between CRT and DRT for these behaviors. The problem, therefore, might be seen as the <u>precision</u> with which the target behaviors can be stated.]

Hively (1974a) and Baker (1974) both emphasize the importance of transfer in constructing items for inclusion in a test domain. The goal of the DRT constructor, according to Hively, is "to create an extensive pool of items that represents, in miniature, the basic characteristics of some important part of the original universe of knowledge... The basic notions that guide this activity are those of general-

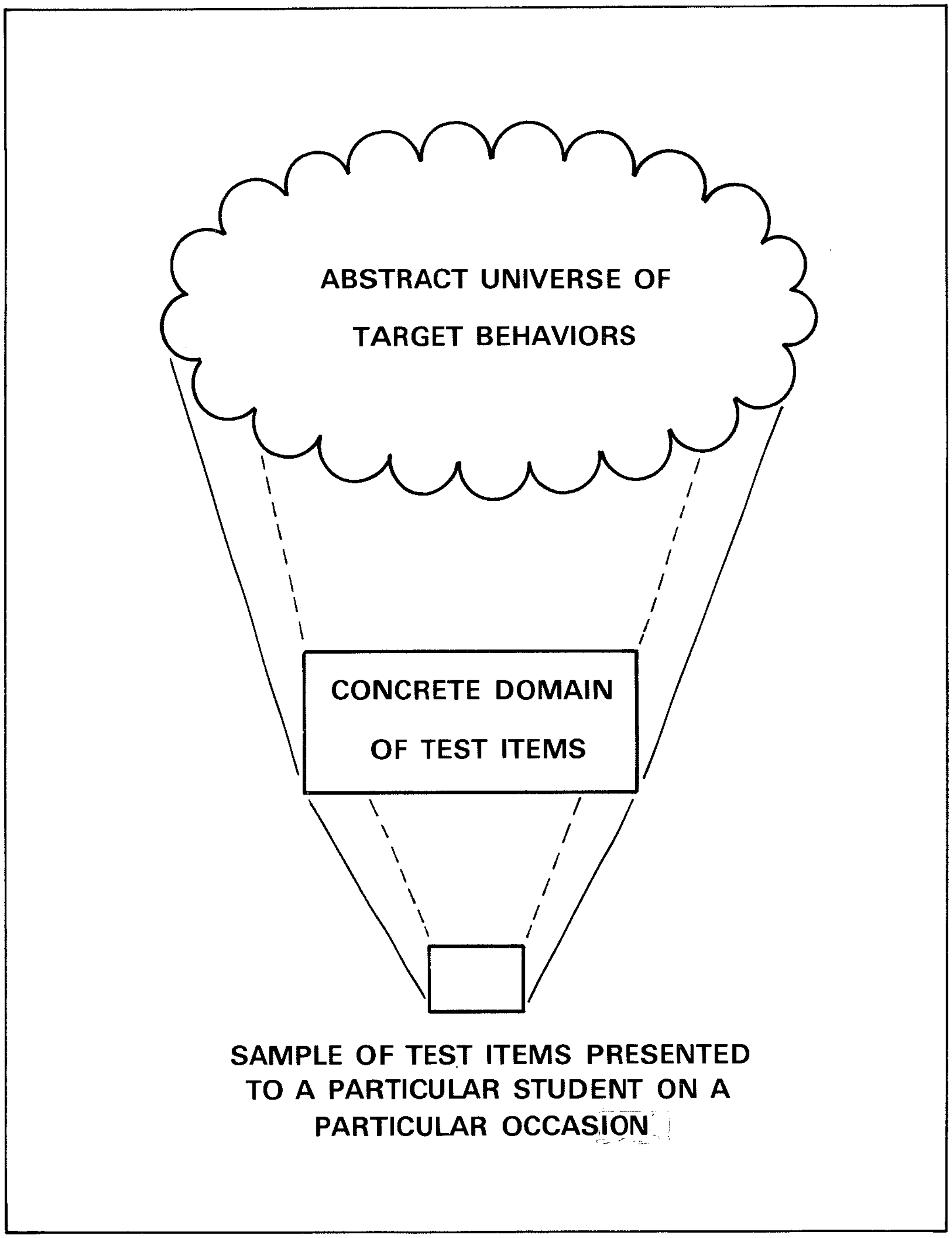


Figure 1

HIVELY'S DOMAIN-REFERENCED TESTING MODEL

(after Hively, 1974b)

<u>ization</u>, <u>transfer</u>, and <u>subject matter structure</u>" [emphasis in original].

<u>Decision analysis models</u>. No matter how one approaches the task of criterion-referenced testing, the basic goal remains the same: to devise a scheme that will classify learners as either masters or non-masters on a set of behaviorally-defined objectives. All such classification schemes are subject to the two types of errors described during the discussion of Adams' work. Ferguson and Novick (1973) have also discussed these errors in similar terms:

> A Type I error occurs when an examinee has sufficient proficiency in a skill but the outcome of the testing suggests that he does not. A Type II error occurs whenever the examinee, in fact, lacks proficiency in a skill but on the basis of test results is said to have sufficient proficiency.

For the purposes of this literature review, a decision analysis CRT model is defined as any scheme that attempts to identify the <u>rates</u> at which these errors occur and, preferably, one that also suggests a method for limiting their occurrence.

Millman (1974) proposes that allowance be made for the error of testing by computing an "Uncertainty Band" to help interpret test scores as follows:

$$UB = 2\sqrt{\frac{N - n}{N - 1}\left(\frac{P_0(1-P_0)}{n}\right)}$$

where UB is the size of the raw score uncertainty band,
N is the number of items in the domain,
n is the number of items in the test, and
P_0 is the passing standard in percent correct.

It is interesting to note that as the number of items in the domain (N) approaches infinity, the term (N-n)/(N-1) approaches 1, and the above equation simplifies to:

$$UB = 2\sqrt{\frac{P_0(1-P_0)}{n}}$$

Millman claims that "when scores fall outside of the Uncertainty Band, correct decisions are made [on the learner's mastery state] over 95% of the time".

Emrick (1971) approaches the problem from the other side, i.e., given the error probabilities and test length, what is the optimal passing standard? This model includes a factor called the "Ratio of Regret", which is computed by summing quantitative expressions of the Bayes risks associated with each of the two types of decision errors. Emrick's formula is:

$$K = \frac{\log\frac{a}{1-b} + \frac{1}{n} \times \log(RR)}{\log\frac{ab}{(1-a)(1-b)}}$$

where K is the passing standard in percent correct,
a is the probability that a Type I error will occur,

b is the probability that a Type II error will occur,
RR is the Ratio of Regret of Type I errors to Type II errors, and
n is the test length in number of items.

Meskauskas (1976) concludes that:

> Emrick's model ... seems worthwhile to pursue. However, empirical quantification of the variables is likely to be a difficult and time-consuming matter.

Ferguson (1971) has developed a Bayesian decision analysis model for computing two criterion scores, P_0 and P_1, each of which is a percentage of correct responses. A learner is said to have "sufficient proficiency" (mastery) on the skill being tested if his or her score is greater than P_0, and "insufficient proficiency" (non-mastery) if the score is less than P_1. The area between P_0 and P_1 is similar to Millman's Uncertainty Band. The probabilities of Type I and Type II errors in this model are respectively expressed as a and b as in Emrick's model. This model is based on the principles of a sequential probability test ratio (Wald, 1947).

The beauty of Ferguson's model is that it allows the test administrator or developer to assign values to P_0, P_1, a, and b to determine the learner's proficiency level to any desired degree of accuracy. This is done as follows. After each test item is administered, the student's score, S, is computed using the formula:

$$S = c \times \log\frac{P_1}{P_0} + w \times \log\frac{1-P_1}{1-P_0}$$

where c is the number of items answered correctly, and
w is the number answered incorrectly.

The learner is said to have "sufficient proficiency" if

$$S \leq \log\frac{b}{1-a}$$

and "insufficient proficiency" if

$$S \geq \log\frac{1-b}{a}$$

If neither of the above inequalities is true, that is, if

$$\log\frac{b}{1-a} < S < \log\frac{1-b}{a}$$

another test item is presented.

As an example of Ferguson's scheme, consider an exam with the following parameters:

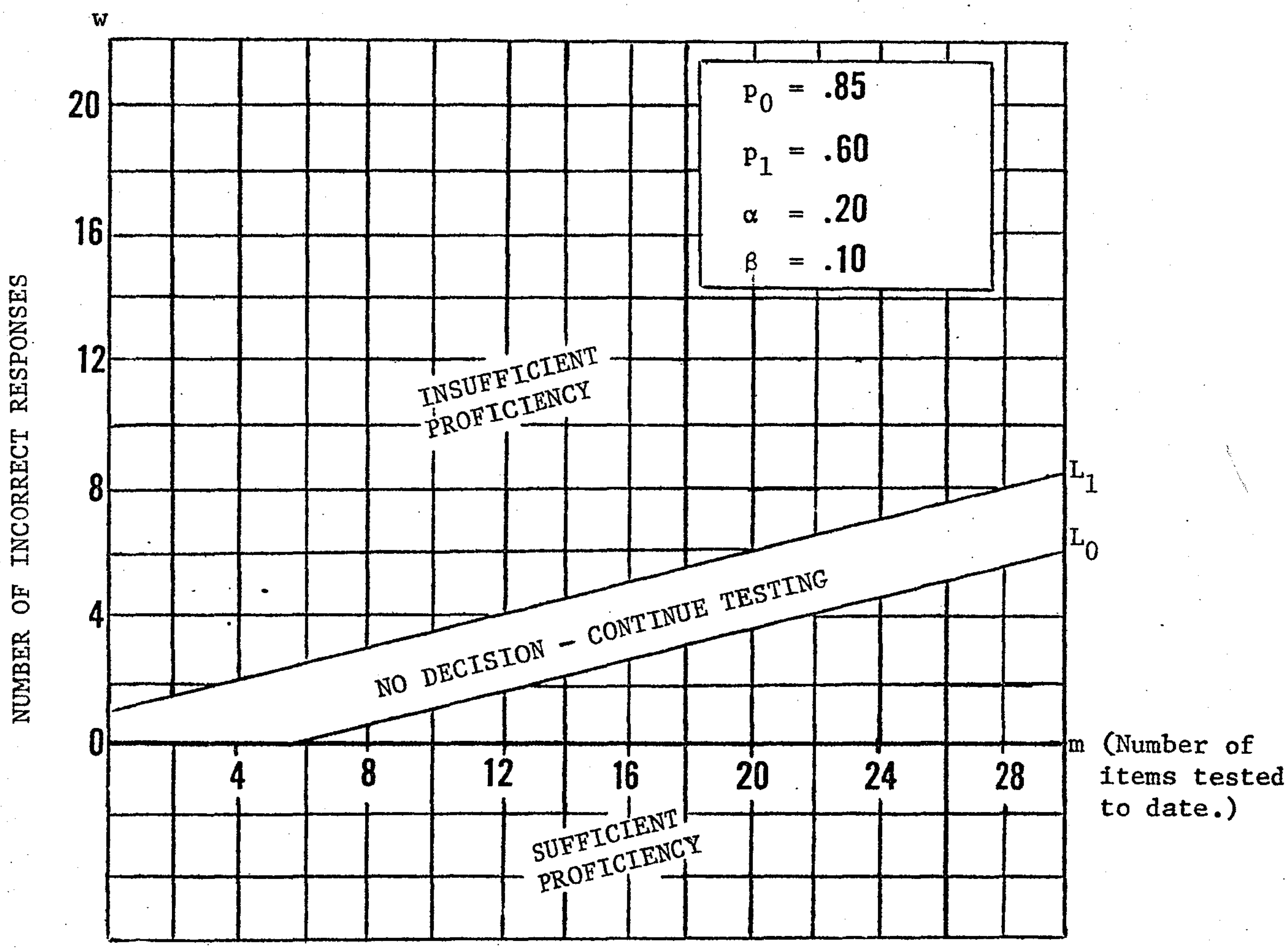
$P_0 = 0.85$

$P_1 = 0.60$

$a = 0.20$

$b = 0.10$

With these values, the graph in Figure 2 can be constructed to illustrate how a learner's test results would be used in determining proficiency. Note that the learner's proficiency state cannot be classified after just one response is made due to the position of the "Uncertainty Band" for the



H_0: p=.85 (Student has sufficient proficiency, omit instruction)

H_1: p=.60 (Student does not have sufficient proficiency, give instruction)

Figure 2

FERGUSON'S METHOD FOR DETERMINING PROFICIENCY ON A CRITERION-REFERENCED TEST

(Ferguson, 1971)

values of P_0, P_1, $\underline{a}$, and $\underline{b}$ chosen. At least two items must be presented and answered incorrectly for a learner to be classified as possessing insufficient proficiency, and at least six must be presented and answered correctly for the opposite classification to be made. By changing the values of these four parameters, the position of the "Uncertainty Band" can be altered. The implementation of these algorithms by Ferguson and other researchers in interactive CAT systems is discussed in a later section. [These algorithms form the basis for the mastery decision model used by the CMI system created for this dissertation.]

CRT reliability. Criterion-referenced tests present special problems in assessing reliability that can make the application of norm-referenced reliability estimates inappropriate. For example, if instruction does indeed take place, a criterion-referenced pretest (administered before instruction) should yield a low mean with a positively skewed distribution, while a criterion-referenced posttest (administered after instruction) should yield a high mean with a negatively skewed distribution (Brown and Pugh, 1976).

Livingston (1972a, 1976) has proposed an adaption to classical norm-referenced reliability coefficients that uses the criterion score as a reference point rather than the mean. His formula for this conversion is:

$$K = \frac{Rs^2 + (m-c)^2}{s^2 + (m-c)^2}$$

where K is the criterion-referenced reliability,
R is the norm-referenced reliability,
s is the standard deviation,
m is the mean, and
c is the criterion (or "cutting") score.

Zieky (1974) interprets this formula as follows:

> ...as the norm-referenced reliability goes up, the criterion-referenced reliability goes up. If the mean score equals the cutting score, then the criterion-referenced reliability equals the norm-referenced reliability. The farther away the mean from the criterion score, the greater the criterion-referenced reliability. Note especially that when all students score exactly at the criterion level, the equation reduces to 0/0 and the [criterion-referenced] reliability is undefined. But if all students get identical scores and that score is <u>not</u> [emphasis in original] equal to the criterion score, the criterion-referenced reliability is 1.

A considerable amount of controversy surrounds the applicability of this formula (Harris, 1972; Livingston, 1972b).

Livingston's formula is a conversion of a correlation coefficient. Haynes and Walker (1976) have looked at another correlation coefficient -- that between test scores and teacher judgement. [Their approach more closely resembles a test of concurrent validity than reliability, but serves well to illustrate a coefficient of classification.] Teachers classified 500 students as either masters or non-masters in two subject areas, based on the objectives of the 1973-74 Florida Statewide Assessment program. When tested on these objectives using 2-5 items per objective, misclassifications occurred at a rate of over 30%, even when an "optimal mastery score" (that score yielding the fewest number of misclassifications) was derived by a posteriori analysis and used as the cutting score.

The concept of criterion-referenced reliability as a measure of the consistency of mastery and non-mastery classifications is one which has received considerable support (Carver, 1970; Hambleton and Novick, 1972; Hansen et al., 1977; Livingston, 1976; Subkovniak, 1976, 1978; Curlette, 1977). Such measures require two sets of test data. The frequencies of agreement between the classification decisions made by both sets of test data may then be represented in a 2x2 table as shown in Figure 3. In this table, A is the number of students who were classified as masters on both T_1 and T_2, and D is the number who were classified as non-masters on both tests. As these frequencies increase, the more the two sets of data agree and the higher the reliability of classification. Conversely, B and C are the disagreement frequencies, and as they increase the reliability of classification decreases.

Carver (1970) points out that reliability of classification does not depend on score variability, and is

		CLASSIFICATION ON T_1	
		Master	Non-Master
CLASSIFICATION ON T_2	Master	A	C
	Non-Master	B	D

Figure 3

FREQUENCIES OF AGREEMENT BETWEEN MASTERY AND NON-MASTERY CLASSIFICATIONS ON TWO SETS OF TEST DATA

therefore useful in assessing the reliability of criterion-referenced tests. The simplest expression of a reliability coefficient based on this concept is the percentage of cases in which both sets of data agree, namely:

$$P_0 = \frac{A+D}{A+B+C+D}$$

This measure varies between 0 and 1 and is referred to as the "percentage of agreement".

Swaminathan et al. (1974) prefer using a refinement of the percentage of agreement known as the kappa coefficient. This expression attempts to correct the percentage of agreement for chance. The computation is:

$$\text{kappa} = \frac{P_0 - P_c}{1 - P_c}$$

where P_0 is the percentage of agreement, and

P_c is $\frac{(A+C)(A+B)+(B+D)(C+D)}{(A+B+C+D)^2}$

Swezey and Pearlstein (1975) prefer a slightly more sophisticated expression called the phi coefficient. This coefficient is really the correlation of two sets of test data using 0 as the non-mastery score and 1 as the mastery score. The computation is:

$$\text{phi} = \frac{AD - BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}}$$

Swezey and Pearlstein suggest that phi $\geq$ 0.5 represents "sufficient reliability", while phi < 0.5 represents "insufficient reliability". Note that if $\underline{B} = \underline{C}$, kappa = phi.

Livingston (1976) analyzed these computations and suggested yet a fourth coefficient. His purpose was to modify the simple percentage of agreement, P_0, so that it varies between -1 and +1 (like the kappa and phi coefficients) and to show that this new coefficient, the G index, more logically reflects the reliability of classification. The computation is:

$$G = 2 \times (P_0 - 0.5)$$

Two examples from Livingston's work suffice to make his point. Consider the data in Figure 4. Livingston argues that the data in Case 1 clearly show that, in most cases, T_1 and T_2 do not agree. Yet the kappa and phi coefficients for these data are +0.12 and +0.25, respectively, which are small but definitely positive. The corresponding G index for the data in Case 1 is -0.20, which, Livingston argues, more accurately indicates the disagreement because it is negative.

Case 1

		T_1	
		M	N-M
T_2	M	20	60
	N-M	0	20

Case 2

		T_1	
		M	N-M
T_2	M	90	5
	N-M	5	0

Figure 4

SAMPLE CLASSIFICATION FREQUENCIES

(after Livingston, 1976)

The data in Case 2 are even more striking: T_1 and T_2 agree in 90% of the testing cases, yet the kappa and phi coefficients are both -0.05. The corresponding G index is 0.80. Here again, Livingston argues, the G index more accurately reflects the correlation of classification because it is positive.

Subkoviak (1978) found that for all four reliability computations, reliability estimates stabilize as test length increases. [All four of these indices will be reported in this dissertation to assess the reliability of classification.]

Computer-Assisted Testing

Computer-assisted testing (CAT) is one of the fastest growing applications of instructional computing. Constructing tests by computer is a relatively straightforward process and can be shown to be cost-effective (Ansfield, 1973; Menne and Lustgraaf, 1974; Prosser, 1975). Lippey (1973) enumerates the major benefits of CAT as follows:

(1) reduces clerical chores required of an instructor,

(2) provides error-free text,

(3) allows the educator to concentrate on content rather than the mechanical aspects of test construction,

(4) eliminates the problems that can arise from students gaining access to the test items

before the formal test (if the item bank is sufficiently large), and

(5) allows input from many users through centralized collection of item statistics, thus improving the quality of the items through experience.

A large variety of CAT systems are currently in use, from those that store only item characteristics (ETS, 1974) to those that construct and administer tests through interactive terminals (Ferguson, 1971). The CAT systems discussed in this section are therefore divided into four levels. Systems at the first three levels generally employ batch processing and include, respectively, systems that store and print teacher-constructed exams, those that automatically construct exams from a given item bank, and those that employ an algorithmic approach to item construction. The fourth level is characterized by interactive systems that make use of branching tests to control the sequence in which items are presented to the student.

Test printing systems. The simplest type of CAT system is one which does the job of a secretary by printing test questions selected by an instructor (Remondini, 1973). The items to be printed may be stored in any machine-readable format, e.g., magnetic tape, disk, or punched cards. In Remondini's system, the computer produces a single copy of the test. This is photocopied and transferred onto ditto masters for duplication. The answer sheets are corrected by a mark-sense device and the computer is used to compute and print an item analysis and punch statistical data for each item onto standard computer cards.

Salisnjack (1973) uses a system almost identical to Remondini's. He claims that it only takes 25 minutes to prepare two forms of a 75-item multiple choice test with the aid of the computer. Salisnjack finds that his CAT system controls the cost of test construction, solves the problem of cheating, and reduces the "edge" provided by fraternity test files. He comments, however, that "attempts at making the complete data bank available to all students as a study guide so far have been unsuccessful -- the cost of providing individual copies is too high, and the copies placed in the library tend to disappear".

MENTREX Enterprises in Los Angeles offers commercial test construction services similar to those offered by the systems of Remondini and Salisnjack (Libaw, 1973). Users request tests through the mail by selecting questions from a "catalog" supplied by the company. The system can produce several forms of the same test by "scrambling" the items or select items for the test based on "keys" specified by the user. Test masters are returned ready for duplication, accompanied by an answer key and machine-readable answer sheets. Answer sheets are later returned to MENTREX for item analysis.

Educational Testing Service (ETS, 1974) is a unique user of CAT due to the sheer size of their operation. They have stated that the following two developments are necessary before they can implement large scale CAT use, and they do not yet see these as part of the current state of the CAT art:

(1) "the development of detailed item classification systems", and

(2) "the delineation of the professional judgements made in building a test from a group of items with detailed content, ability, and statistical specifications in terms precise enough to be translated into computer programs".

ETS currently uses a CAT system to help select items from their huge data banks. The system does not print tests, but simply returns item numbers that fit specified characteristics. These characteristics include:

(1) the item ID number,
(2) its classification,
(3) a history of its use,
(4) up to five sets of statistics,
(5) codes for security level and current activity, and
(6) twelve 15-character keywords.

It is interesting to note that ETS sees the demand for large national selection tests as diminishing. They feel that interactive testing is required for the future, with tests for guidance, placement, and evaluation. Their 1974 paper states that the technology for such systems exists, but that development funds are needed to make them cost-effective.

Test construction systems using item banks. The second level of CAT is characterized by systems that construct tests from stored item banks. In addition to the benefits noted earlier, these systems provide a means for generating multiple forms of the same test. Jensen (1973) has used such a system to generate 4000 different forms for a class of 1500 students. He achieves a form of criterion-referencing by allowing students to take a test on a specific topic as often as they like and counts only the highest grade. His philosophy in this approach is that "...one should ask only what one wishes the student to know, but ask it in so many different ways that the student cannot learn the items without learning the concept".

Prosser (1973) describes a similar test construction system but includes some figures on its cost. This system selects items from predefined "groups" that are specified by the user. To produce 1000 3-page tests, the system requires 20 seconds of CPU time and 3 hours of printer time, making the cost of each form about five cents.

The Classroom Teacher Support System (CTSS) was designed by IBM for the Los Angeles Unified School District (Toggenburger, 1973). This system constructs multiple choice exams according to teacher-specified criteria such as course, category, difficulty level, behavioral level, and keywords. The system can also work with "macro" items, i.e., passages followed by two to nine related questions.

Toggenburger reports that CTSS currently uses an American History item bank of 8000 items that were written by 20 teachers over the period of one summer.

Ansfield (1973) has developed a system similar to CTSS called the Automatic Examination Generator (AEG). Ansfield's report on the AEG, however, also includes data on its cost: the total computer expense for producing the masters for four versions of a 70-item objective test with answer keys is $1.75.

One last item banking system with a somewhat unique character has been developed by Cohen and Cohen (1973). The main purpose of this CAT system is to assure no overlap in the items presented on successive administrations of a test for any one student. Cohen and Cohen have developed two versions of this system, one for batch processing in COBOL, and one for interactive processing in FORTRAN.

Algorithmic approaches to item construction. Olympia (1975) contends that standard item banking has three disadvantages:

> (1) it lacks repeatability [unless the item bank is extremely large], especially when a given item always appears on a test exactly as it is stored,
>
> (2) it requires a large amount of contruction time and storage to create a usable bank, and
>
> (3) it discourages the sharing of one program by various disciplines. [This author feels that the CAT systems discussed in the previous two

sections demonstrate capabilities which clearly contradict this objection.]

To overcome these drawbacks, Olympia devised a system for storing examination items in three "pools": a "keyphrase" pool, a "statementphrase" pool, and a "distractor" pool. The system constructs an item by joining one member of the keyphrase pool with one member of the statementphrase pool and then selects a list of possible answers (including the correct answer) from the distractor pool. As an example, the three pools for constructing items dealing with electron configurations are shown in Table 1.

Denney (1973) describes a system similar to Olympia's. This system stores multiple choice items as a stem

Table 1

POOLS FOR CONSTRUCTING ITEMS ON ELECTRON CONFIGURATIONS

(after Olympia, 1975)

Keyphrase Pool	Statementphrase Pool	Distractor Pool
Chlorine Oxygen Hydrogen Magnesium Helium	has how many valence electrons? has how many L-shell electrons? needs how many more electrons to have an inert gas structure?	0 1 2 3 4 5 6 7

with up to seven distractors. With this data, the system can construct 245 different questions consisting of a correct choice and four alternatives. If the order of the five alternatives is randomized, up to 29,400 different variations of the same question can be generated.

Heines (1974) created an interactive CAT system that generates data to complete item forms or selects one of four previously defined item variations. Regardless of the item generation scheme, the system assures that no student is presented with the same item on two successive administrations of the test. This system is also interesting in that:

(1) all instructions for operating the system are presented to students via an audio tape that resides beside the terminal,

(2) many items are complemented by diagrams and pictures presented by a 35 mm slide projector under student control, and

(3) the system provides an interactive environment for the instructor as well as the student, so that instructors may perform data analysis without any special knowledge of the system.

Interactive, branching tests. O'Reilly, Gorth, and Pinsky (1973) comment on the current state of the CAT art as follows:

> [Current CAT efforts] tend to be largely superficial, poorly grounded in relevant evaluation models and test theory and tend to continue a questionable school and classroom practice... [They] focus on the mechanism of test production via machine, a tendency which works against the need to maintain a precise relationship between

> the intent of instruction and the measurement process.

The CAT systems described thus far definitely reflect the type of system that O'Reilly et al. condemn. Those that follow go beyond "the mechanics of test production" to explore new evaluation techniques that employ the computer as an integral tool.

Ferguson (1971) defines a branching test as "any instrument designed to measure a set of skills or objectives by routing the examinee to items neither too easy nor too difficult for him to solve". One example of this technique was developed by Hansen (1969) and is shown in Figure 5. In this scheme, first Item 1 is presented to the student. The student is then branched to Item 2 if Item 1 is answered correctly, and Item 6 if it is answered incorrectly. Item 1 is designed to have a difficulty index of 0.50, and each successive item is designed to have a difficulty differential of 0.10 from the preceding item. Thus, the most difficult item in the tree (Item 4) will have a difficulty index of 0.80, and the easiest item (Item 13) will have an index of 0.20. Hansen found that this scheme is significantly more reliable than traditional tests and, moreover, is effective at reducing test anxiety.

The criterion-referenced model for Ferguson's work (1971) has been discussed in an earlier section (cf. pages 18-21). By comparing Ferguson's work to that of the other

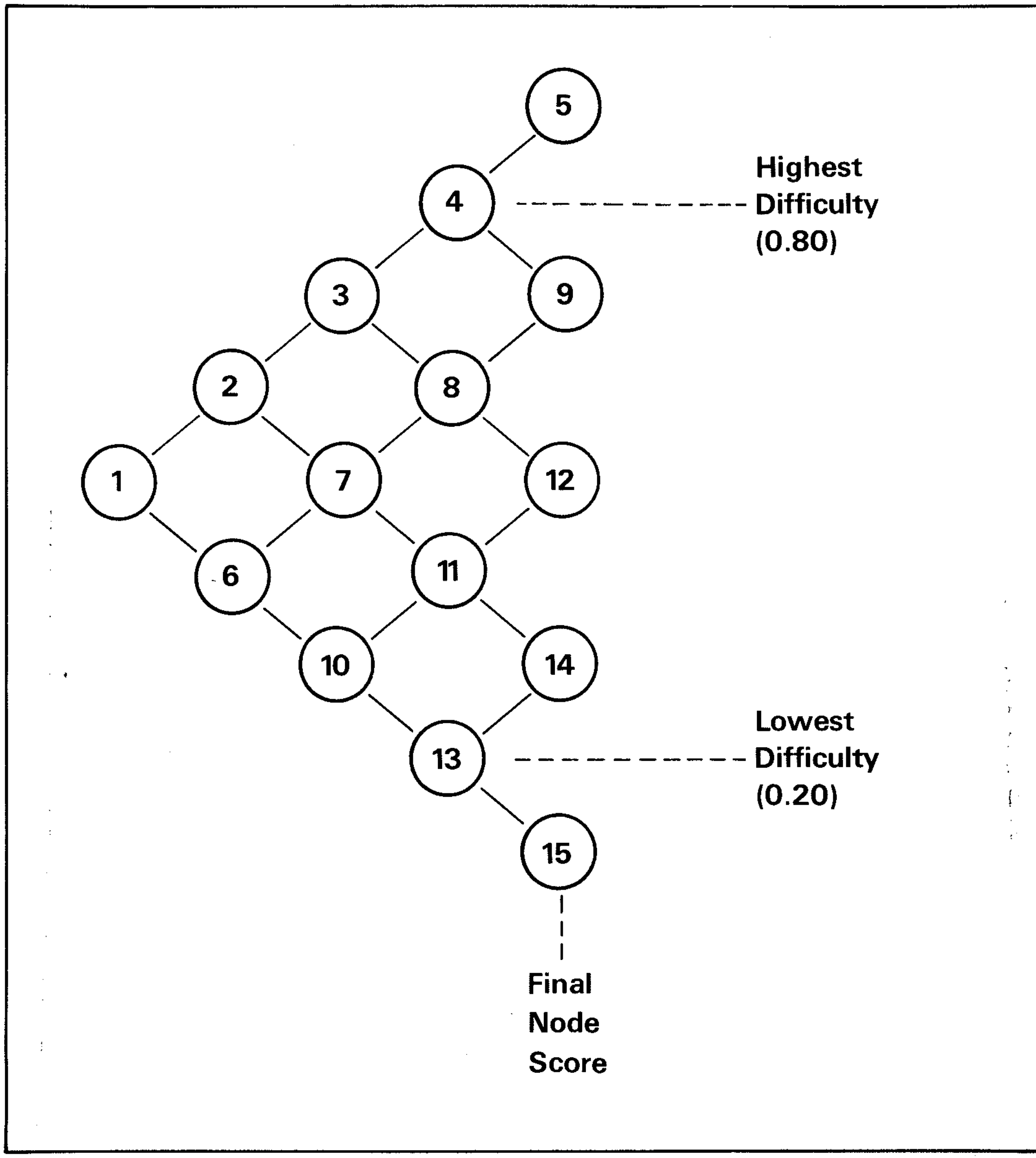


Figure 5

HANSEN'S SEQUENTIAL ITEM TREE NETWORK

(after Hansen, 1969)

CAT researchers discussed so far, it can be seen that Ferguson is one of the few researchers to have created a CAT system as a means for implementing a well-developed theory of evaluation. This system tests objectives in the Individually Prescribed Instruction (IPI) mathematics curriculum, a program that already makes use of comprehensive paper-and-pencil testing and therefore provides a useful measure of the system's success.

Ferguson's CAT system interactively administers a series of items on a specific objective and applies the sequential probability test ratio to the mastery decision process. Students are then branched to test items on either more advanced or more preliminary objectives based on their mastery classification. By this process, Ferguson is able to pinpoint a student's competency level with any desired degree of accuracy and then prescribe instruction to fit the student's needs. Ferguson found that his branching system yields classification decisions that are "consistent with subsequent paper-and-pencil test outcomes approximately 99% of the time". He conjectures that "by employing a sampling technique that permits control over classification errors, the CAT model may increase reliability".

Ferguson discussed three "suggested refinements" to his model. First, he feels that testing must be representative and that this is not always guaranteed by random sampling. He therefore recommends a combination of randomly constructed items with domain-referenced item forms. Second, Ferguson feels that research is needed to achieve a compromise between minimizing Type II (false positive) errors, which he considers the most serious, and reducing the number of items presented for cost-effectiveness. [This refinement is the third subproblem of this dissertation.] Third, he notes that in his model all examinees start at the same point, and therefore highly competent examinees do problems that are too easy while incompetent examinees do ones that are too hard. He suggests that examinees might be allowed to choose their own starting points. Ferguson's paper concludes:

> By tailoring the test to individuals, fewer objectives need to be tested <u>and</u> [emphasis in original] the objectives that are tested are less subject to errors of proficiency classification.

Several researchers have conducted studies to ascertain the time savings and affects on reliability caused by the use of a sequential probability test ratio. Waters (1975) found that this technique yielded item reductions of 50% with time reductions of 40%.

Hansen et al. (1977) performed a study on variable length tests using a "within-subject" design [which was very similar to that used by the study conducted for this dissertation]. Students were administered tests via a computer terminal in which the short sequential probability test moved directly into a long (40 item) test without visibility to the student. This yielded two measures on the same test on the same student. Hansen et al. found no significant differences in the reliability estimates for the two tests, but the shorter tests provided a 30% time savings.

The average response time for a four-alternative multiple choice item presented by Hansen's system was 1.6 minutes. This indicates extremely difficult items. They found that the time and item savings possible with sequential testing were directly proportional to the complexity of the course material and items.

Schneider and Fine (1973) employed a sequential probability test ratio on the TICCIT system and cited three considerations for setting the P_0, P_1, $\underline{a}$, and $\underline{b}$ parameters:

(1) How large is the minimum critical subset of items? [For this dissertation, this was set to two items for non-mastery decisions and one item on each objective for mastery decisions.]

(2) How many items will be required? [For this dissertation, this was determined by *a posteriori* analysis.]

(3) What is the probability of getting an item correct by guessing? [For this dissertation, each item was assigned a weight in the scoring algorithm depending upon its type -- true/false, yes/no, multiple choice with four alternatives, or multiple choice with five alternatives.]

Weiss and his co-workers at the University of Minnesota have conducted a large number of studies on the full range of interactive, branching tests, which they refer to as "adaptive" tests. Their purposes are to devise strategies for matching item difficulties with the student's ability and to sequentially estimate that ability (Weiss, 1973). The bulk of their research has been on selecting items from a precalibrated pool, given the premise that "an individual's ability level will be most accurately estimated when the items administered are as close to his/her ability level as possible". The computer is used to select the next item to be presented based on the student's responses to previous items.

Weiss discussed and compared seven item selection procedures in his 1973 paper. The comparisons assumed that:

(1) a pool of items existed that were precalibrated on item difficulty and grouped into various difficulty "levels",

(2) all items in the pool had equivalent discriminating power, and

(3) the pool consisted of free-response items that would not be answered simply by guessing.

Weiss points out that these assumptions are not without fault. In fact, he notes the following limitations:

(1) random quessing by real subjects introduces inaccuracies,

(2) items with high and low degrees of difficulty cannot have discriminating powers equal to those with medium difficulty, and

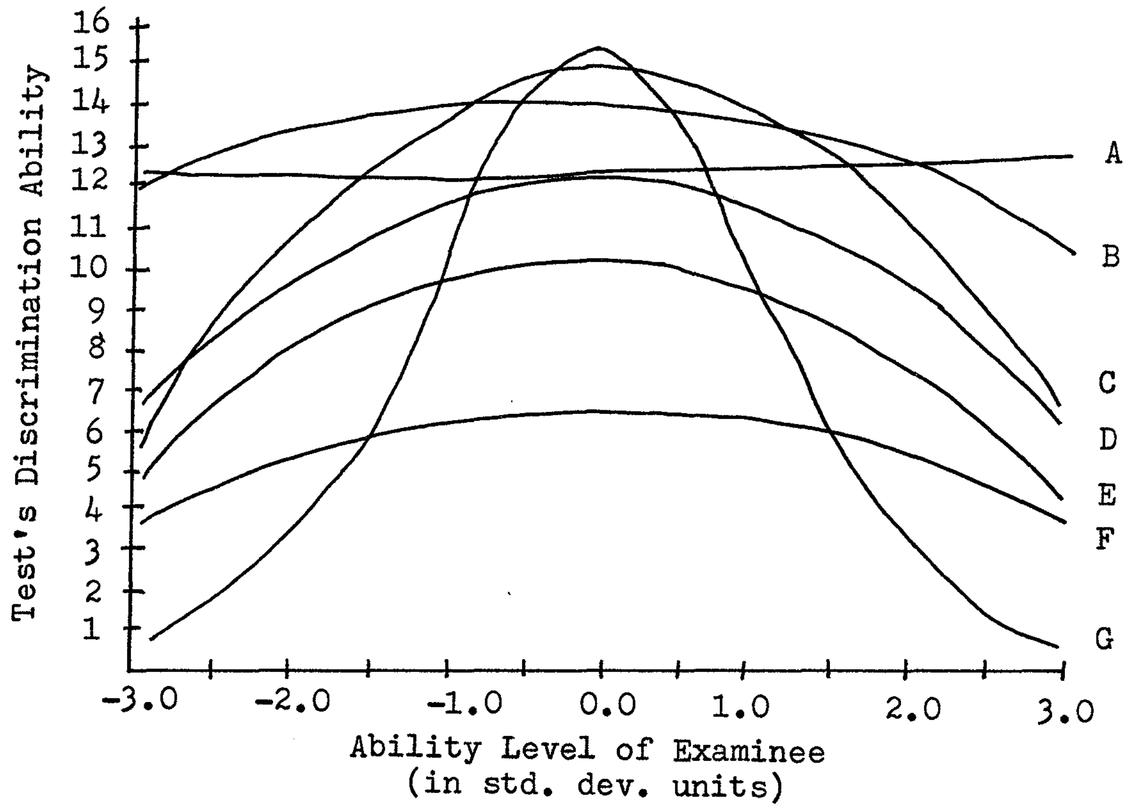
(3) various item selection procedures introduce biases due to differences between individual ability levels and the average ability estimate for all students (used as a starting point in the branching procedure).

The seven item selection procedures discussed by Weiss are summarized in the paragraphs that follow and compared on discrimination ability in Figure 6. Discrimination ability is expressed in arbitrary units, but is considered best when it has a higher value and is constant over a wide range of examinees' ability levels (expressed as standard deviations from the mean).

(1) Rectangular Conventional. A fixed set of items with a wide range of difficulties is presented to all examinees. The discrimination ability of the test as a whole is low, but rather constant over a broad range of ability levels.

(2) Peaked Conventional. Items are selected whose difficulties are very near the center of the ability range. This yields precise measurements for examinees in the middle ability ranges, but very poor measurements for examinees in the extreme ranges. The peaked conventional strategy is "better" than the rectangular conventional strategy in the range of ±1.5 standard deviations, but "worse" outside this range.

(3) Two-Stage. First, a very short test is administered to establish the examinee's general ability level. Second, a longer test is administered that is targeted for that specific ability to pinpoint the examinee's level. This strategy always yields better discrimination results



Curve Legend:
A Stradaptive
B Bayesian
C Pyramidal
D Flexilevel
E Two-Stage
F Rectangular Conventional
G Peaked Conventional

Figure 6

STRATEGIES OF ADAPTIVE TESTING

(after Weiss, 1973)

than the rectangular conventional test and is better than the peaked conventional test except at the very center.

(4) Flexilevel. The first item presented is selected from those in the median difficulty level. Based on the examinee's response to this item, the next item is selected from either the next higher (if response was correct) or next lower (if incorrect) unadministered difficulty level. No test contains two items from the same difficulty level. The intent of this strategy is to avoid misrouting by combining both of the two-stage tests.

(5) Pyramidal. This strategy is similar to flexilevel tests, but a difficulty level may be repeated. It is the same strategy as Hansen's Sequential Item Tree Network (Figure 5). This strategy essentially carries multiple routing to its logical extreme, with one item per stage. Pyramidal tests yield the best overall discrimination ability of any of the preceding four strategies, but this technique still yields better test discrimination with medium ability examinees than with those of high or low ability.

(6) Stradaptive (stratified-adaptive). Items are grouped into "strata" consisting of items with only slight differences in difficulty. Prior knowledge is used to estimate the examinee's ability level, and the first item presented is selected from the stratum at that level. This technique allows multiple entry points. Branching then occurs between adjacent strata, moving to more difficult

items when the examinee responds correctly and easier items when he/she responds incorrectly. The outstanding characteristic of this strategy is that it yields very consistent test discrimination results across all examinee ability levels. In addition, these results are consistently high and surpassed by the peaked conventional and pyramidal tests only in the very center of the ability range.

(7) Bayesian. This strategy is similar to stradaptive tests, but branching is allowed to occur between non-adjacent strata. After each item is administered, the examinee's most probable ability level is computed and an item is presented from the stratum at that level. As the test progresses, the estimate of the examinee's most probable ability level changes and the error associated with that estimate decreases. This is the convergence nature of the Bayesian process. Discrimination abilities of tests using this strategy are somewhat higher than those of stradaptive tests and almost as consistent. They drop off slightly for examinees whose ability levels are more than two standard deviations from the mean.

The implementation of adaptive testing can be difficult because it requires a large pool of precalibrated items and, in some cases, *a priori* estimates of examinee ability that correlate at least .5 with actual ability (Weiss, 1973). Several implementations, however, have been realized. For example, Bejar et al. (1977) used stradaptive tests in a classroom environment and found that these tests not only yielded more precise estimates of achievement than standard tests throughout the entire range, but that they also reduced test length significantly.

Another characteristic of adaptive testing is that it is predominatly norm-referenced. At least one study has been done, however, that attempted to evaluate the use of adaptive testing in a criterion-referenced mode. Vale (1976) applied the Bayesian adaptive testing strategy to the problem of making mastery vs. non-mastery classifications. He found that a conventional test was just as useful as the Bayesian test when there was one cutting score at the mean ability level. With two or more cutting scores, or with the cutting score different from the mean ability level, the adaptive test yielded superior results.

Computer-Managed Instruction

Computer-managed instruction (CMI) goes beyond CAT by providing an analysis of student performances to guide:

(1) the prescription of additional course work,

(2) the improvement of subject matter presentations, and

(3) the revision of the evaluative materials themselves.

Most CMI systems have CAT components, but include more extensive links with the instructional materials themselves.

Techniques for CMI may be broken down into two distinct categories, distinguished by the mode in which the computer is accessed when analyzing data. The following sections define these two modes and discuss sample projects that demonstrate each technique.

Batch CMI. When applied to the instructional process, "batch" computing refers to the use of the computer while neither the student nor the instructor is interacting directly with the computational process. The most common application of this technique involves the use of answer sheets that students mark with a pencil and that are subsequently processed by an optical mark reader. The optical reader may also be coupled to a card punch which prepares standard computer cards for input into data analysis programs. When these cards are processed, extensive test statistics and item analyses may be generated for the instructor and individual student results may be recorded in a master file.

Paul Geisert (1973) extended Butzow's testing technique (discussed under the section on the evaluation of SPI) by using the computer to store students' records. His model employs an open quiz room in which printed tests are distributed by an on-duty assistant, written by the student, and returned. These tests are graded on stated criteria by the specific teaching assistant assigned to the student.

Each student receives a pack of computer cards at the beginning of the semester and submits them for processing as he or she completes the course objectives. These cards are fed directly to any of five packaged computer programs. Thus, this CMI system does not have a CAT component. It uses the computer purely for record keeping. This use is important, however, as Geisert's course has an enrollment of 500 students.

Franke et al. (1972) have demonstrated similar computer usage. Their computer is used for generating examinations (CAT) as well as keeping records. The enrollment in the experimental course is over 3500 students annually, so the role of the computer is even more important than in Geisert's course. This system divides the course into eight "phases", and students may "test out" of any or all phases by means of a written examination. Examinations are generated from a bank of over 3000 questions grouped by phases, topic categories within each phase, and difficulty estimates based on the number of students answering the question correctly. The computer is also used to generate reports documenting each student's achieved scores to date and summaries from master files stored on magnetic tape.

A highly developed system making used of batch CMI techniques has been developed by Kelley (1968, 1972). This system differs from those discussed above in that it produces extremely detailed outputs for students, teaching assistants, and the course professor, and because it generates weekly assignments for each student. Student progress is monitored by "surveys" consisting of ten multiple choice questions each, administered approximately every week. These surveys are not examinations, and Kelley's implementation does not use them for grading purposes (although they may be so used). Rather, results of the surveys are fed to a computer program which prints out individualized assignments contingent upon the decision rules specified by the professor. These decision rules take into account each student's weekly survey score, his or her previous course background, and his or her specific goals (Kelley, 1973). Summary item analysis reports are generated for the teaching assistants and professor to guide class discussions and the revision of course materials. The beauty of Kelly's design lies in its flexibility in applying the decision rules specified by the professor and the clarity of its reports.

Batch CMI has the advantage of being able to handle a virtually unlimited number of students and uses the computer in its most efficient mode. On the other hand, use of the system in this manner generally requires precise specification of program parameters (Kelly, 1973) and at least several hours' wait (Kelly, 1972). An error in parameter specification can easily abort a run and double or triple the time required to receive results.

Interactive CMI. Interactive programming has been used to overcome some of the CMI utilization problems enumerated by Kelly. This technique requires that the user interact directly with the program through a computer terminal while the computational process is taking place. In this mode, program parameters are specified at the request of the computer and mistakes may be immediately detected and corrected. Responses to the users' requests are also immediate, so that they may be led interactively through all steps necessary to generate the reports of their choosing. The only unguided acts that users are required to perform are the connection of the terminal to the computer system and the loading and execution of the desired program. From there on, users are always prompted as to what data is required by the program. Interactive CMI systems are often coupled to interactive computer-assisted instruction (CAI) programs which administer instruction to students under a time-sharing environment.

A complete interactive CMI system for both teachers and students has been developed by Ghesquire (1973) on the PLATO system. This system creates both tabular and graphic displays of data collected during student CAI on a video display screen. Unfortunately, the unique characteristics of the PLATO system make most of Ghesquire's work non-exportable to other systems without major revisions.

In a rare article on CMI in industrial training, Buchanan (1978) reported on another PLATO CMI system being used by United Airlines in training programs for newly-hired pilots. Additional research on interactive CMI has been carried out on the TICCIT system (Schneider and McMurchie, 1973).

Dean (1978) has reported on the use of CMI at IBM and the cost savings that they have realized through this application. The IBM Field Training System administers practice "quizzes" and more formal "examinations". The examinations provide quality control on self-paced training that is delivered at branch offices in the form of self-study guides, audio tapes, and video tapes. The cost of running IBM's CMI system is about $15 per student per day. The cost of running a full CAI system would be about $60 per student per day, while the cost of instruction via conventional stand-up lecture is about $100 per student per day. (All costs are net per student with development factored out.) Due to the large size of their training operation, Dean claims that "compared to a conventional instruction system, the Field Training System helps IBM realize a conservative $10 million annual cost avoidance".

[This dissertation deals with the design, implementation, and evaluation of an interactive CMI system.]

Chapter 3

DESIGN AND IMPLEMENTATION OF THE STUDY

The study was conceived as a two-part project. The first part involved the development of a self-paced course whose objectives were measureable via computer-assisted testing. The second part involved the development of a computer-managed instruction (CMI) system to maintain quality control on the use of the self-paced course. These two components, taken together, formed a single training package.

The study was conducted in two public school systems and one junior college for approximately two months. Data on the CMI system's use was collected throughout this period and analyzed with respect to the dissertation's subproblems.

Description of the Self-Paced Course

The self-paced course to which this study relates is entitled The BASIC Primer (Digital Equipment Corporation, 1979). The course covers the concepts and syntax of BASIC language programming on Digital computer systems.

The BASIC Primer consists of an introductory Student Guide and 15 modules on BASIC language programming. These modules are intended to be studied in accordance with the hierarchy shown in Figure 7. This hierarchy was derived from the prerequisite relationships of the objectives. Modules 1, 2, 4, 6, 8, 9, and 10 form the "core material" for the course, while the other nine modules form the "extended material".

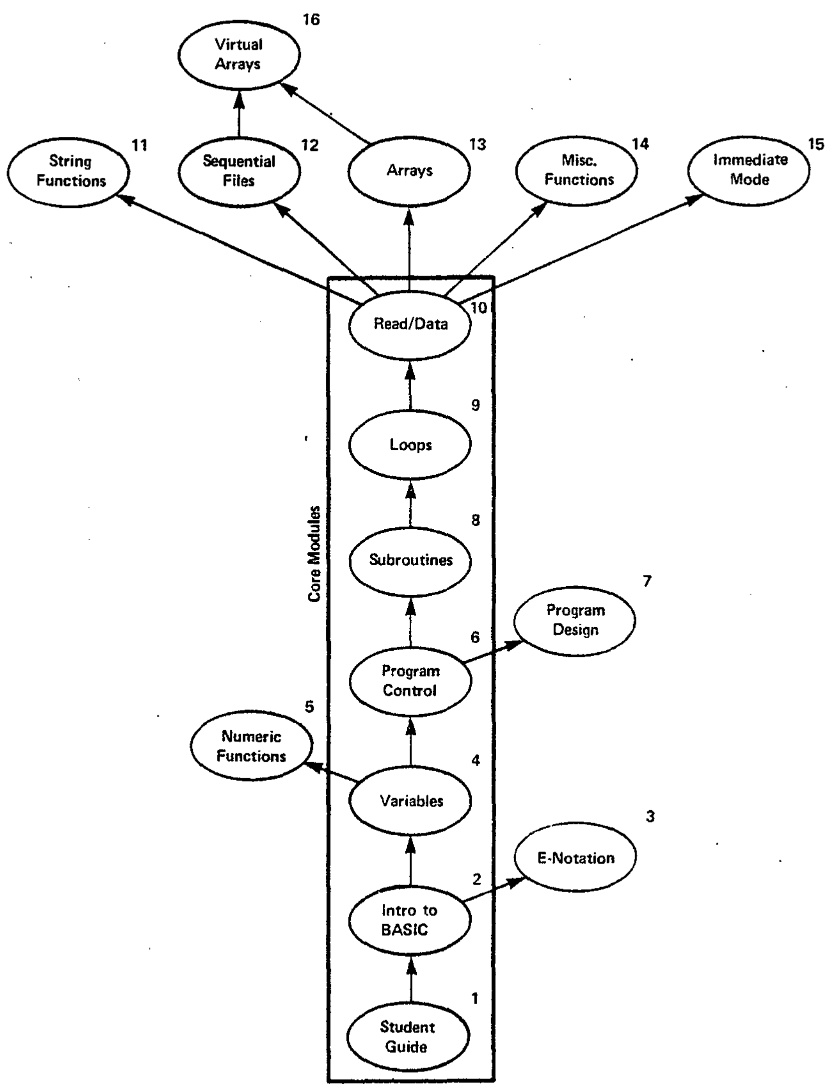


Figure 7

MODULE HIERARCHY FOR <u>THE BASIC PRIMER</u>

Figure 8 shows how students worked through The BASIC Primer. Before studying each module, students were given the opportunity to take a pretest. If they could demonstrate mastery on this test, they were branched to the pretest for the next module in the hierarchy. This loop continued until students came to a test on which they could not demonstrate mastery. At this point, students were instructed to study that module off-line, and to return to the CMI system when they were ready for the posttest.

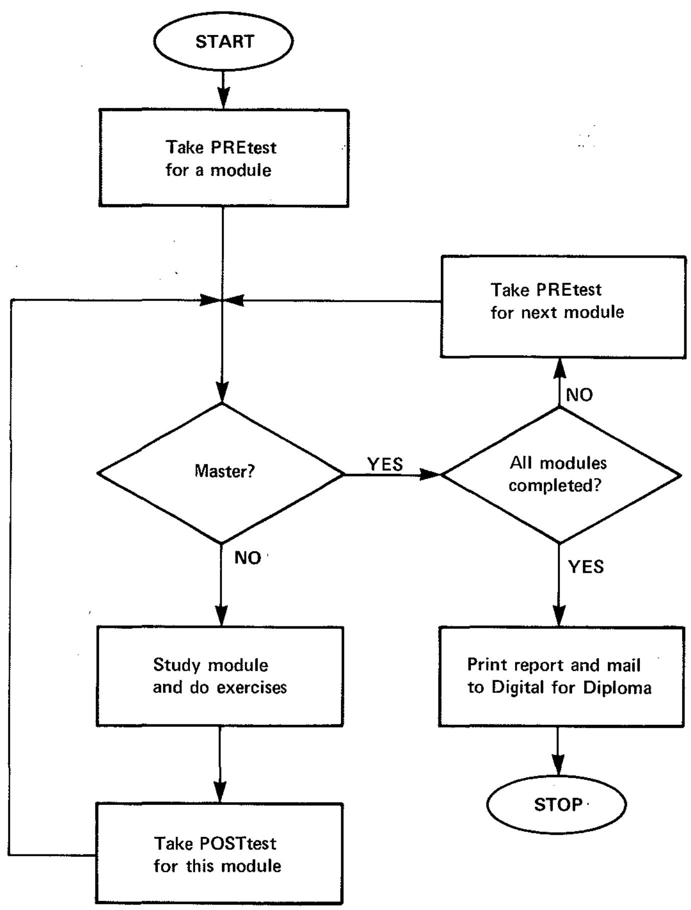


Figure 8

RELATION BETWEEN ON-LINE CMI SYSTEM AND OFF-LINE LEARNING MODULES

<u>Description of the CMI System</u>

The student portion of the computer-managed instruction system consists of three main programs. These are the Registration Program (CMI), the Router Program (ROUTER), and the Computer-Assisted Test Administration Program (CATSTR). Two other subprograms are also included in this system: the New Student Registration Program (REGSTR) and a Feedback Program (FEEDBK). Students moved through these programs as illustrated in Figure 9.

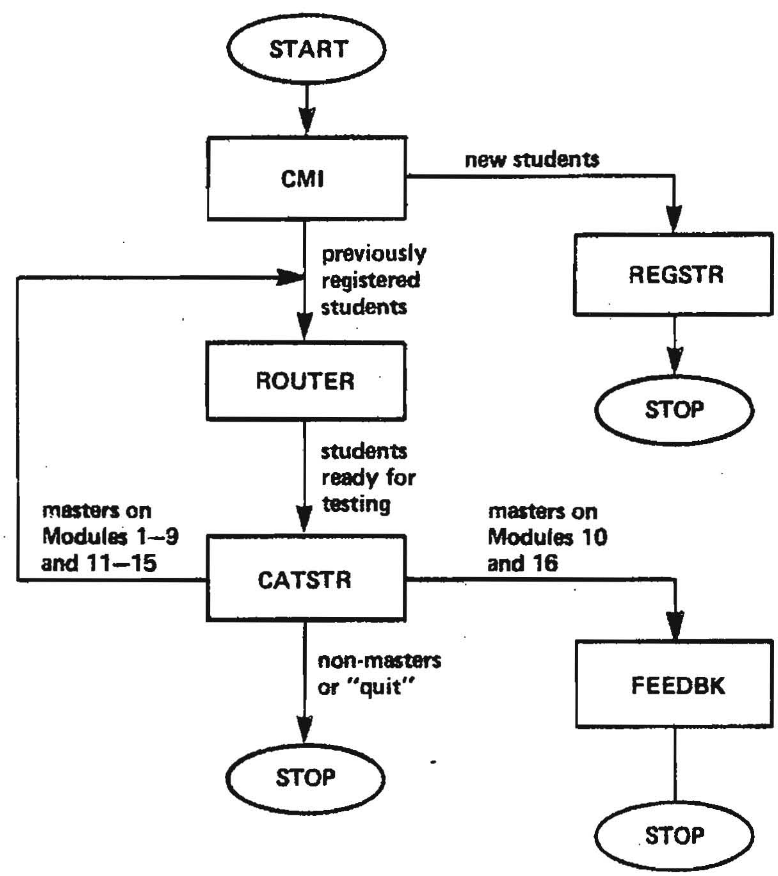


Figure 9

PROGRAM COMPONENTS OF THE CMI SYSTEM

<u>The registration programs</u>. The main Registration Program (CMI) performs two functions. First, it asks students to identify themselves with their code names (see Figure 10). These uniquely identify each student so that data stored on their work will be confidential. Second, program CMI uses this code name as a key to search for student-specific data in the roster file. This data includes students' first and last names and the type of terminal that each student is using. Students who have not yet registered are branched to the New Student Registration Program.

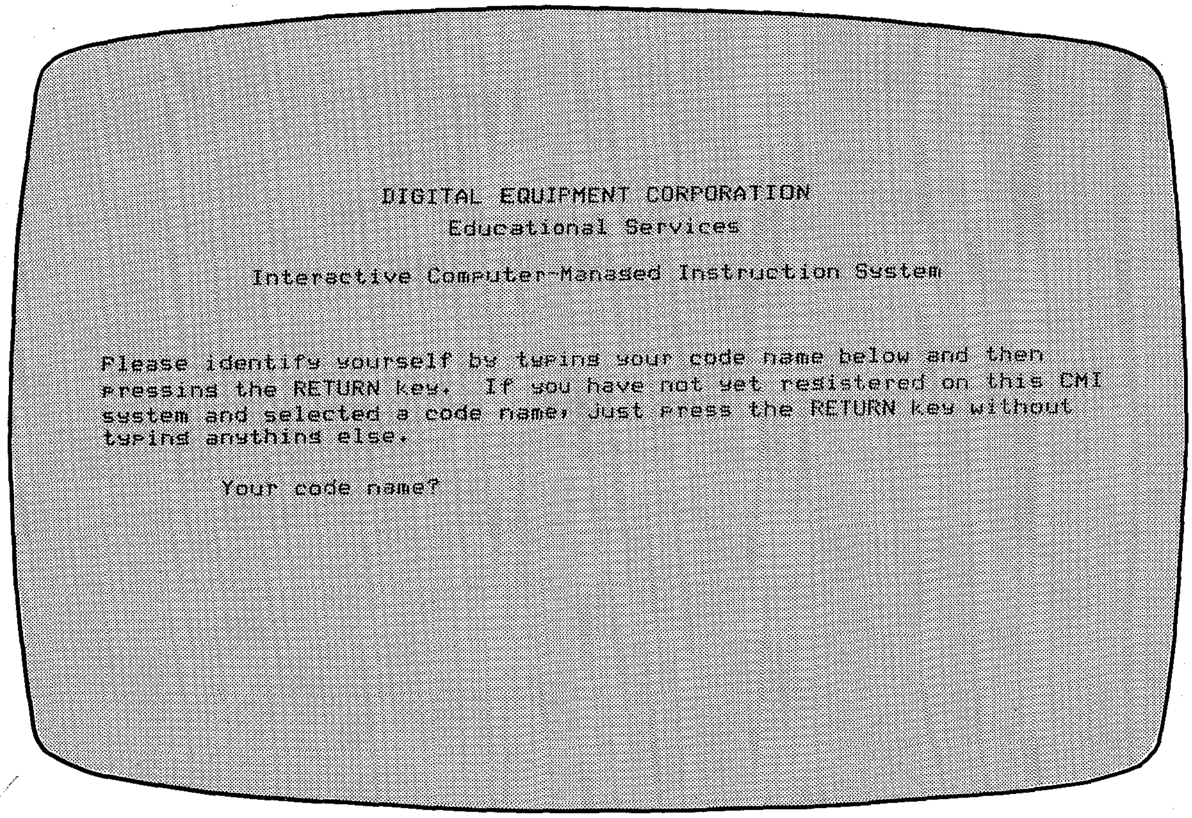


Figure 10

FIRST DISPLAY GENERATED BY THE CMI PROGRAM

The New Student Registration Program (REGSTR) allows students to register themselves on the system. This feature was necessary because the training package was used in a self-paced mode. Students entered their terminal types, their names, and their addresses. They were then asked demographic questions regarding their age, education, and motivation for taking the course. This demographic data is used to interpret the system usage data.

The New Student Registration Program also produced a registration form from the data that it collected from students. Students were instructed to print this form on paper and mail it to the course development group. These forms were acknowledged and the demographic data were extracted.

The router program. The purpose of the Router Program (ROUTER) is to identify what tests the student is ready to take. This program reads the prerequisite modules for each of the 16 modules from a data file. This data, combined with student status data, is used to determine the modules for which the student has met the prerequisites.

If a student has met the prerequisites for only one module, the appropriate pretest or posttest for that module is generated (by the Test Administration Program). If a student has met the prerequisites for more than one module, he or she is given the choice of the module to be tested. Thus the system assures that students take the tests in the order specified by the established module hierarchy. This constraint strengthens the instructional design of the course by requiring students to possess the prerequisite knowledge for each module that they study. The system does not allow students to take tests on modules for which they have not yet mastered all of the prerequisite modules.

The Router Program also gives students the option of displaying their status in the course. The status for each module is reported as one of the following:

(1) Not attempted

(2) Pretest tried but not completed satisfactorily

(3) Pretest completed satisfactorily; posttest skipped

(4) Posttest tried but not completed satisfactorily

(5) Posttest completed satisfactorily

Students are allowed to take a pretest for a module only once. All subsequent tests are automatically interpreted as posttests, because it is assumed that students who do not demonstrate mastery on a pretest will go study the corresponding module. Students are allowed to take posttests as many times as necessary to complete them satisfactorily, i.e., demonstrate mastery on all of the objectives in that module. (Tests are generated interactively in real time, so no two tests are exactly alike. See the discussion below.)

<u>The testing program</u>. The Computer-Assisted Test Administration Program (CATSTR) generates both pretests and posttests. These tests are administered to students at a computer terminal. The Test Administration Program presents true/false, yes/no, and multiple choice items (with either four or five alternatives). A typical item display is shown in Figure 11.

The items are randomly selected from files for each module in which the items are categorized by objective.

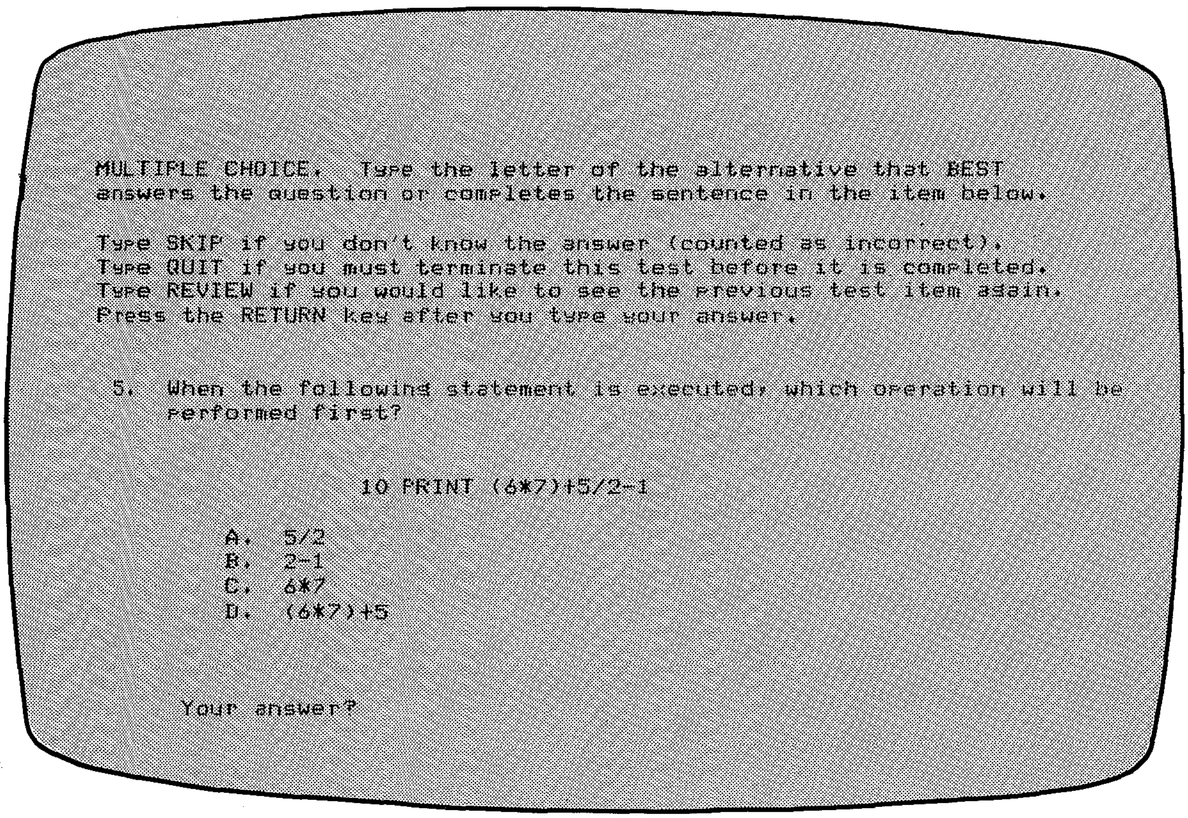


Figure 11

SAMPLE DISPLAY OF A MULTIPLE CHOICE ITEM
AS PRESENTED TO A STUDENT

Each objective is assigned a flag which is initialized to "reset" before each test administration. The item selection algorithm then proceeds as follows. First, an objective number is selected randomly and the flag for that objective is "set". Second, an item for that objective is selected randomly from the pool of items that have not yet been presented on the current test. For the next item, an objective number is selected from the pool of objectives whose flags are <u>not</u> set. Item selection proceeds as above. If no mastery decision can be made before all of the objective flags become set, these flags are simply reset and the process begins anew. However, individual items are never selected more than once on any single administration of a test.

The system is "human-engineered" to make it as foolproof as possible. For example, it provides the options "SKIP" and "QUIT" (see Figure 11). It provides "error" messages in plain English, e.g., if a student enters an unrecognizable response, the system prints the allowable responses, erases the student's previous response, and makes room for another one. If a student enters "REVIEW", a display similar to the one shown in Figure 12 is generated. (Students may review their work, but not change responses.)

Students who demonstrate mastery on a test are branched back to the Router Program. They then select the next module on which they would like to be tested. Students

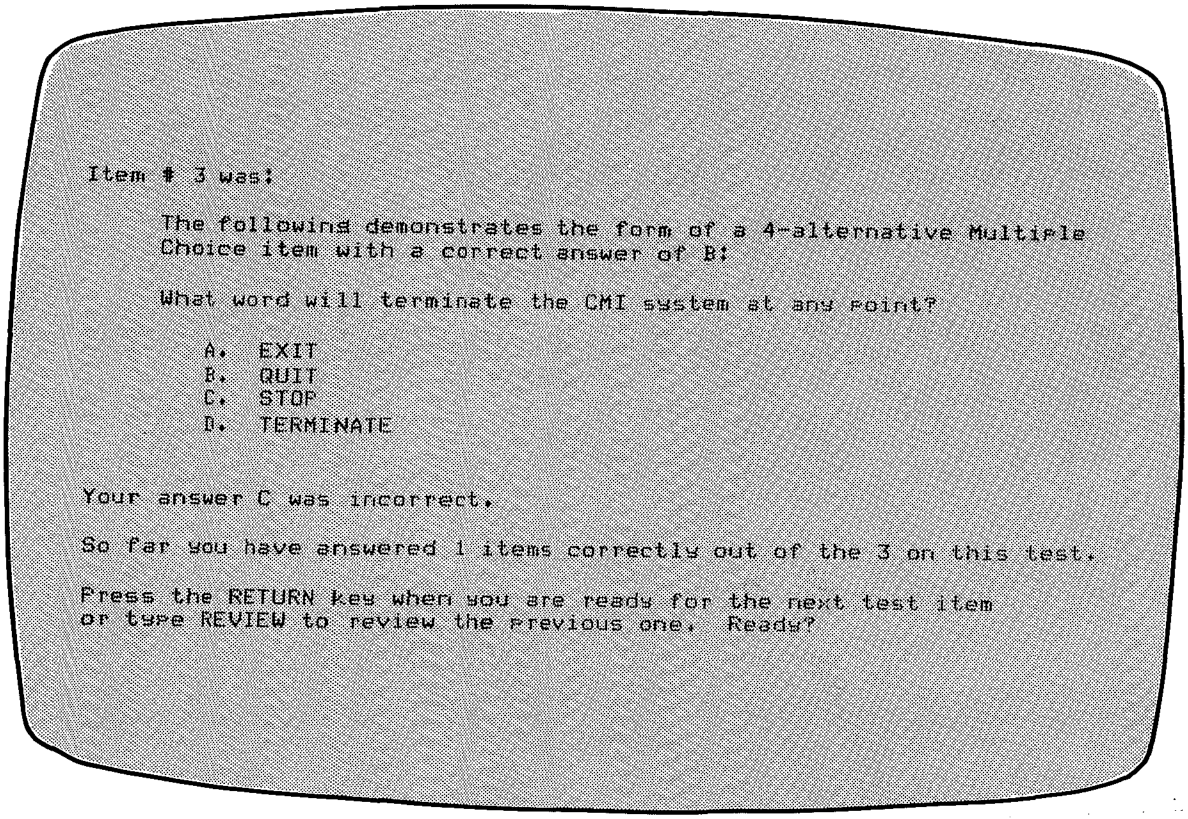


Figure 12

SAMPLE DISPLAY OF A TEST ITEM IN REVIEW MODE

who do not demonstrate mastery on a test are told the objectives on which they missed items. The CMI system then halts, and students are directed to do additional study off-line.

<u>The feedback program</u>. Students who demonstrate mastery on the tests for Modules 10 or 16 are branched to

the Feedback Program. Module 10 is the final module in the sequence of "core" modules, while Module 16 represents the highest level in the hierarchy of "extended" modules. This program asks for student comments on the course and generates another registration file. This file verifies that the student has earned a diploma and contains other data that is needed to evaluate the training package. This data was used as a backup to the data received from the course administrators on magnetic tape and to verify the validity and completeness of the tape data. Students were directed to print this file on paper and mail it to the course development group. Students were then sent their diplomas.

The Mastery Decision Model

The purpose of the Computer-Assisted Test Administration Program is to classify students as either masters or non-masters on the specific module being tested, and to make this classification in a minimum amount of time. To accomplish this, the Test Administration Program evaluates each student response with a sequential probability test ratio.

Computations. Ferguson's sequential probability scoring algorithm was designed for tests in which the probability of getting an item correct by guessing is the same for all items. Since the CMI system designed for this study presents true/false, yes/no, and four- and five-alternative multiple choice items, which have varying probabilities of getting them correct by guessing, the algorithm had to be modified. Each item is therefore assigned a weight, "W", according to the formula:

$$W = \frac{.25}{P_g}$$

where P_g is the probability of getting the item correct by guessing

Using this formula, true/false and yes/no items were assigned a weight of .25/.50 or 0.50. Four-alternative multiple choice items were assigned a weight of .25/.25 or 1.00, and five-alternative multiple choice items a weight of .25/.20 or 1.25.

After each test item is administered, the student's score, "S", is computed using the formula (cf. page 19):

$$S = C \times \log(P_1/P_0) + (T-C) \times \log((1-P_1)/(1-P_0))$$

where C is the sum of the weights of the items answered correctly, and
T is the sum of the weights of all items that have been presented (thus, T-C is the sum of the weights of the items answered incorrectly)

The student is classified as a master and testing is terminated if

$$S \leq \log(b/(1-a))$$

where a is the probability that a Type I error will occur, and b is the probability that a Type II error will occur

and at least one item has been presented on each objective in the module. (A Type I error is false negative error and occurs when a true master is classified as a non-master by the test. A Type II error is a false positive error and occurs when a true non-master is classified as a master.) If the above inequality is true but all objectives have not been tested, another item is presented.

The student is classified as a non-master and testing is terminated if

$$S \geq \log((1-b)/a)$$

regardless of the number of items presented on each objective. If neither of these inequalities is true, that is, if

$$\log(b/(1-a)) < S < \log((1-b)/a)$$

another test item is presented. The system continues in this manner until one of the first two inequalities becomes true or until 30 items have been administered. If no decision can be made after 30 items, the system classifies the student based on the differences between his or her score and the two criteria. The student is classified in the category whose criterion score is closest to his or her computed score after 30 items.

Derivation of test parameters. The CMI system generates both pretests and posttests. For this reason, it is important to realize that the seriousness of making Type I and Type II errors is different on pretests and posttests. If the system makes a Type II (false positive) error on a pretest, it will tell a student who has not studied the corresponding module to skip instruction that he or she really needs. This same error on a posttest is not as serious, because the student will have already studied the module at least once, and one can assume that at least some minimal learning has taken place. A Type I (false negative) error is never as serious as a Type II error, because this situation simply asks a student to repeat instruction that he or she does not really need. This wastes some time, but one can assume that it does not decrease the learner's proficiency level.

To take the relative importance of these errors into consideration, the CMI system uses the parameters shown in Table 2. These parameters were chosen for the following reasons. First, the pretest and posttest mastery and non-mastery criteria were set to span the "normal" mastery criterion percentage score of 70-80% that most criterion-referenced tests use when only one cutting score is employed. Second, the mastery criterion for pretests was increased 5% over that for posttests to reflect a slightly

Table 2

SEQUENTIAL TESTING PARAMETERS FOR PRETESTS AND POSTTESTS

Parameter	For Pretests	For Posttests
Mastery criterion (P_0)	0.90	0.85
Non-mastery criterion (P_1)	0.65	0.60
Prob. of Type I error (*a*)	0.058	0.104
Prob. of Type II error (*b*)	0.025	0.050

more stringent criterion for mastery if a module has not yet been studied. The non-mastery criterion for pretests was also increased 5% to keep the differences between these two criteria equal for both types of tests. (This was necessary because the difference between the two criteria is itself a factor in determining test length. As the difference increases, the number of test items required to make a decision at any given level of certainty decreases. Conversely, as the difference between the two percentage criterion levels decreases, the number of required test items increases.)

Third, the allowable probabilities of Type II (false positive) errors were set to 0.025 and 0.050, respectively, for pretests and posttests. The factor of 2 separating these parameters reflects the relative seriousness of making this type of error on pretests versus its seriousness on posttests. That is, it is estimated that the seriousness of making a Type II error on a pretest is twice as great as that on a posttest, so the allowable probability of this error on pretests was decreased by a factor of 2. Finally, the probabilities of Type I (false negative) errors were derived by computing the highest value that would usually require at least three items to be presented before any decision is made, unless the first two items are both five-alternative multiple choice items (with weights of 1.25 each). This philosophy was adopted because the author felt that students would distrust the system if they were regularly judged non-masters after only two items had been presented.

The absolute values of the error probabilities also warrant some discussion. Ferguson (1970) allowed probabilities of 0.20 and 0.10, respectively, for his Type I and Type II errors. These values reflect the same 2:1 ratio used in this study, but the absolute values are approximately twice the size. The main reason for selecting lower probabilities is that Ferguson's testing unit was the objective, while this study's testing unit is the module (a group of 2 to 20 objectives). It was felt that when working at the module level, the consequences of errors of classification are more serious than at the objective lever. Thus, the absolute values of the allowable error probabilities were lowered.

Figure 13 shows how these parameters reflect the mastery decision model in terms of raw scores. Figure 13a is a graph of the pretest decision rules, while Figure 13b shows the posttest decision rules. Note the difference in the sizes of the two master areas and the specific points labelled. The point labelled "(2.5,0)" in both graphs indicates that the earliest that a non-master decision could be made on either test is after the sum of the weights of all items presented totals at least 2.5. If, at this time, the student has not answered any items correctly, he or she is classified as a non-master.

In Figure 13a, the point labelled "(11.5,11.5)" indicates that the earliest that a master decision could be made on a pretest is after items having a total weight of 11.5 have been presented and all items have been answered correctly. Contrast this point with the one labelled "(8.5,8.5)" in Figure 13b. The latter indicates that the earliest that a master decision could be made on a posttest is after items having a total weight of at least 8.5 have been presented and answered correctly. Therefore, the posttest mastery criterion is less stringent than the pretest mastery criterion. This relationship is exactly the one desired, because it reflects that classifying a student as a master on a posttest is less serious than doing so on a pretest.

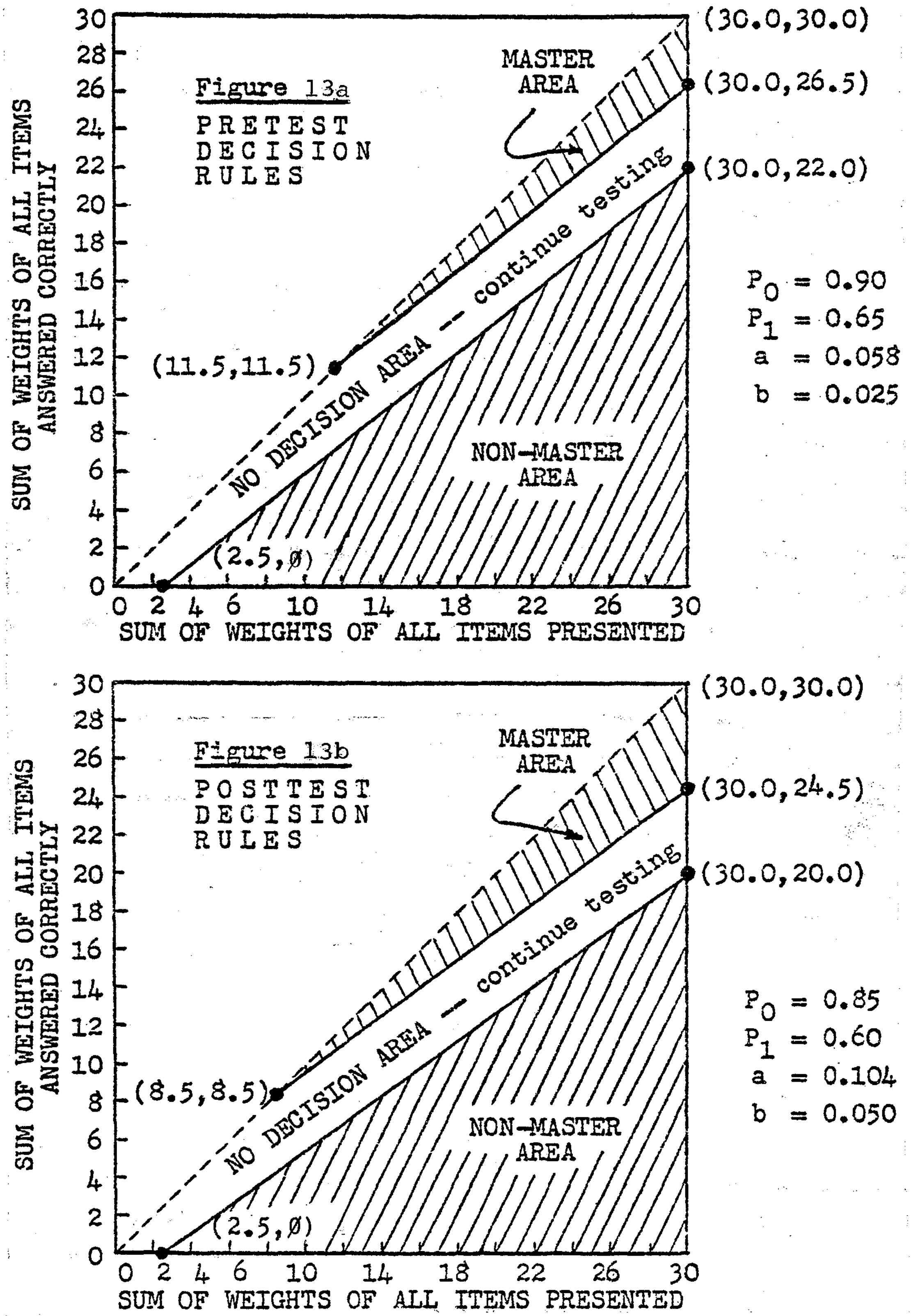


Figure 13

GRAPHS OF THE PRETEST AND POSTTEST DECISION RULES

The Reliability Model

This study assesses criterion-referenced test reliability as a reliability of classification. The two sets of test data used to assess this reliability were the mastery decisions made on the normal (variable length) versions of specific tests and those made on the same tests when they were extended to 30 items. This is a within-subject design because it yields two measures for a single student on a single test. To do this, every fifth test presented to any particular student was extended to 30 items in length, regardless of the test parameters. When the scoring algorithm made its initial decision, a tentative mastery classification was recorded. The system then continued presenting test items until the maximum of 30 had been presented, at which time the final master classification was recorded. (The transition from short test to long test was imperceptible to the student being tested.) This data was analyzed to determine the percentage of agreement between the two classifications, and the kappa, phi, and G indices (cf. pages 25-26).

The following data were also recorded by the system so that the reliability data could be put into perspective:

(1) the total amount of time that each student was logged in and running the CMI programs,

(2) the number of times that each student logged into the CMI system,

(3) the total amount of time that each student spent taking tests,

(4) the number of test items answered by each student,

(5) the amount of time spent on the shorter portion of each extended test and the amount of time spent on the entire 30-item test,

(6) the number of items at which testing could have been terminated for each extended test,

(7) tallies of the number of tests that contained each possible number of items (1-30) for

- pretests that resulted in master classifications
- pretests that resulted in non-master classifications
- posttests that resulted in master classifications, and
- posttests that resulted in non-master classifications,

(8) complete criterion-referenced item analysis data for both pretests and posttests, and

(9) a chronological history of each test administered to each student.

Conducting the Study

Selection of test sites. Test sites for the study were selected on the following criteria:

(1) the possession of a Digital PDP-11 computer running the RSTS/E operating system,

(2) geographic proximity to the author's place of business,

(3) potential for extensive use of the training package,

(4) willingness to pay a nominal fee for the use of the training package and to sign a li-

> censing and non-disclosure agreement binding them not to advertise their use of The BASIC Primer until the product had been announced by Digital Equipment Corporation.

Three test sites were selected that met these criteria. They were Wachusett Regional School District in Holden, Massachusetts, Rhode Island Junior College in Lincoln, Rhode Island, and Falmouth High School in Falmouth, Massachusetts.

System installation. After each test site submitted a purchase order to Digital to make the licensing agreement legally binding, the author personally delivered The BASIC Primer software on magnetic tape and copies of the self-paced course manuals. He discussed the study with representatives of each site and observed the system managers as they followed the instructions in The BASIC Primer Course Administrator's Guide to install the CMI software on their own computer systems. This involved loading and compiling the CMI system programs, defining site-dependent system parameters, building the test files, and initializing the data files. All of these tasks except loading and compiling were accomplished by an interactive, menu-driven applications program with many human-engineering features similar to those in the student CMI programs. The installation was the only time that the author visited any of the sites. All subsequent communication was accomplished by telephone and letter.

Data collection and analysis. The system manager at each test site was requested to copy all of the CMI data files onto a magnetic tape every week. These tapes were mailed to the author, who loaded them onto a master computer system. Here the data was combined and processed by analysis programs.

The output of these analysis programs is discussed in the next chapter.

Chapter 4

RESULTS OF THE STUDY

The First Subproblem

The first subproblem was to develop a workable CMI system that could be customer-installed and used by computer-naive learners. Another aspect of this subproblem was to implement a scheme for gathering system usage data from customer sites for analysis.

Ease of customer installation. The data relating to the first aspect of this subproblem is subjective. The CMI system software was delivered personally to each test site by the author. The first installation was performed at Wachusett Regional School District, where the system manager is a knowledgeable and experienced user of the RSTS/E operating system. He elected to build the CMI system on a private disk pack, and this required some modifications to the administrative software components. These modifications were relatively minor, however, and the author was able to make them at the site. Once the modifications were made, the Wachusett system manager was able to follow the Course Administrator's Guide and build the system.

The second and third installations were performed at Rhode Island Junior College and Falmouth High School, respectively. In contrast to Wachusett, the system managers at these sites were both new RSTS/E users. They were, however, both able to build the CMI system with the author supplying only "yes" and "no" answers to simple questions. These installations went smoothly, each requiring approximately two hours. Thus the author feels that the CMI system can be easily and reliably installed by customers, but that several improvements in the Course Administrator's Guide are warranted. These include an administrative command summary and at least one example of a complete system build.

Usability by computer-naive learners. The data relating to the second aspect of this subproblem is also somewhat subjective. A total of 65 learners used the CMI system at all three test sites. Since the timing of the study did not coincide with normal semesters, almost all of the learners were adult faculty members (mean age = 33.8), although it was also used by a 9-year-old who completed the first 3 modules and a 12-year-old who completed all of the core modules and earned a diploma. Some of the learners had previous computer experience, but virtually none of them had previous experience with the RSTS/E operating system or with interactive computer-managed instruction. All of the learners took the course for their own edification; none took it as part of a formal instructional program. The distribution of the learners' academic levels is shown in Table 3.

One measure of the system's usability by these learners was the number of calls that the author received regarding problems with the software. In the two months that the system was used, the author received only one such call. The problem was caused by an incorrectly dimensioned array, and the system managers were able to correct this problem by "patching" a single line of code. Instructions

Table 3

DISTRIBUTION OF LEARNERS' ACADEMIC LEVELS

N = 65

Academic Level	Frequency	Percentage
Not yet completed high school	6	9.2 %
Completed high school	7	10.8 %
Currently enrolled in an undergraduate college or university program	5	7.7 %
Completed a bachelor's program	11	16.7 %
Currently enrolled in a graduate college or university program	4	6.2 %
Completed a graduate program	27	41.5 %
No response	5	7.7 %

for doing this were provided to the test sites via letter. No calls were received which questioned how the system worked. A copy of the directions that students received for using the system is provided in Appendix A.

Another measure of students' success in using the system was the completeness of the system usage data. This data was complete except for the registration forms, which were sometimes not returned and sometimes not filled out entirely (some students failed to answer some questions). All statistical data stored automatically by the CMI system was complete.

Feasibility of data collection. System managers were asked to dump the data files on their systems to magnetic tape once each week and to mail their tapes to the author. As it turned out, they mailed tapes every two or three weeks, depending upon their systems' usage levels.

All tapes received by the author (a total of 11) were in excellent condition. These tapes were easily loaded onto a master computer and read with no difficulty or loss of data. All of the system usage data reported in this dissertation was derived from the tapes submitted by users. Thus it may be concluded that a viable scheme for gathering system usage data was implemented: the mailing of magnetic tapes.

The Second Subproblem

The second subproblem was to design the CMI system so that learners were forced to work through the modules in the sequence prescribed by the module hierarchy and so that excessive test taking would be discouraged.

Data for this subproblem is in the form of a chronological history of each test that was administered. Before each test was administered, the system stored a test history record containing the following data:

(1) the student's number,

(2) the number of the module to be tested,

(3) whether the ensuing test was to be a "normal" test or one "extended" to 30 items (for assessing reliability),

(4) the mastery criteria to be used ("pretest" or "posttest"),

(5) the date, and

(6) the time of day.

A computer program was written to display this chronological test history data by student. The output of this program for the Falmouth High School test site is shown in Figure 14. In addition to the six data items described above, Figure 14 also shows the differential time, "D", between successive test administrations when this time was less than 30 minutes. This data is typical of the other two test sites

[Text continues on page 87.]

Data for FALMOUTH HIGH SCHOOL 02-Feb-79, page 77

TEST HISTORY DATA

STUDENT 1

STUDENT 2

Mod 1	Normal	Pretest	17-Nov-78	12:22	PM	
	Normal	Posttest	17-Nov-78	12:29	PM	

STUDENT 3

Mod 1	Normal	Pretest	19-Nov-78	11:38	AM	
Mod 2	Normal	Pretest	19-Nov-78	11:50	AM	
	Extended	Posttest	20-Nov-78	10:40	AM	
Mod 3	Normal	Pretest	20-Nov-78	11:00	AM	
	Normal	Posttest	20-Nov-78	11:16	AM	D = 16 min
Mod 4	Normal	Posttest	20-Nov-78	11:20	AM	
	Normal	Posttest	28-Nov-78	02:56	PM	
Mod 5	Extended	Posttest	28-Nov-78	03:04	PM	
Mod 6	Normal	Pretest	29-Nov-78	09:14	AM	
	Normal	Posttest	29-Nov-78	02:13	PM	
Mod 7	Normal	Posttest	29-Nov-78	02:32	PM	
Mod 8	Normal	Posttest	30-Nov-78	09:08	AM	
Mod 9	Extended	Pretest	04-Dec-78	09:02	AM	
	Normal	Posttest	04-Dec-78	09:21	AM	D = 19 min
	Normal	Posttest	05-Dec-78	09:08	AM	
	Normal	Posttest	08-Dec-78	07:00	PM	
	Normal	Posttest	14-Dec-78	09:01	AM	

Figure 14

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

Mod 10	Extended	Pretest	14-Dec-78	09:13	AM	
	Normal	Posttest	15-Dec-78	09:22	AM	
	Normal	Posttest	15-Dec-78	09:23	AM	D = 1 min
Mod 13	Normal	Pretest	15-Dec-78	11:04	AM	
Mod 7	Normal	Posttest	15-Dec-78	11:28	AM	
	Extended	Posttest	18-Dec-78	09:00	AM	
Mod 13	Normal	Posttest	18-Dec-78	09:10	AM	
Mod 11	Normal	Pretest	20-Dec-78	09:34	AM	
	Normal	Posttest	21-Dec-78	09:12	AM	
Mod 12	Normal	Pretest	21-Dec-78	11:24	AM	
Mod 14	Extended	Pretest	02-Jan-79	02:50	PM	
	Normal	Posttest	03-Jan-79	09:03	AM	
Mod 13	Normal	Posttest	03-Jan-79	09:21	AM	
Mod 15	Normal	Pretest	03-Jan-79	11:27	AM	
	Normal	Posttest	09-Jan-79	09:06	AM	
	Extended	Posttest	09-Jan-79	09:10	AM	D = 4 min
Mod 12	Normal	Posttest	09-Jan-79	09:18	AM	
	Normal	Posttest	16-Jan-79	09:24	AM	
	Normal	Posttest	16-Jan-79	09:26	AM	D = 2 min

STUDENT 4

Mod 1	Normal	Posttest	20-Nov-78	08:07	AM	
Mod 2	Extended	Posttest	20-Nov-78	08:12	AM	
Mod 4	Normal	Posttest	21-Nov-78	07:31	AM	
	Normal	Posttest	02-Dec-78	10:35	PM	
	Normal	Posttest	02-Dec-78	10:39	PM	D = 4 min
	Normal	Posttest	03-Dec-78	08:40	PM	
	Extended	Posttest	03-Dec-78	09:16	PM	

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

	Normal	Posttest	03-Dec-78	09:53 PM	
Mod 6	Normal	Posttest	04-Dec-78	10:02 PM	
	Normal	Posttest	05-Dec-78	08:53 PM	
Mod 3	Normal	Posttest	05-Dec-78	10:40 PM	
Mod 5	Extended	Posttest	10-Dec-78	09:53 PM	
Mod 7	Normal	Posttest	17-Dec-78	03:34 PM	
Mod 8	Normal	Posttest	17-Dec-78	03:56 PM	
Mod 9	Normal	Posttest	17-Dec-78	09:43 PM	
Mod 5	Normal	Posttest	21-Dec-78	12:45 PM	
Mod 9	Extended	Posttest	21-Dec-78	01:41 PM	
	Normal	Posttest	23-Dec-78	02:30 PM	
Mod 10	Normal	Posttest	23-Dec-78	03:40 PM	
Mod 11	Normal	Posttest	25-Dec-78	04:44 PM	
Mod 12	Normal	Posttest	27-Dec-78	05:21 PM	

STUDENT 5

Mod 1	Extended	Posttest	29-Nov-78	09:05 AM	
Mod 2	Normal	Pretest	29-Nov-78	09:17 AM	
	Normal	Posttest	04-Dec-78	08:21 AM	
Mod 4	Normal	Pretest	04-Dec-78	08:27 AM	
	Normal	Posttest	08-Dec-78	09:04 AM	
	Extended	Posttest	08-Dec-78	10:51 AM	
	Normal	Posttest	08-Dec-78	11:21 AM	D = 30 min
	Normal	Posttest	08-Dec-78	12:29 PM	
Mod 3	Normal	Pretest	08-Dec-78	12:35 PM	

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

Mod 5	Extended	Pretest	08-Dec-78	02:39 PM	
	Normal	Posttest	08-Dec-78	02:57 PM	D = 18 min
Mod 6	Normal	Pretest	08-Dec-78	03:02 PM	
	Normal	Posttest	20-Dec-78	03:05 PM	
	Extended	Posttest	21-Dec-78	10:54 AM	
	Normal	Posttest	21-Dec-78	11:19 AM	D = 25 min
	Normal	Posttest	21-Dec-78	12:25 PM	
Mod 7	Normal	Pretest	29-Dec-78	02:33 PM	
Mod 8	Normal	Pretest	09-Jan-79	03:38 PM	

STUDENT 6

Mod 1	Normal	Pretest	21-Nov-78	01:50 PM	
Mod 2	Normal	Pretest	21-Nov-78	02:04 PM	
	Normal	Posttest	04-Dec-78	09:30 PM	
	Normal	Posttest	10-Dec-78	08:13 PM	
	Extended	Posttest	16-Dec-78	11:37 AM	
Mod 4	Normal	Pretest	16-Dec-78	12:01 PM	
	Normal	Posttest	30-Dec-78	01:59 PM	

STUDENT 7

Mod 1	Normal	Posttest	28-Nov-78	12:27 PM	
Mod 2	Normal	Pretest	28-Nov-78	12:45 PM	
	Normal	Posttest	01-Dec-78	02:31 PM	
	Extended	Posttest	01-Dec-78	02:38 PM	D = 7 min

STUDENT 8

Mod 1	Normal	Posttest	28-Nov-78	10:37 AM	

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

Data for FALMOUTH HIGH SCHOOL 02-Feb-79, page 81

Mod	2	Normal	Pretest	28-Nov-78	10:45 AM	
		Extended	Posttest	04-Dec-78	10:38 AM	
		Normal	Posttest	05-Dec-78	10:56 AM	
Mod	4	Normal	Posttest	06-Dec-78	10:29 AM	
		Normal	Posttest	11-Dec-78	03:03 PM	
		Normal	Posttest	11-Dec-78	03:12 PM	D = 9 min
Mod	3	Extended	Posttest	11-Dec-78	03:28 PM	
Mod	5	Normal	Posttest	11-Dec-78	03:45 PM	
Mod	6	Normal	Pretest	11-Dec-78	03:54 PM	
		Normal	Posttest	05-Jan-79	02:02 PM	
		Normal	Posttest	05-Jan-79	02:37 PM	
		Extended	Posttest	05-Jan-79	02:44 PM	D = 7 min
Mod	7	Normal	Pretest	05-Jan-79	03:05 PM	
Mod	8	Normal	Pretest	05-Jan-79	03:08 PM	
Mod	7	Normal	Posttest	05-Jan-79	03:23 PM	

STUDENT 9

Mod	1	Normal	Posttest	28-Nov-78	10:07 AM	
Mod	2	Extended	Posttest	28-Nov-78	10:16 AM	
		Normal	Posttest	10-Jan-79	10:47 AM	

STUDENT 10

Mod	1	Extended	Pretest	28-Nov-78	12:04 PM	
Mod	2	Normal	Posttest	01-Dec-78	02:58 PM	
		Normal	Posttest	01-Dec-78	03:12 PM	D = 14 min

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

STUDENT 11

Mod	1	Normal	Pretest	29-Nov-78	09:21 AM	
		Normal	Posttest	29-Nov-78	09:28 AM	
Mod	2	Normal	Posttest	12-Dec-78	09:06 AM	
		Normal	Posttest	12-Dec-78	09:14 AM	D = 8 min

STUDENT 12

Mod	1	Normal	Pretest	28-Nov-78	01:44 PM	
		Normal	Posttest	28-Nov-78	01:49 PM	
Mod	2	Normal	Pretest	30-Nov-78	07:53 AM	
		Extended	Posttest	30-Nov-78	07:54 AM	D = 1 min
Mod	4	Normal	Posttest	01-Dec-78	07:49 AM	
		Normal	Posttest	01-Dec-78	07:52 AM	D = 3 min
Mod	6	Normal	Posttest	04-Dec-78	07:54 AM	
Mod	8	Normal	Pretest	04-Dec-78	08:14 AM	
		Extended	Posttest	11-Dec-78	07:50 AM	
Mod	9	Normal	Pretest	11-Dec-78	11:44 AM	
		Normal	Posttest	18-Dec-78	07:53 AM	
Mod	10	Normal	Posttest	18-Dec-78	11:35 AM	
Mod	7	Normal	Pretest	21-Dec-78	02:33 PM	
Mod	3	Extended	Posttest	03-Jan-79	02:03 PM	
Mod	5	Normal	Posttest	03-Jan-79	02:13 PM	
Mod	7	Normal	Posttest	16-Jan-79	02:02 PM	

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

STUDENT 13

Mod	1	Normal	Pretest	29-Nov-78	07:13	AM	
Mod	2	Normal	Pretest	29-Nov-78	02:00	PM	
		Extended	Posttest	30-Nov-78	07:13	AM	
Mod	4	Normal	Posttest	04-Dec-78	07:07	AM	
		Normal	Posttest	06-Dec-78	03:10	PM	
		Normal	Posttest	07-Dec-78	03:15	PM	
Mod	5	Normal	Pretest	08-Dec-78	02:10	PM	
Mod	3	Extended	Pretest	08-Dec-78	02:23	PM	
Mod	5	Normal	Posttest	08-Dec-78	02:52	PM	
Mod	6	Normal	Pretest	08-Dec-78	03:03	PM	
		Normal	Posttest	11-Dec-78	03:03	PM	
		Normal	Posttest	11-Dec-78	03:31	PM	D = 28 min
Mod	7	Extended	Pretest	11-Dec-78	03:35	PM	
Mod	8	Normal	Pretest	11-Dec-78	04:14	PM	
		Normal	Posttest	13-Dec-78	07:06	AM	
		Normal	Posttest	15-Dec-78	01:59	PM	
		Normal	Posttest	15-Dec-78	02:08	PM	D = 9 min
Mod	9	Extended	Pretest	15-Dec-78	02:24	PM	
		Normal	Posttest	19-Dec-78	06:07	PM	
		Normal	Posttest	22-Dec-78	07:04	AM	
		Normal	Posttest	09-Jan-79	03:39	PM	

STUDENT 14

Mod	1	Normal	Posttest	03-Dec-78	11:46	AM
Mod	2	Extended	Posttest	03-Dec-78	11:54	AM

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

Mod	4	Normal	Posttest	30-Dec-78	02:34 PM	

STUDENT 15

Mod	1	Extended	Pretest	06-Dec-78	10:25 AM	
Mod	2	Normal	Pretest	07-Dec-78	10:22 AM	
		Normal	Posttest	13-Dec-78	10:33 AM	

STUDENT 16

Mod	1	Normal	Posttest	06-Dec-78	02:35 PM	
Mod	2	Normal	Posttest	15-Dec-78	02:35 PM	
Mod	3	Normal	Pretest	15-Dec-78	02:49 PM	
Mod	4	Normal	Pretest	05-Jan-79	09:01 AM	
		Extended	Posttest	05-Jan-79	09:07 AM	D = 6 min
		Normal	Posttest	05-Jan-79	02:34 PM	
		Normal	Posttest	05-Jan-79	02:43 PM	D = 9 min
		Normal	Posttest	09-Jan-79	02:59 PM	
		Normal	Posttest	09-Jan-79	03:20 PM	D = 21 min
		Normal	Posttest	09-Jan-79	03:23 PM	D = 3 min
Mod	5	Normal	Posttest	10-Jan-79	03:05 PM	

STUDENT 17

Mod	1	Normal	Pretest	11-Jan-79	02:48 PM	
Mod	2	Normal	Pretest	11-Jan-79	02:55 PM	
		Normal	Posttest	12-Jan-79	10:21 AM	
Mod	4	Extended	Pretest	12-Jan-79	01:59 PM	
		Normal	Posttest	14-Jan-79	04:12 PM	

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

Data for FALMOUTH HIGH SCHOOL 02-Feb-79, page 85

Mod 6	Normal	Pretest	14-Jan-79	04:31 PM
Mod 8	Normal	Posttest	16-Jan-79	04:08 PM

SUMMARY

Number of Normal Pretests =	40		
Extended Pretests =	10		
Total Number of Pretests =		50	(28.6%)
Number of Normal Posttests =	102		
Extended Posttests =	23		
Total Number of Posttests =		125	(71.4%)

Total Number of Tests Taken =		175	

Instances of Repeated Tests =	74	
No. of Elapsed Times < 15 m =	15	(20.3%)
No. of Elapsed Times < 30 m =	22	(29.7%)

Figure 14 (continued)

TEST HISTORY DATA FOR FALMOUTH HIGH SCHOOL

and is used for the discussion that follows. (Test history for the other two sites is provided in Appendix B.)

Module sequences. The data in Figure 14 shows that the CMI system was 100% successful at assuring that students took the module tests in sequences that were compatible with the prescribed module hierarchy (see Figure 7). The data shows that students did take advantage of the course's non-linear structure and skipped around quite freely when they had met the prerequisites for more than one module. This skipping, however, was always within the confines of the prescribed module hierarchy. That is, no tests were administered on modules for which a student had not yet mastered the prerequisite modules.

Excessive test taking. The BASIC Primer modules were designed to require an average study time of approximately 30 minutes. Data stored by the CMI system showed that the average time required to take a test was 10.7 minutes. Thus it may be conjectured that students who took two successive tests on the same module with less than 10-15 minutes in between probably did not study or review the module, but rather simply repeated the test as soon as a non-mastery decision was made.

The data for Falmouth High School in Figure 14 indicate that there were 74 instances (out of 175) in which a student repeated a module test. (The data for Module 1 is ignored because the test for this module simply demonstrated the use of the CMI system. See Appendix A.) In 15 (20.3%) of the 74 instances, the elapsed time between the two successive tests was less then 15 minutes. In a total of 22 instances (29.7%), the elapsed time was less than 30 minutes. In 15 (68.2%) of these 22 instances, the student demonstrated mastery on the second test administration.

Thus the CMI system did not prevent students from repeating tests without studying in between. However, it can be seen that when students did this, the majority were able to demonstrate mastery on their second try. These data may be interpreted as indicating that even though students could repeat tests as soon as they were judged non-masters, very few did so unless they were quite certain that they could demonstrate mastery the second time around. Thus it may be suggested that the CMI system did discourage excessive test taking.

Other inferences from test history data. Three other inferences can be drawn from the test history data. First, the data shows that students in the test population required a total elapsed time of approximately 4 weeks to complete The BASIC Primer "core" modules (see Figure 7). Second, the test history data confirms that, within certain constraints, every fifth test presented to each student was extended to a 30 item test for assessing reliability as planned. Third, the test history data shows that students often skipped the pretests. This may have been because students preferred to study the modules before being tested on them or because they claimed previous knowledge of the subject matter. In any event, the ratio of pretests to posttests presented by the system was always less than 1. The latter two facts are important in supporting the validity of the reliability calculations presented in the discussion of the fourth subproblem.

The Third Subproblem

The third subproblem was to design the CMI system so that its use required a minimum amount of time.

The CMI system stored two types of data relating to this subproblem. The first is test time data composed of:

(1) the amount of time that each student spent on-line,

(2) the number of times that each student logged into the CMI system,

(3) the amount of time that each student spent taking tests,

(4) the number of tests that each student took, and

(5) the number of test items presented to each student.

These data and related values are summarized in Table 4. The average time per test was computed by dividing the total student testing time by the number of tests taken. The

Table 4

SUMMARY TEST TIME AND RELATED DATA FOR EACH SITE

	All Sites Combined	Wachusett Regional School District	Rhode Island Junior College	Falmouth High School
Total Student Time On-Line (hours:minutes)	175:54	15:41	124:09	36:04
Number of Student Logins	659	86	343	230
Total Student Testing Time (hours:minutes)	124:43	11:01	85:46	27:56
Number of Tests Taken	708	108	420	180
Number of Test Items Presented	10900	1271	6585	3044
Average Time Per Test (minutes)	10.6m	6.1m	12.3m	9.3m
Average Login Overhead (percent)	41.0%	42.4%	44.7%	29.1%
Average Time per Item (seconds)	41.2s	31.2s	46.9s	33.0s

login overhead was computed by dividing the difference between the total student time on-line and the total student testing time by the total student testing time. The average time per item was computed by dividing the total student testing time by the number of test items presented.

The data in Table 4 shows that:

(1) students typically required 10.6 minutes to take one test,

(2) students typically took only one or two tests each time they logged in to the CMI system (659 tests taken with 708 logins), and

(3) students typically spent about half as much time logging into the system and displaying their status as they did taking tests (41.0% login overhead).

One may therefore conservatively estimate that students typically required approximately 15 minutes to log into the system and take a test. This figure is highly supportive of the contention that use of the CMI system required a minimum amount of time, especially when one considers that the test time data includes tests extended arbitrarily to 30 items (about 20% of all test administered) to allow assessment of reliability.

The second type of data that relates to this sub-problem is test length data. It was stated in Chapter 3 that the CMI system employed a sequential probability test ratio in an attempt to reduce the number of test items required to assess mastery. To assess the effect of this mastery algorithm, the system stored tallies of the number of tests that resulted in each possible test length (1-30 items). A program was then written to display and analyze this data. The output of this program for all sites combined is shown in Figure 15. (The output for each individual site is provided in Appendix C.)

The data in Figure 15 are broken down by pretests and posttests and masters and non-masters. (Note that the total number of tests taken reported in Figure 15 is less than that reported in Table 4. This is because Table 4 includes data for Module 1, the demonstration module, while Figure 15 only includes data for Modules 2 through 16.) The totals are also corrected for extended and student-terminated tests. (Student-terminated tests are those in which students entered "quit" before the test was completed.)

Corrections were made as follows. First, the number of tests 1 item in length ("19" in Column 3 and "30" in Column 5) were subtracted from the "Actual Totals" because the earliest that any classification could be made using the sequential probability test ratio was on a test 2 items in length. The number of tests 1 item in length must therefore all have been student-terminated tests. Second, the number of tests 30 items in length in each column was reduced by the total number of extended master or non-master tests times the fraction of 30-item tests in the column.

No. of Pretests	= 216	No. of Extended Masters	= 46
No. of Posttests	= 397	No. of Extended Non-Masters	= 49
Total No. of Tests	= 613	Total No. of Extended Tests	= 95

Test Length in No. Items	No. of Pretest Masters	No. of Pretest Non-Masters	No. of Posttest Masters	No. of Posttest Non-Masters
1	0	19	0	30
2	0	2	0	7
3	0	13	0	11
4	0	10	0	18
5	0	10	0	11
6	0	8	0	9
7	0	13	0	4
8	0	4	4	7
9	0	9	9	6
10	0	3	14	4
11	2	8	10	4
12	1	2	3	4
13	1	3	10	5
14	1	4	4	3
15	1	3	5	0
16	0	1	5	2
17	1	2	7	1
18	0	3	12	2
19	0	4	9	2
20	1	3	5	4
21	1	3	5	4
22	2	4	1	2
23	0	2	4	1
24	1	1	9	2
25	0	1	3	0
26	1	1	6	1
27	1	2	3	3
28	2	6	5	4
29	5	3	5	3
30	25	23	84	21
Actual Totals:	46	170	222	175
Corrected Totals:	35	128	187	124
Mean Length:	25	11	21	10
Median Length:	29	9	20	8

Figure 15

TEST LENGTH DATA FOR ALL SITES COMBINED
ON MODULES 2 THROUGH 16

For example, the actual number of 30-item tests in Column 2, pretest masters, is 25, and the actual number of 30-item tests in Column 4, posttest masters, is 84. The total number of 30-item master tests is therefore 25 + 84 = 109. The number of extended master tests derived from the test history data is 46 (shown at the top right of the figure). The portion of these 46 that were subtracted from the 25 in Column 2 was 46 * (25/109) = 11. These 11 account for the correction of the Column 2 total from 46 to 35. Likewise, the proportion subtracted from the 84 in Column 4 was 46 * (84/109) = 35. The Column 4 total was therefore corrected to 222 - 35 = 187.

In Columns 3 and 5, the total number of 30-item non-master tests is 23 + 21 = 44. The number derived from the test history data, however, is 49. This indicates that at least 5 of the extended non-master tests were terminated by students prematurely. Therefore, all of the 44 30-item tests were assumed to be extended tests and were subtracted out. The corrected Column 3 total was computed as 170 (actual) - 23 (30-item) - 19 (1-item) = 128. The corrected Column 5 total was 175 - 21 - 30 = 124. The mean and median test lengths were then computed with the corrected totals.

The data in Figure 15 shows that, in 71.6% of the tests that resulted in mastery classifications and in 100% of the tests that resulted in non-mastery classifications, the sequential probability test ratio was able to terminate the test before it reached 30 items in length. This indicates that the sequential probability test ratio contributed significantly to reducing test lengths and shortening test time.

The test length statistics shown at the bottom of Figure 15 also reflect this contention because the mean and median test lengths for each of the four types of tests are all less than 30 items. In addition, the test length statistics also accurately reflect the various certainty criteria (error probabilities) for each of four types of tests as set by the author. This relationship is shown in Table 5. Further data on the 95 tests extended to 30 items indicates that, on the average, these extended test could have been terminated after 18.5 items had been presented if the sequential probability test ratio had been applied. These

Table 5

A PRIORI ERROR PROBABILITIES AND
A POSTERIORI MEDIAN TEST LENGTHS

Test Type and Classification	Critical Score	Allowable Error Probability	Median Test Length
Pretest Master	90 %	.025	29
Posttest Master	85 %	.050	20
Pretest Non-Master	65 %	.058	9
Posttest Non-Master	60 %	.104	8

early terminations would have resulted in an average time saving of 6.2 minutes on each extended test. It may therefore be concluded that:

(1) the CMI system designed for this study was at least moderately successful in meeting its objective of requiring a minimum amount of time for its use, and

(2) the sequential probability test ratio was a significant factor in reducing test lengths and testing time.

The Fourth Subproblem

The fourth subproblem was to implement an algorithm that maintained reliability with short tests and to devise and implement a scheme for assessing the validity of this algorithm.

The algorithm selected to maintain reliability while reducing test length was the sequential probability test ratio. The data in Figure 15 and Table 5 show that this algorithm was correctly implemented. Figure 15 shows that test lengths did indeed vary widely, and Table 5 shows that these variances were consistent with the *a priori* error probabilities set by the author, i.e., the median test length varies inversely with the allowable error probability.

The scheme devised for assessing the reliability of the sequential probability test ratio was to extend every fifth test presented to each individual student to 30 items, regardless of the test parameters. The system then stored:

(1) the system's mastery decision on the extended test,

(2) the decision that would have been made if the test had been terminated early.

These data can be arranged in a 2x2 table as shown in Table 6. Using the data in this format, the percentage of agreement, kappa, phi, and G indices can be computed using the formula discussed in Chapter 2 (cf. pages 25-26).

Table 6

MODEL OF RELIABILITY DATA

		EARLY DECISION		
		Master	Non-Master	
EXTENDED DECISION	Master	A	C	A+C
	Non-Master	B	D	B+D
		A+B	C+D	N=A+B+C+D

Table 7 presents the combined test reliability data for all sites. (The data for each individual site is included in Appendix D.) This data shows that there were a total of 95 extended tests and that in 92 (96.8%) of these the decisions on the short tests and extended tests agreed. In 3 cases, the system would have made a Type I or false negative error (by classifying a true master as a non-

Table 7

TEST RELIABILITY DATA FOR ALL SITES COMBINED ON MODULES 2-16

		EARLY DECISION		
		Master	Non-Master	
EXTENDED DECISION	Master	43	3	43
	Non-Master	0	49	49
		43	52	N = 95

master) if the early decision had been allowed to stand. The system would not have made any Type II (false positive) errors if it had accepted its early decisions. That is, it would never have classified a true non-master as a master.

The corresponding reliability indices for the data in Table 7 are as follows:

Percentage of Agreement = 0.968
Kappa = 0.937
Phi = 0.939
G = 0.937

These indices indicate that the sequential probability test ratio yielded highly reliable classifications, even when the tests were shortened. In fact, as discussed in relation to the third subproblem, average savings of 11.5 items and 6.2 minutes per test could have been realized if the early decisions had been accepted.

Some question may arise as to whether the observed results in Table 7 vary significantly from the theoretical results based on the *a priori* error probabilities. This question may be addressed as follows.

The expected values in each cell cannot be computed simply from the marginal totals, because the expected ratios of the values in cells *A* to *B* and *C* to *D* are not 1:1. Rather, these ratios are a function of the *a priori* error probabilities. For example, if all extended tests were posttests, the expected value in cell *A* would be

$$.95 \times (A+B)$$

because .95 = 1 - .05 (the allowable probability of a false positive error on posttests). The expected value in cell *B* would be

$$.05 \times (A+B)$$

However, not all extended tests were posttests. The pretest and posttest fractions were derived from the test history data by counting the number of extended pretests and extended posttests for Modules 2 through 16. For all sites combined, these fractions were .345 and .655, respectively. Given these fractions and the *a priori* error probabilities shown in Table 5, the expected values for each of the four cells may be computed as follows:

Cell _A_: $((.975x.345)+(.950x.655)) \times 43 = 41.221$

Cell _B_: $((.025x.345)+(.050x.655)) \times 43 = 1.779$

Cell _C_: $((.058x.345)+(.104x.655)) \times 52 = 4.583$

Cell _D_: $((.942x.345)+(.896x.655)) \times 52 = 47.417$

Using the observed and expected values, chi-square may be computed as follows (Ary et al., 1972):

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 2.512$$

With one degree of freedom, this value is not significant at the .05 level of confidence. It may therefore be concluded that, using the chi-square test, the observed data do not vary significantly from the theoretical data, and thus the reported reliability indices may be taken as valid measures.

However, the use of the chi-square test in this situation may be challenged. Several researchers (Issac and Michael, 1971; Lewis and Burke, 1949; Siegel, 1956) have pointed out that chi-square is not valid with "low" expected frequencies, specifically, frequencies less than 5. Even though the total N for Table 7 is 96, the expected frequencies for cells _B_ and _C_ do fall below 5. In such cases, Siegel (1956) and Issac and Michael (1971) recommend the use of the Fisher Exact Probability Test. However, the author does not feel that this test is appropriate for the data in Table 7 because its computation is based solely on the marginal totals and assumes that the expected ratios of the values in cells A to B and C to D are 1:1. As shown previously, this assumption is not valid for the data at hand.

There is one other test that can be applied to the data in Table 7: the Kolmogorov-Smirnov One-Sample Test. This test provides a measure of "the degree of agreement between the distribution of a set of sample values and some specified theoretical distribution" (Siegel, 1956). Moreover, Siegel claims, the Kolmogorov-Smirnov test is applicable for very small samples. The critical value in this test, D, is the maximum difference between the cumulative percentage of theoretical values for each category and the corresponding cumulative percentage of observed values. The data for computing the Kolmogorov-Smirnov D for the data in Table 7 is shown in Table 8.

The maximum difference, D, is 0.019. The significance of this value of D for an N of 95 is evaluated by dividing standard values (Siegel, 1956) by the square root of N. At the .05 level, the standard value is 1.36, and thus the critical value for significance is 0.140. Since 0.019 is less than 0.140, the Kolmogorov-Smirnov test also indicates that the observed data in Table 7 does not vary significantly from the theoretical cell frequencies.

It may therefore be concluded from the data in Figure 15 and Table 7 that the shortening of tests by applying a sequential probability test ratio did significantly reduce

Table 8

COMPUTATION OF THE KOLMOGOROV-SMIRNOV D
FOR THE DATA IN TABLE 7

N = 95

Cell	Theoretical Cumulative Percentage	Observed Cumulative Percentage	Difference
A	0.434	0.453	0.019
B	0.453	0.453	0.000
C	0.484	0.501	0.017
D	1.000	1.000	0.000

test lengths, but that it did not significantly reduce test reliability.

The Fifth Subproblem

The fifth subproblem was to devise and implement a scheme for assessing the validity of the CMI system's test items.

The CMI system stored tallies of the number of students choosing each alternative for each test item in four categories:

(1) pretests on which master decisions were made,

(2) pretests on which non-master decisions were made,

(3) posttests on which master decisions were made, and

(4) postests on which non-master decisions were made.

This data was used to generate a complete criterion-referenced item analysis for each test item. The item analysis data for Module 2 is presented in Figure 16. (Due to their volumes, data for Modules 3-16 are not included in this dissertation.)

The indices in Figure 16 were computed as follows. The difficulty indices are the number of masters and non-masters who responded correctly divided by the number of item administrations. (Asterisks are used to mark the correct response for each item.) The pretest and posttest discrimination indices are the differences between the percent of correct masters and the percent of correct non-masters. The pretest/posttest discrimination index is the percent of correct posttest masters minus the percent of correct pretest non-masters. This last index provides one of the best measures of the validity of a criterion-referenced test item (Glaser, 1963).

The data in Figure 16 show that Items 1-2 (Objective 1, Item 2), 1-5, 7-7, and 7-12 may be defective because they have negative pretest/posttest discrimination indices. Close examination of these items revealed indicators of the

[Text continues on page 109.]

Item Analysis Data for ALL SITES COMBINED on Module 2

22-Jan-79, page 1

Obj No.	Itm No.	Item Type	Test Type	No. of Admins	Percnt Corr Mast	Percnt Corr N-Mast	Diff. Index	Disc. Index	-- Number of Masters Responding -- T,Y A	F,N B	C	D	E	Skip	Number of Non-Masters Responding T,Y A	F,N B	C	D	E	Skip
1	1	MC4	Pre	13	1ØØ%	89%	Ø.92	Ø.11	Ø	Ø	4 *	Ø	--	Ø	Ø	Ø	8 *	1	--	Ø
			Post	21	1ØØ%	67%	Ø.95	Ø.33	Ø	Ø	18	Ø	--	Ø	1	Ø	2	Ø	--	Ø
			Pre/Pst	34	1ØØ%	83%	Ø.94	Ø.11	Ø	Ø	22	Ø	--	Ø	1	Ø	1Ø	1	--	Ø
1	2	MC4	Pre	16	5Ø%	42%	Ø.44	Ø.Ø8	1	2 *	Ø	1	--	Ø	2	5 *	Ø	5	--	Ø
			Post	18	23%	Ø%	Ø.17	Ø.23	3	3	Ø	6	--	1	4	Ø	Ø	1	--	Ø
			Pre/Pst	34	29%	29%	Ø.29	-Ø.19	4	5	Ø	7	--	1	6	5	Ø	6	--	Ø
1	3	MC4	Pre	5	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	Ø	Ø	2 *	Ø	--	Ø	Ø	Ø	3 *	Ø	--	Ø
			Post	17	1ØØ%	67%	Ø.88	Ø.33	Ø	Ø	11	Ø	--	Ø	1	1	4	Ø	--	Ø
			Pre/Pst	22	1ØØ%	78%	Ø.91	Ø.ØØ	Ø	Ø	13	Ø	--	Ø	1	1	7	Ø	--	Ø
1	4	TF	Pre	9	75%	1ØØ%	Ø.89	-Ø.25	1	3 *	--	--	--	Ø	Ø	5 *	--	--	--	Ø
			Post	17	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	Ø	13	--	--	--	Ø	Ø	4	--	--	--	Ø
			Pre/Pst	26	94%	1ØØ%	Ø.96	Ø.ØØ	1	16	--	--	--	Ø	Ø	9	--	--	--	Ø
1	5	TF	Pre	9	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	4 *	Ø	--	--	--	Ø	5 *	Ø	--	--	--	Ø
			Post	23	85%	1ØØ%	Ø.87	-Ø.15	17	3	--	--	--	Ø	3	Ø	--	--	--	Ø
			Pre/Pst	32	88%	1ØØ%	Ø.91	-Ø.15	21	3	--	--	--	Ø	8	Ø	--	--	--	Ø
1	6	YN	Pre	8	1ØØ%	5Ø%	Ø.63	Ø.5Ø	2 *	Ø	--	--	--	Ø	3 *	3	--	--	--	Ø
			Post	22	88%	1ØØ%	Ø.91	-Ø.12	15	2	--	--	--	Ø	5	Ø	--	--	--	Ø
			Pre/Pst	3Ø	89%	73%	Ø.83	Ø.38	17	2	--	--	--	Ø	8	3	--	--	--	Ø
1	7	YN	Pre	1Ø	1ØØ%	86%	Ø.9Ø	Ø.14	Ø	3 *	--	--	--	Ø	1	6 *	--	--	--	Ø
			Post	19	93%	25%	Ø.79	Ø.68	1	14	--	--	--	Ø	3	1	--	--	--	Ø
			Pre/Pst	29	94%	64%	Ø.83	Ø.Ø8	1	17	--	--	--	Ø	4	7	--	--	--	Ø
1	8	MC5	Pre	8	1ØØ%	86%	Ø.88	Ø.14	Ø	Ø	Ø	1 *	Ø	Ø	Ø	Ø	Ø	6 *	1	Ø
			Post	12	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	Ø	Ø	Ø	9	Ø	Ø	Ø	Ø	Ø	3	Ø	Ø
			Pre/Pst	2Ø	1ØØ%	9Ø%	Ø.95	Ø.14	Ø	Ø	Ø	1Ø	Ø	Ø	Ø	Ø	Ø	9	1	Ø
1	9	MC5	Pre	7	1ØØ%	83%	Ø.86	Ø.17	Ø	Ø	Ø	Ø	1 *	Ø	1	Ø	Ø	Ø	5 *	Ø
			Post	2Ø	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	Ø	Ø	Ø	Ø	17	Ø	Ø	Ø	Ø	Ø	3	Ø
			Pre/Pst	27	1ØØ%	89%	Ø.96	Ø.17	Ø	Ø	Ø	Ø	18	Ø	1	Ø	Ø	Ø	8	Ø
1	1Ø	MC5	Pre	13	1ØØ%	38%	Ø.62	Ø.63	Ø	5 *	Ø	Ø	Ø	Ø	4	3 *	Ø	1	Ø	Ø
			Post	19	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	Ø	16	Ø	Ø	Ø	Ø	Ø	3	Ø	Ø	Ø	Ø
			Pre/Pst	32	1ØØ%	55%	Ø.84	Ø.63	Ø	21	Ø	Ø	Ø	Ø	4	6	Ø	1	Ø	Ø

2	1	MC4	Pre	8	1ØØ%	67%	Ø.75	Ø.33	2 *	Ø	Ø	Ø	--	Ø	4 *	1	Ø	1	--	Ø
			Post	25	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	22	Ø	Ø	Ø	--	Ø	3	Ø	Ø	Ø	--	Ø
			Pre/Pst	33	1ØØ%	78%	Ø.94	Ø.33	24	Ø	Ø	Ø	--	Ø	7	1	Ø	1	--	Ø
2	2	MC4	Pre	9	1ØØ%	5Ø%	Ø.67	Ø.5Ø	Ø	Ø	Ø	3 *	--	Ø	Ø	1	1	3 *	--	1
			Post	2Ø	77%	57%	Ø.7Ø	Ø.2Ø	2	Ø	1	1Ø	--	Ø	1	Ø	2	4	--	Ø
			Pre/Pst	29	81%	54%	Ø.69	Ø.27	2	Ø	1	13	--	Ø	1	1	3	7	--	1

Figure 16

Obj No.	Itm No.	Item Type	Test Type	No. of Admins	Percnt Corr Mast	Percnt Corr N-Mast	Diff. Index	Disc. Index	-- Number of Masters Responding -- T,Y A	F,N B	C	D	E	Skip	Number of Non-Masters Responding T,Y A	F,N B	C	D	E	Skip
2	3	TF	Pre	15	1ØØ%	83%	Ø.87	Ø.17	3 *	Ø	--	--	--	Ø	1Ø *	2	--	--	--	Ø
			Post	24	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	17	Ø	--	--	--	Ø	7	Ø	--	--	--	Ø
			Pre/Pst	39	1ØØ%	89%	Ø.95	Ø.17	2Ø	Ø	--	--	--	Ø	17	2	--	--	--	Ø
2	4	TF	Pre	6	1ØØ%	8Ø%	Ø.83	Ø.2Ø	1 *	Ø	--	--	--	Ø	4 *	1	--	--	--	Ø
			Post	12	8Ø%	1ØØ%	Ø.83	-Ø.2Ø	8	2	--	--	--	Ø	2	Ø	--	--	--	Ø
			Pre/Pst	18	82%	86%	Ø.83	Ø.ØØ	9	2	--	--	--	Ø	6	1	--	--	--	Ø
2	5	TF	Pre	13	75%	56%	Ø.62	Ø.19	3 *	Ø	--	--	--	1	5 *	3	--	--	--	1
			Post	2Ø	94%	1ØØ%	Ø.95	-Ø.Ø6	15	1	--	--	--	Ø	4	Ø	--	--	--	Ø
			Pre/Pst	33	9Ø%	69%	Ø.82	Ø.38	18	1	--	--	--	1	9	3	--	--	--	1
2	6	MC4	Pre	13	1ØØ%	33%	Ø.54	Ø.67	Ø	4 *	Ø	Ø	--	Ø	2	3 *	2	Ø	--	2
			Post	24	95%	75%	Ø.92	Ø.2Ø	Ø	19	1	Ø	--	Ø	Ø	3	Ø	1	--	Ø
			Pre/Pst	37	96%	46%	Ø.78	Ø.62	Ø	23	1	Ø	--	Ø	2	6	2	1	--	2
2	7	TF	Pre	7	1ØØ%	5Ø%	Ø.71	Ø.5Ø	Ø	3 *	--	--	--	Ø	2	2 *	--	--	--	Ø
			Post	15	1ØØ%	67%	Ø.93	Ø.33	Ø	12	--	--	--	Ø	1	2	--	--	--	Ø
			Pre/Pst	22	1ØØ%	57%	Ø.86	Ø.5Ø	Ø	15	--	--	--	Ø	3	4	--	--	--	Ø
2	8	YN	Pre	11	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	2 *	Ø	--	--	--	Ø	9 *	Ø	--	--	--	Ø
			Post	23	1ØØ%	67%	Ø.91	Ø.33	17	Ø	--	--	--	Ø	4	2	--	--	--	Ø
			Pre/Pst	34	1ØØ%	87%	Ø.94	Ø.ØØ	19	Ø	--	--	--	Ø	13	2	--	--	--	Ø
2	9	YN	Pre	15	1ØØ%	64%	Ø.73	Ø.36	Ø	4 *	--	--	--	Ø	3	7 *	--	--	--	1
			Post	24	89%	67%	Ø.83	Ø.22	2	16	--	--	--	Ø	2	4	--	--	--	Ø
			Pre/Pst	39	91%	65%	Ø.79	Ø.25	2	2Ø	--	--	--	Ø	5	11	--	--	--	1
3	1	MC4	Pre	14	1ØØ%	4Ø%	Ø.57	Ø.6Ø	4 *	Ø	Ø	Ø	--	Ø	4 *	2	2	2	--	Ø
			Post	36	81%	78%	Ø.81	Ø.Ø4	22	1	1	3	--	Ø	7	Ø	Ø	2	--	Ø
			Pre/Pst	5Ø	84%	58%	Ø.74	Ø.41	26	1	1	3	--	Ø	11	2	2	4	--	Ø
3	2	TF	Pre	22	1ØØ%	88%	Ø.91	Ø.13	Ø	6 *	--	--	--	Ø	2	14 *	--	--	--	Ø
			Post	38	97%	88%	Ø.95	Ø.Ø9	1	29	--	--	--	Ø	1	7	--	--	--	Ø
			Pre/Pst	6Ø	97%	88%	Ø.93	Ø.Ø9	1	35	--	--	--	Ø	3	21	--	--	--	Ø
3	3	TF	Pre	23	1ØØ%	76%	Ø.83	Ø.24	Ø	6 *	--	--	--	Ø	4	13 *	--	--	--	Ø
			Post	37	87%	86%	Ø.86	Ø.Ø1	4	26	--	--	--	Ø	1	6	--	--	--	Ø
			Pre/Pst	6Ø	89%	79%	Ø.85	Ø.1Ø	4	32	--	--	--	Ø	5	19	--	--	--	Ø
3	4	TF	Pre	2Ø	83%	71%	Ø.75	Ø.12	5 *	1	--	--	--	Ø	1Ø *	4	--	--	--	Ø
			Post	37	71%	33%	Ø.62	Ø.38	2Ø	8	--	--	--	Ø	3	6	--	--	--	Ø
			Pre/Pst	57	74%	57%	Ø.67	Ø.ØØ	25	9	--	--	--	Ø	13	1Ø	--	--	--	Ø
3	5	MC4	Pre	19	8Ø%	86%	Ø.84	-Ø.Ø6	1	Ø	4 *	Ø	--	Ø	1	Ø	12 *	1	--	Ø
			Post	35	97%	1ØØ%	Ø.97	-Ø.Ø3	1	Ø	29	Ø	--	Ø	Ø	Ø	5	Ø	--	Ø
			Pre/Pst	54	94%	89%	Ø.93	Ø.11	2	Ø	33	Ø	--	Ø	1	Ø	17	1	--	Ø

Figure 16 (continued)

Obj No.	Itm No.	Item Type	Test Type	No. of Admins	Percnt Corr Mast	Percnt Corr N-Mast	Diff. Index	Disc. Index	-- Number of Masters Responding -- T,Y A	F,N B	C	D	E	Skip	Number of Non-Masters Responding T,Y A	F,N B	C	D	E	Skip
4	1	TF	Pre	15	100%	40%	0.60	0.60	5 *	0	--	--	--	0	4 *	5	--	--	--	1
			Post	26	71%	80%	0.73	-0.09	15	6	--	--	--	0	4	1	--	--	--	0
			Pre/Pst	41	77%	53%	0.68	0.31	20	6	--	--	--	0	8	6	--	--	--	1
4	2	TF	Pre	18	100%	64%	0.72	0.36	0	4 *	--	--	--	0	4	9 *	--	--	--	1
			Post	29	90%	63%	0.83	0.28	2	19	--	--	--	0	3	5	--	--	--	0
			Pre/Pst	47	92%	64%	0.79	0.26	2	23	--	--	--	0	7	14	--	--	--	1
4	3	TF	Pre	20	80%	40%	0.50	0.40	1	4 *	--	--	--	0	8	6 *	--	--	--	1
			Post	30	79%	50%	0.73	0.29	5	19	--	--	--	0	3	3	--	--	--	0
			Pre/Pst	50	79%	43%	0.64	0.39	6	23	--	--	--	0	11	9	--	--	--	1
4	4	TF	Pre	15	100%	78%	0.87	0.22	6 *	0	--	--	--	0	7 *	2	--	--	--	0
			Post	27	95%	100%	0.96	-0.05	21	1	--	--	--	0	5	0	--	--	--	0
			Pre/Pst	42	96%	86%	0.93	0.18	27	1	--	--	--	0	12	2	--	--	--	0
4	5	TF	Pre	13	25%	0%	0.08	0.25	1 *	3	--	--	--	0	0 *	8	--	--	--	1
			Post	37	31%	0%	0.22	0.31	8	18	--	--	--	0	0	11	--	--	--	0
			Pre/Pst	50	30%	0%	0.18	0.31	9	21	--	--	--	0	0	19	--	--	--	1
4	6	MC4	Pre	17	75%	31%	0.41	0.44	1	0	3 *	0	--	0	5	1	4 *	2	--	1
			Post	35	93%	50%	0.83	0.43	1	0	25	1	--	0	1	1	4	1	--	1
			Pre/Pst	52	90%	38%	0.69	0.62	2	0	28	1	--	0	6	2	8	3	--	2
5	1	TF	Pre	20	86%	69%	0.75	0.16	1	6 *	--	--	--	0	4	9 *	--	--	--	0
			Post	39	87%	56%	0.79	0.31	4	26	--	--	--	0	4	5	--	--	--	0
			Pre/Pst	59	86%	64%	0.78	0.17	5	32	--	--	--	0	8	14	--	--	--	0
5	2	TF	Pre	19	83%	46%	0.58	0.37	1	5 *	--	--	--	0	7	6 *	--	--	--	0
			Post	36	78%	33%	0.67	0.44	6	21	--	--	--	0	6	3	--	--	--	0
			Pre/Pst	55	79%	41%	0.64	0.32	7	26	--	--	--	0	13	9	--	--	--	0
5	3	TF	Pre	22	100%	81%	0.86	0.19	6 *	0	--	--	--	0	13 *	2	--	--	--	1
			Post	33	96%	86%	0.94	0.10	25	1	--	--	--	0	6	1	--	--	--	0
			Pre/Pst	55	97%	83%	0.91	0.15	31	1	--	--	--	0	19	3	--	--	--	1
5	4	TF	Pre	19	100%	81%	0.84	0.19	3 *	0	--	--	--	0	13 *	1	--	--	--	2
			Post	40	84%	88%	0.85	-0.03	27	5	--	--	--	0	7	1	--	--	--	0
			Pre/Pst	59	86%	83%	0.85	0.03	30	5	--	--	--	0	20	2	--	--	--	2
5	5	MC4	Pre	20	25%	25%	0.25	0.00	2	0	1	1 *	--	0	8	0	2	4 *	--	2
			Post	38	55%	14%	0.47	0.41	9	4	1	17	--	0	3	0	3	1	--	0
			Pre/Pst	58	51%	22%	0.40	0.30	11	4	2	18	--	0	11	0	5	5	--	2
6	1	MC4	Pre	21	100%	100%	1.00	0.00	0	0	4 *	0	--	0	0	0	17 *	0	--	0
			Post	33	100%	100%	1.00	0.00	0	0	26	0	--	0	0	0	7	0	--	0
			Pre/Pst	54	100%	100%	1.00	0.00	0	0	30	0	--	0	0	0	24	0	--	0

Figure 16 (continued)

Obj No.	Itm No.	Item Type	Test Type	No. of Admins	Percnt Corr Mast	Percnt Corr N-Mast	Diff. Index	Disc. Index	-- Number of Masters Responding -- T,Y A	F,N B	C	D	E	Skip	Number of Non-Masters Responding T,Y A	F,N B	C	D	E	Skip
6	2	TF	Pre	15	100%	56%	0.73	0.44	6 *	0	--	--	--	0	5 *	4	--	--	--	0
			Post	34	92%	67%	0.85	0.25	23	2	--	--	--	0	6	3	--	--	--	0
			Pre/Pst	49	94%	61%	0.82	0.36	29	2	--	--	--	0	11	7	--	--	--	0
6	3	MC4	Pre	16	100%	36%	0.56	0.64	5 *	0	0	0	--	0	4 *	1	1	5	--	0
			Post	32	92%	50%	0.81	0.42	22	0	0	2	--	0	4	0	0	4	--	0
			Pre/Pst	48	93%	42%	0.73	0.55	27	0	0	2	--	0	8	1	1	9	--	0
6	4	MC4	Pre	18	100%	43%	0.56	0.57	0	4 *	0	0	--	0	2	6 *	3	2	--	1
			Post	30	81%	0%	0.70	0.81	0	21	3	2	--	0	1	0	1	2	--	0
			Pre/Pst	48	83%	33%	0.65	0.38	0	25	3	2	--	0	3	6	4	4	--	1
6	5	TF	Pre	9	67%	50%	0.56	0.17	1	2 *	--	--	--	0	3	3 *	--	--	--	0
			Post	29	91%	67%	0.86	0.25	2	21	--	--	--	0	2	4	--	--	--	0
			Pre/Pst	38	88%	58%	0.79	0.41	3	23	--	--	--	0	5	7	--	--	--	0
6	6	TF	Pre	19	75%	67%	0.68	0.08	3 *	1	--	--	--	0	10 *	3	--	--	--	2
			Post	31	71%	43%	0.65	0.28	17	7	--	--	--	0	3	4	--	--	--	0
			Pre/Pst	50	71%	59%	0.66	0.04	20	8	--	--	--	0	13	7	--	--	--	2

7	1	MC4	Pre	10	100%	29%	0.50	0.71	0	0	3 *	0	--	0	1	3	2 *	1	--	0
			Post	9	86%	50%	0.78	0.36	0	1	6	0	--	0	0	1	1	0	--	0
			Pre/Pst	19	90%	33%	0.63	0.57	0	1	9	0	--	0	1	4	3	1	--	0
7	2	TF	Pre	9	67%	17%	0.33	0.50	1	2 *	--	--	--	0	5	1 *	--	--	--	0
			Post	16	73%	40%	0.63	0.33	3	8	--	--	--	0	3	2	--	--	--	0
			Pre/Pst	25	71%	27%	0.52	0.56	4	10	--	--	--	0	8	3	--	--	--	0
7	3	MC4	Pre	8	100%	57%	0.63	0.43	0	0	1 *	0	--	0	1	0	4 *	0	--	2
			Post	15	100%	33%	0.87	0.67	0	0	12	0	--	0	1	1	1	0	--	0
			Pre/Pst	23	100%	50%	0.78	0.43	0	0	13	0	--	0	2	1	5	0	--	2
7	4	MC4	Pre	9	83%	100%	0.89	-0.17	1	0	5 *	0	--	0	0	0	3 *	0	--	0
			Post	12	100%	100%	1.00	0.00	0	0	9	0	--	0	0	0	3	0	--	0
			Pre/Pst	21	93%	100%	0.95	0.00	1	0	14	0	--	0	0	0	6	0	--	0
7	5	MC4	Pre	7	100%	100%	1.00	0.00	0	3 *	0	0	--	0	0	4 *	0	0	--	0
			Post	14	100%	100%	1.00	0.00	0	12	0	0	--	0	0	2	0	0	--	0
			Pre/Pst	21	100%	100%	1.00	0.00	0	15	0	0	--	0	0	6	0	0	--	0
7	6	MC4	Pre	9	0%	78%	0.78	-0.78	0	0	0 *	0	--	0	0	0	7 *	1	--	1
			Post	17	100%	33%	0.88	0.67	0	0	14	0	--	0	0	2	1	0	--	0
			Pre/Pst	26	100%	67%	0.85	0.22	0	0	14	0	--	0	0	2	8	1	--	1
7	7	MC4	Pre	7	100%	83%	0.86	0.17	0	1 *	0	0	--	0	1	5 *	0	0	--	0
			Post	10	71%	33%	0.60	0.38	1	5	0	1	--	0	1	1	1	0	--	0
			Pre/Pst	17	75%	67%	0.71	-0.12	1	6	0	1	--	0	2	6	1	0	--	0

Figure 16 (continued)

Obj No.	Itm No.	Item Type	Test Type	No. of Admins	Percnt Corr Mast	Percnt Corr N-Mast	Diff. Index	Disc. Index	Masters T,Y A	Masters F,N B	Masters C	Masters D	Masters E	Masters Skip	Non-Masters T,Y A	Non-Masters F,N B	Non-Masters C	Non-Masters D	Non-Masters E	Non-Masters Skip
7	8	YN	Pre	9	Ø%	63%	Ø.56	-Ø.63	1	Ø *	--	--	--	Ø	2	5 *	--	--	--	1
			Post	16	79%	1ØØ%	Ø.81	-Ø.21	3	11	--	--	--	Ø	Ø	2	--	--	--	Ø
			Pre/Pst	25	73%	7Ø%	Ø.72	Ø.16	4	11	--	--	--	Ø	2	7	--	--	--	1
7	9	YN	Pre	8	67%	6Ø%	Ø.63	Ø.Ø7	1	2 *	--	--	--	Ø	2	3 *	--	--	--	Ø
			Post	17	73%	1ØØ%	Ø.76	-Ø.27	4	11	--	--	--	Ø	Ø	2	--	--	--	Ø
			Pre/Pst	25	72%	71%	Ø.72	Ø.13	5	13	--	--	--	Ø	2	5	--	--	--	Ø
7	1Ø	MC4	Pre	6	Ø%	83%	Ø.83	-Ø.83	Ø *	Ø	Ø	Ø	--	Ø	5 *	1	Ø	Ø	--	Ø
			Post	22	1ØØ%	71%	Ø.91	Ø.29	15	Ø	Ø	Ø	--	Ø	5	1	1	Ø	--	Ø
			Pre/Pst	28	1ØØ%	77%	Ø.89	Ø.17	15	Ø	Ø	Ø	--	Ø	1Ø	2	1	Ø	--	Ø
7	11	TF	Pre	8	67%	2Ø%	Ø.38	Ø.47	1	2 *	--	--	--	Ø	4	1 *	--	--	--	Ø
			Post	11	56%	Ø%	Ø.45	Ø.56	4	5	--	--	--	Ø	2	Ø	--	--	--	Ø
			Pre/Pst	19	58%	14%	Ø.42	Ø.36	5	7	--	--	--	Ø	6	1	--	--	--	Ø
7	12	TF	Pre	8	1ØØ%	1ØØ%	1.ØØ	Ø.ØØ	2 *	Ø	--	--	--	Ø	6 *	Ø	--	--	--	Ø
			Post	16	92%	1ØØ%	Ø.94	-Ø.Ø8	12	1	--	--	--	Ø	3	Ø	--	--	--	Ø
			Pre/Pst	24	93%	1ØØ%	Ø.96	-Ø.Ø8	14	1	--	--	--	Ø	9	Ø	--	--	--	Ø
7	13	TF	Pre	5	1ØØ%	33%	Ø.6Ø	Ø.67	Ø	2 *	--	--	--	Ø	2	1 *	--	--	--	Ø
			Post	14	78%	4Ø%	Ø.64	Ø.38	2	7	--	--	--	Ø	3	2	--	--	--	Ø
			Pre/Pst	19	82%	38%	Ø.63	Ø.44	2	9	--	--	--	Ø	5	3	--	--	--	Ø

=========

Number of items in Module 2 = 54

Figure 16 (continued)

types of misunderstandings that students had. For example, Item 1-2 was:

> Which PRINT statement will display the numeric 1024 when executed?
>
> A. 10 PRINT "1024"
> B. 10 PRINT 1024
> C. 10 PRINT "1"+"0"+"2"+"4"
> D. (Both A and B are correct.)

The correct answer for this item is B. However, 7 out of the 16 masters chose D. This may show that the instructional materials did not satisfactorily differentiate between numeric and string constants.

It is difficult to assess the validity of all 733 items in the item banks for Modules 2 through 16 using the item analysis data. The data collected on the Extended Modules (numbers 3, 5, 7, and 11 through 16, see Figure 7) was inconclusive due to the small number of test administrations for these modules. The data for the Core Modules is more extensive, but it is impossible to state categorically what discrimination index separates "good" items from "bad". The summary data on the 289 items in the Core Modules is therefore presented in Table 9 in ranges of their observed discrimination indices. This data indicates that a number of items in the banks of these modules warrant scrutiny and possible revision. The number of items in the lower discrimination ranges might have been decreased if additional

Table 9

OBSERVED RANGES OF DISCRIMINATION INDICES
FOR ITEMS IN THE CORE MODULES

N = 289

("di" denotes "discrimination index")

Range of the Discrimination Index	Number of Items	Percent of Total
-1.0 <= di <= -0.1	23	8.0 %
-0.1 < di <= +0.1	87	30.1
+0.1 < di <= +0.3	82	28.4
+0.3 < di <= +0.5	58	20.1
+0.5 < di <= +0.7	24	8.3
+0.7 < di <= +1.0	15	5.2

formative evaluation had been done before the banks were released to the test sites.

Chapter 5
DISCUSSION

Summary of Results

The data presented in Chapter 4 have shown that this study was successful in designing, developing, and implementing a CMI system that could be installed by customers and that could generate easily transportable usage data for analysis. Due to the nature of the student population, the study was not able to show conclusively that the CMI system was easily usable by completely computer-naive learners.

Test history data showed that the system was successful in forcing users to work through the module tests in the prescribed hierarchy and in discouraging excessive test taking. Test length and reliability data demonstrated that a mastery algorithm based on a sequential probability test ratio was successful in reducing test lengths significantly (on all tests except pretests on which mastery decisions were made) without sacrificing reliability.

General Comments

The customer field testing of a piece of computer software is often a humbling experience for the software developer because customers invariably uncover bugs and oversights that require extensive reprogramming. The first installation of The BASIC Primer was just such an experience, but the author was able to "patch" the software in a matter of hours at the user site. All of the required corrections were made in a matter of days and delivery of the revised software to the other two installations was made on schedule. Once all three sites were up and running, only one programming bug that affected users was uncovered during the entire duration of the study (November 1978 through January 1979). Thus, it may be said that considering the number of possible problems that could have arisen, the study itself went very smoothly.

A posteriori analysis of the data stored by the system showed that more *a priori* analysis should have been done to improve the design of the system's data base. For example, the exact numbers of extended pretests and extended posttests should have been stored so that the ratio calculation in computing expected cell frequencies could have been eliminated. In addition, this data would have simplified or eliminated the "correction" process required to analyze the test length data.

Even with the CMI system, there was little control over student motivation in the self-paced environment. During the period of this study, 31% of the students who began the course completed the core modules. (Some of the learners had just started the course when the final data was collected, but others had definitely been working on it for some time.) Thus, even though the CMI system was instrumental in assuring that learning does take place for students who choose to study the course, the question of motivation was still largely unaddressed by this study. For self-paced training to be truly effective, this question must be addressed.

Another uncontrolled factor was the order in which students studied the self-paced modules. The CMI system only controlled the order in which students were tested on the modules. Additional data might have been collected to ascertain the sequence in which students studied, perhaps in the form of an on-line questionnaire.

Directions for Future Study

The four sequential probability test parameters (P_0, P_1, $\underline{a}$, and $\underline{b}$) used in this study were fixed for both pretests and posttests. Test length data indicates that these data may have been a bit too stringent, especially for pretests on which mastery decisions were made, which had a median length of 29 items. Further research should be conducted into varying the parameters so that the balance between test length and reliability can be optimized.

As explained in Chapter 3, P_0 and P_1 were set at different values for pretests and posttests. This introduced an unnecessary complication in analyzing the test length data because it made it difficult to ascertain the amount of the differences in median test lengths that was attributable to differences in error parameters alone. A further study might be done in which P_0 and P_1 are held constant on the two tests and only $\underline{a}$ and $\underline{b}$ are changed.

Since the test item banks were not precalibrated before they were used at the test sites, it was impossible to weight the items based on observed difficulty indices. A weighting algorithm based on item type was therefore implemented (cf. page 62). This algorithm has been criticized on the grounds that 4-alternative multiple choice items may not prove twice as difficult as true/false or yes/no items. Further analysis of this criticism is warranted, including the more global question of using several types of items in a single sequential probability test.

Another issue related to weighting is the varying importance of objectives. In the current study, the importance of each objective (and therefore the probability of selecting an item for that objective) was equal. A further study might modify the existing software to allow objectives, as well as items to be weighted.

Due to the timing of the study and the uniqueness of The BASIC Primer as a product of Digital Equipment Corporation, the training package was not used as part of any formal course. It was used mainly for teacher training. While its value in this type of environment was established, the application of the CMI technique in a more structured course environment could not be evaluated. Most of the studies quoted in Chapter 2 were conducted in formal course environments, and further research should be conducted using the CMI system designed for this study to determine whether course structure is a factor in its applicability.

Appendix A

DIRECTIONS TO STUDENTS ON USING THE CMI SYSTEM

The pages that follow are an excerpt from Module 1 of *The BASIC Primer*, the Student Guide. These pages guided students through learning how to use the CMI system. (This excerpt is copyrighted by Digital Equipment Corporation.)

NOTE: Please read the entire exercise BEFORE going to the terminal.

Exercise 1B: Using the CMI System

1. Log in to your computer system.

2. Type RUN CMI <ret>.

3. The computer should generate the display shown on the facing page. If it does not, see your course administrator for assistance.

4. The first question asks you to identify yourself. Since this is your first time on the CMI system, you have not yet selected a code name. Therefore, just press the RETURN key alone as the computer requests.

5. The system will now ask you several questions to register you for The BASIC Primer course. Read and answer these questions. (NOTE: You may type in either upper or lower case).

6. When the program has asked all of its registration questions, it will tell you that it has created a file containing your registration data. Ask your course administrator to show you how to print this file on paper. Mail the resultant output to:

   Educational Services
   Computer-Based Course Development Group
   DIGITAL EQUIPMENT CORPORATION
   Crosby Drive
   Bedford, MA 01730

7. After you have printed your registration data and put it in an envelope for mailing, type RUN CMI <ret> to run the CMI system again.

8. This time, when the system asks you to identify yourself, type your code name. (Don't forget to press RETURN.)

9. The system should recognize your code name, ask one or two additional questions, and tell you that you are registered for this terminal session.

10. The system should then ask if you would like to see your status on each of the modules.

11. Type Y <ret>. The system will print your name, your course, and your status on each of the 16 modules. Your status should show that you have not attempted any of the modules.

12. The system will then ask if you are ready for the pretest for Module 1. Type Y <ret>.

13. When it asks you if you have studied the material for this module yet, type N <ret>.

14. The system will identify the module to be tested and print the first test item. This may be a true/false item, a yes/no item, or a multiple choice item. The display contains three sections, as shown on the facing page.

15. Type the correct answer that is indicated in the first sentence of the test item on your terminal and press RETURN. The system will say "Correct" and go on to the next item.

16. The directions at the top of your screen may change, and the REVIEW option will be added to the list of options in Section 2 of the display. Before trying out those options, type an incorrect answer for the item on your terminal and press RETURN. The system will say "Incorrect" and go on to the next test item.

17. Now let's look at the system's other features. First, type a response that the system does not expect, for example, J <ret>. The system will print a message and ask you to enter another answer. You therefore do not have to worry about small typing mistakes, they should not hurt your score.

18. Type SKIP <ret>. The system will print "OK" and go on to the next test item. This option will save you time in pretests, when you may see a number of test items which you cannot answer.

19. Type REVIEW <ret>. As shown on the facing page, the system will display the previous item, what you answered, its judgement of that answer, and your current score. You can continue typing REVIEW to step back through the entire test. Try this, and then press RETURN to get back to the item on which

you first typed REVIEW.

20. The only option left is QUIT. Type QUIT <ret>. The system will print "OK", pause, and then print a final message. When the printing stops, the CMI system has terminated, and returned you to the BASIC operating system.

21. Type RUN CMI <ret> to run the CMI system once again.

22. Identify yourself with your code name, answer any other questions that may be presented, and then ask the system to display the status on your work.

23. Your status should show that you tried the pretest for Module 1, but that you did not complete it satisfactorily. You must therefore take the posttest for this module, so type Y <ret> for this test.

24. The first item of the posttest will now be displayed in the same format as the pretest. The same options are also available. Generally, the CMI system presents between 3 and 30 items on a test. Simply keep answering the items until the system tells you that you have or have not mastered the module being tested.

> On this particular test you can type OK <ret> to shortcut the testing procedure. This will cause the system to tell you that you have mastered Module 1 and can go on to Module 2. This only works on the posttest for Module 1.

25. Type OK <ret>. The system will print several messages. Display your status on the modules and you will see that you have been given credit for completing Module 1 satisfactorily.

26. The system will tell you that you have met the prerequisites for Module 2. If you are now ready to take this test, answer Y to the first question. If you type N, the system will terminate the terminal session and you can log off the system and try it some other time. When you log back on, the system will remember that you are ready for a test on Module 2. It will not help you to guess on the test items -- use the SKIP option instead. After you take each test, follow the system's directions.

Appendix B

TEST HISTORY DATA FOR WACHUSETT REGIONAL SCHOOL DISTRICT AND RHODE ISLAND JUNIOR COLLEGE

The pages that follow contain computer listings of the test history data for Wachusett Regional School District and Rhode Island Junior College. The data for Falmouth High School in Figure 14.

Data for WACHUSETT REG'L. SCHOOL DIST.   02-Feb-79, page 19

TEST HISTORY DATA

STUDENT 1

Mod 1	Normal	Pretest	10-Nov-78	02:09	PM
	Normal	Posttest	10-Nov-78	02:49	PM
	Normal	Posttest	10-Nov-78	02:55	PM
	Normal	Pretest	24-Nov-78	11:15	AM

STUDENT 2

Mod 1	Normal	Posttest	21-Nov-78	01:13	PM
Mod 2	Normal	Pretest	21-Nov-78	01:18	PM
	Normal	Posttest	22-Nov-78	10:05	AM
Mod 3	Extended	Posttest	22-Nov-78	10:11	AM
Mod 4	Normal	Pretest	22-Nov-78	10:25	AM
Mod 5	Normal	Posttest	30-Nov-78	01:59	PM
Mod 6	Normal	Posttest	30-Nov-78	02:04	PM
Mod 7	Normal	Pretest	01-Dec-78	12:17	PM
Mod 8	Extended	Posttest	02-Jan-79	01:57	PM

STUDENT 3

Mod 1	Normal	Pretest	21-Nov-78	10:11	AM
	Normal	Posttest	21-Nov-78	10:28	AM
Mod 2	Extended	Posttest	21-Nov-78	01:36	PM
Mod 3	Normal	Posttest	22-Nov-78	09:36	AM
Mod 4	Normal	Pretest	22-Nov-78	09:48	AM
	Normal	Posttest	27-Nov-78	11:16	AM
Mod 5	Normal	Pretest	27-Nov-78	11:25	AM
Mod 6	Extended	Posttest	30-Nov-78	05:38	PM
	Normal	Posttest	05-Dec-78	10:55	AM

Data for WACHUSETT REG'L. SCHOOL DIST. 02-Feb-79, page 20

STUDENT 4

Mod 1	Normal	Pretest	20-Nov-78	03:20	PM	
Mod 2	Extended	Posttest	20-Nov-78	03:40	PM	
Mod 3	Normal	Posttest	20-Nov-78	03:49	PM	
Mod 4	Normal	Posttest	20-Nov-78	03:56	PM	
	Normal	Posttest	21-Nov-78	04:28	PM	
Mod 5	Extended	Pretest	22-Nov-78	09:49	AM	
	Normal	Posttest	22-Nov-78	09:51	AM	D = 2 min
Mod 6	Normal	Posttest	22-Nov-78	09:59	AM	
Mod 7	Normal	Pretest	27-Nov-78	09:53	AM	
Mod 8	Normal	Pretest	27-Nov-78	10:03	AM	
Mod 9	Extended	Posttest	27-Nov-78	02:41	PM	
Mod 10	Normal	Pretest	27-Nov-78	03:13	PM	
	Normal	Posttest	28-Nov-78	09:45	AM	
	Normal	Posttest	28-Nov-78	09:47	AM	D = 2 min
	Normal	Posttest	28-Nov-78	02:30	PM	
	Extended	Posttest	28-Nov-78	02:31	PM	D = 1 min
	Normal	Posttest	28-Nov-78	02:33	PM	D = 2 min
	Normal	Posttest	28-Nov-78	02:38	PM	D = 5 min
	Extended	Posttest	28-Nov-78	02:57	PM	D = 19 min
	Extended	Posttest	28-Nov-78	03:06	PM	D = 9 min
	Extended	Posttest	05-Dec-78	10:29	AM	
	Extended	Posttest	05-Dec-78	05:08	PM	
	Extended	Posttest	06-Dec-78	10:35	AM	
	Extended	Posttest	06-Dec-78	04:10	PM	
Mod 11	Extended	Pretest	06-Dec-78	06:52	PM	
	Normal	Posttest	07-Dec-78	01:35	PM	
Mod 12	Normal	Pretest	07-Dec-78	04:04	PM	
Mod 11	Normal	Posttest	08-Dec-78	10:17	AM	
Mod 12	Normal	Posttest	08-Dec-78	02:37	PM	
Mod 13	Extended	Pretest	08-Dec-78	05:23	PM	
	Normal	Posttest	08-Dec-78	07:26	PM	

Data for WACHUSETT REG'L. SCHOOL DIST.   02-Feb-79, page 21

Mod 14	Normal	Pretest	08-Dec-78	07:29 PM	
Mod 15	Normal	Pretest	08-Dec-78	07:41 PM	
Mod 16	Normal	Pretest	08-Dec-78	07:53 PM	

STUDENT 5

Mod 1	Extended	Pretest	21-Nov-78	10:35 AM	
	Normal	Posttest	21-Nov-78	10:46 AM	
Mod 2	Normal	Pretest	29-Nov-78	01:53 PM	
	Normal	Posttest	06-Dec-78	02:26 PM	
Mod 4	Normal	Pretest	06-Dec-78	02:37 PM	

STUDENT 6

STUDENT 7

Mod 1	Normal	Pretest	22-Nov-78	12:45 PM	
Mod 2	Normal	Pretest	22-Nov-78	12:51 PM	
Mod 4	Normal	Pretest	27-Nov-78	09:38 AM	
	Extended	Posttest	27-Nov-78	11:20 AM	
	Normal	Posttest	28-Nov-78	09:42 AM	
Mod 3	Normal	Pretest	28-Nov-78	09:54 AM	
	Normal	Posttest	28-Nov-78	11:34 AM	
Mod 5	Normal	Pretest	28-Nov-78	11:36 AM	
Mod 6	Extended	Pretest	29-Nov-78	11:34 AM	
	Normal	Posttest	29-Nov-78	02:29 PM	
Mod 3	Normal	Posttest	30-Nov-78	09:54 AM	
	Normal	Posttest	30-Nov-78	10:00 AM	D = 6 min
Mod 5	Normal	Posttest	30-Nov-78	10:05 AM	
Mod 3	Extended	Posttest	01-Dec-78	05:53 PM	
	Normal	Posttest	06-Dec-78	05:04 PM	
	Normal	Posttest	06-Dec-78	05:29 PM	D = 25 min
	Normal	Posttest	06-Dec-78	05:31 PM	D = 2 min

Mod	5	Normal	Posttest	06-Dec-78	05:36	PM	
Mod	7	Extended	Pretest	07-Dec-78	06:47	PM	
		Normal	Posttest	07-Dec-78	06:59	PM	D = 12 min
Mod	5	Normal	Posttest	07-Dec-78	07:08	PM	
		Normal	Posttest	08-Dec-78	07:47	PM	
		Normal	Posttest	12-Dec-78	05:28	PM	
		Normal	Posttest	12-Dec-78	05:50	PM	D = 22 min
		Normal	Posttest	12-Dec-78	05:51	PM	D = 1 min
Mod	8	Normal	Pretest	12-Dec-78	05:57	PM	
		Extended	Posttest	12-Dec-78	07:11	PM	
Mod	9	Normal	Pretest	13-Dec-78	01:39	PM	
		Normal	Posttest	15-Dec-78	07:46	PM	

STUDENT 8

Mod	1	Normal	Pretest	22-Nov-78	12:17	PM
Mod	2	Normal	Pretest	22-Nov-78	12:27	PM

STUDENT 9

Mod	1	Normal	Pretest	07-Dec-78	04:18	PM
Mod	2	Extended	Pretest	07-Dec-78	04:26	PM
Mod	3	Normal	Pretest	07-Dec-78	04:44	PM
Mod	4	Normal	Pretest	07-Dec-78	05:31	PM

STUDENT 11

Mod	1	Normal	Pretest	08-Jan-79	02:41	PM
Mod	2	Normal	Pretest	08-Jan-79	02:55	PM

Data for WACHUSETT REG'L. SCHOOL DIST. 02-Feb-79, page 23

SUMMARY

Number of Normal Pretests =	32	
Extended Pretests =	7	
Total Number of Pretests =		39 (37.5%)
Number of Normal Posttests =	43	
Extended Posttests =	16	
Total Number of Posttests =		59 (56.7%)

Total Number of Tests Taken =		104
Instances of Repeated Tests =	35	
No. of Elapsed Times < 15 m =	10 (28.6%)	
No. of Elapsed Times < 30 m =	13 (37.1%)	

TEST HISTORY DATA

STUDENT 1

Mod 1	Normal	Pretest	15-Nov-78	12:20	PM	
	Normal	Posttest	15-Nov-78	12:32	PM	
Mod 2	Normal	Pretest	15-Nov-78	12:36	PM	
	Normal	Posttest	29-Nov-78	08:52	AM	
Mod 4	Extended	Pretest	29-Nov-78	09:33	AM	
	Normal	Posttest	29-Nov-78	01:26	PM	
	Normal	Posttest	12-Dec-78	08:30	AM	
Mod 6	Normal	Pretest	12-Dec-78	10:03	AM	
	Normal	Posttest	12-Dec-78	11:00	AM	
Mod 8	Normal	Pretest	12-Dec-78	11:30	AM	
Mod 9	Normal	Pretest	12-Dec-78	01:13	PM	
Mod 10	Normal	Pretest	12-Dec-78	03:28	PM	
	Extended	Posttest	13-Dec-78	09:20	AM	

STUDENT 2

Mod 1	Normal	Pretest	15-Nov-78	12:28	PM	
	Normal	Posttest	15-Nov-78	12:48	PM	
	Normal	Pretest	17-Nov-78	10:27	AM	
	Normal	Posttest	17-Nov-78	10:38	AM	
Mod 2	Normal	Pretest	17-Nov-78	10:40	AM	
	Extended	Posttest	03-Dec-78	10:01	AM	
Mod 4	Normal	Posttest	03-Dec-78	10:14	AM	
Mod 6	Normal	Posttest	03-Dec-78	10:23	AM	
	Normal	Posttest	03-Dec-78	10:40	AM	D = 17 min
	Normal	Posttest	03-Dec-78	10:56	AM	D = 16 min
Mod 8	Extended	Posttest	03-Dec-78	01:01	PM	
Mod 9	Normal	Posttest	03-Dec-78	01:26	PM	

Mod 10	Normal	Posttest	03-Dec-78	02:01	PM	

STUDENT 3

Mod 1	Normal	Pretest	22-Nov-78	03:36	PM	
Mod 2	Normal	Posttest	22-Nov-78	03:50	PM	
Mod 3	Extended	Pretest	22-Nov-78	04:07	PM	
	Normal	Posttest	24-Nov-78	08:52	AM	
Mod 4	Normal	Pretest	24-Nov-78	09:07	AM	
Mod 5	Normal	Pretest	24-Nov-78	10:25	AM	
Mod 6	Extended	Pretest	24-Nov-78	10:46	AM	
Mod 8	Normal	Pretest	27-Nov-78	03:58	PM	
Mod 9	Normal	Pretest	27-Nov-78	04:20	PM	
Mod 10	Normal	Pretest	28-Nov-78	01:44	PM	
	Extended	Posttest	28-Nov-78	01:52	PM	D = 8 min
	Normal	Posttest	28-Nov-78	01:58	PM	D = 6 min
	Normal	Posttest	28-Nov-78	02:01	PM	D = 3 min
	Normal	Posttest	29-Nov-78	05:29	PM	
Mod 5	Extended	Posttest	30-Nov-78	05:52	PM	
Mod 7	Normal	Pretest	30-Nov-78	06:20	PM	
Mod 11	Normal	Pretest	30-Nov-78	06:53	PM	
	Normal	Posttest	30-Nov-78	07:14	PM	D = 21 min
Mod 12	Normal	Pretest	30-Nov-78	07:46	PM	
	Extended	Posttest	01-Dec-78	08:01	AM	
	Normal	Posttest	01-Dec-78	10:14	AM	
Mod 13	Normal	Pretest	01-Dec-78	10:35	AM	
	Normal	Posttest	01-Dec-78	04:59	PM	
Mod 14	Normal	Posttest	01-Dec-78	05:42	PM	
Mod 15	Extended	Posttest	01-Dec-78	06:08	PM	
Mod 16	Normal	Posttest	01-Dec-78	06:48	PM	

STUDENT 4

Mod 1	Normal	Pretest	28-Nov-78	02:52	PM	
Mod 2	Extended	Posttest	02-Dec-78	09:19	AM	
Mod 4	Normal	Pretest	02-Dec-78	09:43	AM	
	Normal	Posttest	02-Dec-78	10:21	AM	
Mod 3	Normal	Pretest	02-Dec-78	10:42	AM	
Mod 4	Normal	Posttest	02-Dec-78	10:43	AM	
Mod 3	Extended	Posttest	02-Dec-78	10:51	AM	
Mod 6	Normal	Posttest	06-Dec-78	04:40	PM	
	Normal	Posttest	06-Dec-78	04:44	PM	D = 4 min
	Normal	Posttest	06-Dec-78	05:11	PM	D = 27 min
Mod 8	Normal	Posttest	08-Dec-78	08:23	AM	
Mod 3	Normal	Posttest	08-Dec-78	08:34	AM	
Mod 5	Normal	Pretest	08-Dec-78	08:44	AM	
Mod 7	Normal	Pretest	08-Dec-78	08:48	AM	
	Normal	Posttest	08-Dec-78	12:55	PM	
Mod 9	Normal	Pretest	08-Dec-78	01:02	PM	
Mod 5	Extended	Posttest	08-Dec-78	01:09	PM	
	Normal	Posttest	08-Dec-78	01:13	PM	D = 4 min
Mod 9	Normal	Posttest	14-Dec-78	11:15	AM	
Mod 10	Normal	Pretest	14-Dec-78	11:30	AM	
	Normal	Posttest	14-Dec-78	11:34	AM	D = 4 min
	Extended	Posttest	14-Dec-78	11:47	AM	D = 13 min
	Normal	Posttest	14-Dec-78	11:55	AM	D = 8 min
	Normal	Posttest	16-Dec-78	10:12	AM	
Mod 16	Normal	Posttest	12-Jan-79	11:40	AM	

Data for RHODE ISLAND JUNIOR COLLEGE 02-Feb-79, page 45

STUDENT 5

Mod 1	Extended	Pretest	28-Nov-78	03:01	PM	
Mod 2	Normal	Pretest	28-Nov-78	03:16	PM	
	Normal	Posttest	04-Dec-78	05:48	PM	
Mod 3	Normal	Pretest	04-Dec-78	05:57	PM	
Mod 4	Normal	Pretest	04-Dec-78	06:03	PM	
	Extended	Posttest	05-Dec-78	04:13	PM	
	Normal	Posttest	05-Dec-78	04:31	PM	D = 18 min
	Normal	Posttest	05-Dec-78	04:38	PM	D = 7 min
	Normal	Posttest	05-Dec-78	04:48	PM	D = 10 min
	Normal	Posttest	05-Dec-78	05:05	PM	D = 17 min
	Normal	Posttest	05-Dec-78	05:07	PM	D = 2 min
	Normal	Posttest	05-Dec-78	05:16	PM	D = 9 min
Mod 6	Normal	Pretest	05-Dec-78	05:32	PM	
	Normal	Posttest	06-Dec-78	07:35	AM	
	Normal	Posttest	06-Dec-78	08:06	AM	
Mod 8	Extended	Pretest	06-Dec-78	05:53	PM	
	Normal	Posttest	06-Dec-78	06:31	PM	
Mod 7	Normal	Pretest	06-Dec-78	06:44	PM	
Mod 5	Normal	Pretest	06-Dec-78	06:47	PM	
Mod 9	Normal	Pretest	06-Dec-78	06:50	PM	
	Extended	Posttest	08-Dec-78	07:30	AM	
	Normal	Posttest	08-Dec-78	01:27	PM	
	Normal	Posttest	13-Dec-78	04:02	PM	
	Normal	Posttest	13-Dec-78	04:04	PM	D = 2 min
Mod 10	Normal	Pretest	13-Dec-78	04:31	PM	
	Extended	Posttest	14-Dec-78	05:08	PM	
Mod 3	Normal	Posttest	16-Dec-78	09:13	AM	
Mod 7	Normal	Posttest	19-Dec-78	07:41	AM	
Mod 15	Normal	Pretest	19-Dec-78	12:48	PM	
Mod 14	Normal	Pretest	19-Dec-78	12:55	PM	

Mod 11	Extended	Pretest	19-Dec-78	01:00 PM
Mod 12	Normal	Pretest	19-Dec-78	01:03 PM
Mod 13	Normal	Pretest	19-Dec-78	01:06 PM
Mod 15	Normal	Posttest	19-Dec-78	03:36 PM

STUDENT 6

Mod 1	Normal	Pretest	28-Nov-78	03:43 PM
Mod 2	Normal	Pretest	29-Nov-78	01:13 PM
	Normal	Posttest	01-Dec-78	02:39 PM
Mod 4	Normal	Pretest	01-Dec-78	03:03 PM

STUDENT 7

Mod 1	Normal	Pretest	29-Nov-78	12:01 AM
	Normal	Posttest	29-Nov-78	12:18 AM
Mod 2	Normal	Pretest	01-Dec-78	11:10 PM
	Extended	Posttest	05-Dec-78	11:41 PM
Mod 4	Normal	Pretest	06-Dec-78	12:07 AM
	Normal	Posttest	06-Dec-78	11:27 PM
Mod 6	Normal	Pretest	07-Dec-78	12:02 AM
	Normal	Posttest	07-Dec-78	11:33 PM
Mod 8	Extended	Pretest	08-Dec-78	12:04 AM
	Normal	Posttest	11-Dec-78	10:44 PM
Mod 9	Normal	Pretest	11-Dec-78	11:06 PM
	Normal	Posttest	12-Dec-78	11:04 PM
	Normal	Posttest	13-Dec-78	09:21 PM
	Extended	Posttest	13-Dec-78	10:56 PM
Mod 10	Normal	Pretest	13-Dec-78	11:47 PM
	Normal	Posttest	14-Dec-78	10:47 PM

STUDENT 8

Mod 1	Normal	Pretest	29-Nov-78	10:43	AM	
	Normal	Posttest	29-Nov-78	10:48	AM	
Mod 2	Extended	Pretest	29-Nov-78	10:51	AM	
Mod 3	Normal	Pretest	29-Nov-78	11:01	AM	
	Normal	Posttest	29-Nov-78	12:53	PM	
	Normal	Posttest	29-Nov-78	01:01	PM	D = 8 min
Mod 4	Normal	Pretest	29-Nov-78	01:25	PM	
Mod 6	Extended	Pretest	29-Nov-78	01:33	PM	
Mod 8	Normal	Pretest	29-Nov-78	01:53	PM	
Mod 9	Normal	Pretest	29-Nov-78	02:07	PM	
Mod 10	Normal	Pretest	29-Nov-78	02:39	PM	
Mod 5	Normal	Posttest	05-Dec-78	02:05	PM	
Mod 7	Extended	Posttest	05-Dec-78	02:17	PM	
Mod 11	Normal	Posttest	05-Dec-78	02:30	PM	
Mod 12	Normal	Posttest	05-Dec-78	02:44	PM	
Mod 13	Normal	Posttest	05-Dec-78	02:55	PM	
Mod 15	Normal	Posttest	05-Dec-78	03:05	PM	
Mod 14	Extended	Posttest	05-Dec-78	03:15	PM	
Mod 16	Normal	Posttest	06-Dec-78	11:12	AM	

STUDENT 9

Mod 1	Normal	Posttest	29-Nov-78	11:21	AM
Mod 2	Extended	Posttest	29-Nov-78	11:22	AM
Mod 4	Normal	Posttest	29-Nov-78	11:32	AM

Mod	5	Normal	Posttest	29-Nov-78	11:46 AM
Mod	6	Normal	Posttest	29-Nov-78	12:05 PM
Mod	8	Normal	Posttest	29-Nov-78	08:51 PM
Mod	9	Extended	Posttest	29-Nov-78	09:01 PM
Mod	10	Normal	Posttest	29-Nov-78	09:15 PM
Mod	3	Normal	Posttest	04-Dec-78	03:45 PM
Mod	7	Normal	Posttest	04-Dec-78	03:55 PM
Mod	11	Normal	Posttest	04-Dec-78	04:02 PM
Mod	12	Extended	Posttest	04-Dec-78	04:10 PM
Mod	13	Normal	Posttest	05-Dec-78	12:18 PM
Mod	14	Normal	Posttest	05-Dec-78	12:44 PM
Mod	15	Normal	Posttest	05-Dec-78	12:55 PM
Mod	14	Normal	Posttest	05-Dec-78	03:57 PM
Mod	16	Extended	Posttest	05-Dec-78	04:12 PM

STUDENT 10

Mod	1	Extended	Pretest	29-Nov-78	11:33 AM
		Normal	Posttest	29-Nov-78	11:42 AM
Mod	2	Normal	Posttest	30-Nov-78	11:17 AM
Mod	3	Normal	Posttest	30-Nov-78	11:26 AM
Mod	4	Normal	Posttest	30-Nov-78	11:31 AM
		Extended	Posttest	30-Nov-78	12:39 PM
Mod	6	Normal	Pretest	01-Dec-78	11:26 AM
		Normal	Posttest	04-Dec-78	03:12 PM
		Normal	Posttest	04-Dec-78	03:54 PM
Mod	3	Normal	Posttest	04-Dec-78	04:28 PM

Mod	6	Extended	Posttest	05-Dec-78	12:48	PM	
Mod	8	Normal	Pretest	05-Dec-78	01:13	PM	
		Normal	Posttest	05-Dec-78	01:24	PM	D = 11 min
Mod	9	Normal	Pretest	07-Dec-78	12:06	PM	
Mod	10	Normal	Pretest	07-Dec-78	12:53	PM	
		Extended	Posttest	08-Dec-78	11:11	AM	
Mod	5	Normal	Posttest	11-Dec-78	03:01	PM	
Mod	7	Normal	Posttest	11-Dec-78	03:14	PM	
Mod	11	Normal	Posttest	12-Dec-78	11:14	AM	
Mod	12	Normal	Posttest	12-Dec-78	04:06	PM	
		Extended	Posttest	18-Dec-78	06:07	PM	
Mod	13	Normal	Pretest	20-Dec-78	11:01	AM	
		Normal	Posttest	20-Dec-78	11:13	AM	D = 12 min

STUDENT 11

Mod	1	Normal	Pretest	30-Nov-78	12:56	PM
Mod	2	Normal	Pretest	04-Dec-78	10:52	AM

STUDENT 12

Mod	1	Normal	Pretest	29-Nov-78	02:07	PM	
Mod	2	Normal	Pretest	01-Dec-78	02:46	PM	
Mod	4	Normal	Pretest	06-Dec-78	01:14	PM	
		Extended	Posttest	06-Dec-78	02:11	PM	
Mod	6	Normal	Pretest	07-Dec-78	01:06	PM	
		Normal	Posttest	08-Dec-78	12:58	PM	
		Normal	Posttest	08-Dec-78	01:10	PM	D = 12 min
Mod	8	Normal	Pretest	11-Dec-78	01:54	PM	
		Extended	Posttest	11-Dec-78	02:22	PM	D = 28 min
Mod	9	Normal	Pretest	13-Dec-78	10:00	AM	
		Normal	Posttest	13-Dec-78	02:31	PM	

Mod 10	Normal	Pretest	14-Dec-78	01:55	PM		
Mod 3	Normal	Pretest	15-Dec-78	01:04	PM		
Mod 5	Extended	Pretest	19-Dec-78	01:34	PM		
Mod 7	Normal	Pretest	21-Dec-78	01:00	PM		
Mod 12	Normal	Posttest	12-Jan-79	10:05	AM		
	Normal	Posttest	12-Jan-79	10:44	AM		
Mod 13	Normal	Posttest	17-Jan-79	08:32	AM		
Mod 14	Extended	Posttest	18-Jan-79	09:03	AM		

STUDENT 13

Mod 1	Normal	Pretest	29-Nov-78	03:41	PM		
	Normal	Posttest	29-Nov-78	03:48	PM		
Mod 2	Extended	Pretest	29-Nov-78	03:50	PM		
	Normal	Posttest	01-Dec-78	02:27	PM		
	Normal	Posttest	01-Dec-78	02:42	PM	D = 15 min	
Mod 4	Normal	Pretest	01-Dec-78	02:55	PM		
	Extended	Posttest	05-Dec-78	03:20	PM		
	Normal	Posttest	05-Dec-78	03:42	PM	D = 22 min	
	Normal	Posttest	05-Dec-78	03:45	PM	D = 3 min	
	Normal	Posttest	05-Dec-78	03:46	PM	D = 1 min	
	Normal	Posttest	06-Dec-78	02:26	PM		
	Normal	Posttest	06-Dec-78	02:33	PM	D = 7 min	
	Normal	Posttest	06-Dec-78	02:47	PM	D = 14 min	
	Normal	Posttest	06-Dec-78	02:49	PM	D = 2 min	
	Normal	Posttest	06-Dec-78	03:05	PM	D = 16 min	
	Normal	Posttest	06-Dec-78	03:20	PM	D = 15 min	
	Normal	Posttest	06-Dec-78	03:27	PM	D = 7 min	
	Normal	Posttest	11-Dec-78	02:31	PM		
Mod 6	Normal	Pretest	11-Dec-78	02:58	PM		
	Normal	Posttest	12-Dec-78	02:28	PM		
Mod 8	Normal	Pretest	12-Dec-78	02:54	PM		
Mod 9	Normal	Pretest	12-Dec-78	03:19	PM		
	Normal	Posttest	13-Dec-78	03:22	PM		
	Normal	Posttest	13-Dec-78	03:23	PM	D = 1 min	
	Normal	Posttest	13-Dec-78	03:34	PM	D = 11 min	

	Normal	Posttest	13-Dec-78	03:36 PM	D = 2 min
	Normal	Posttest	13-Dec-78	03:40 PM	D = 4 min
Mod 10	Normal	Pretest	14-Dec-78	02:27 PM	
	Normal	Posttest	15-Dec-78	01:41 PM	

STUDENT 14

Mod 1	Normal	Pretest	30-Nov-78	04:19 PM
Mod 2	Extended	Pretest	30-Nov-78	09:41 PM
	Normal	Posttest	30-Nov-78	10:28 PM

STUDENT 15

Mod 1	Extended	Pretest	01-Dec-78	03:18 PM
	Normal	Posttest	01-Dec-78	03:24 PM
Mod 2	Normal	Pretest	01-Dec-78	03:26 PM
	Normal	Posttest	08-Dec-78	09:13 AM
	Normal	Posttest	14-Dec-78	10:22 AM
Mod 3	Extended	Pretest	14-Dec-78	10:50 AM

STUDENT 16

Mod 1	Normal	Pretest	04-Dec-78	01:36 PM
	Normal	Posttest	04-Dec-78	01:42 PM
Mod 2	Normal	Pretest	04-Dec-78	01:45 PM
	Normal	Posttest	05-Dec-78	09:12 AM
Mod 4	Extended	Pretest	05-Dec-78	09:18 AM
Mod 3	Normal	Pretest	05-Dec-78	09:36 AM
Mod 5	Normal	Pretest	05-Dec-78	12:55 PM
Mod 6	Normal	Pretest	05-Dec-78	12:59 PM
Mod 8	Normal	Pretest	05-Dec-78	01:07 PM
Mod 9	Extended	Pretest	05-Dec-78	01:17 PM
Mod 7	Normal	Pretest	05-Dec-78	01:33 PM

Mod 9	Normal	Posttest	06-Dec-78	09:33	AM		
Mod 10	Normal	Posttest	06-Dec-78	09:37	AM		
Mod 11	Extended	Pretest	06-Dec-78	09:53	AM		
Mod 12	Normal	Pretest	07-Dec-78	11:41	AM		
Mod 5	Normal	Posttest	13-Dec-78	08:50	AM		
Mod 7	Normal	Posttest	13-Dec-78	09:03	AM		
Mod 11	Normal	Posttest	13-Dec-78	09:15	AM		
	Normal	Posttest	14-Dec-78	01:56	PM		
	Normal	Posttest	14-Dec-78	02:00	PM	D =	4 min
Mod 12	Normal	Posttest	18-Dec-78	04:20	PM		
Mod 13	Normal	Pretest	18-Dec-78	04:31	PM		
	Normal	Posttest	19-Dec-78	01:26	PM		
Mod 14	Extended	Pretest	19-Dec-78	01:41	PM		
Mod 15	Normal	Posttest	19-Dec-78	03:47	PM		
Mod 16	Normal	Posttest	19-Dec-78	03:52	PM		

STUDENT 17

Mod 1	Normal	Pretest	04-Dec-78	01:36	PM		
	Normal	Posttest	04-Dec-78	01:43	PM		
Mod 2	Normal	Pretest	04-Dec-78	01:47	PM		
	Extended	Posttest	06-Dec-78	02:21	PM		
Mod 4	Normal	Posttest	13-Dec-78	02:05	PM		
Mod 6	Normal	Pretest	13-Dec-78	02:21	PM		
	Normal	Posttest	19-Dec-78	12:22	PM		
	Normal	Posttest	19-Dec-78	12:24	PM	D =	2 min
	Extended	Posttest	19-Dec-78	12:29	PM	D =	5 min
Mod 8	Normal	Pretest	19-Dec-78	01:18	PM		

Data for RHODE ISLAND JUNIOR COLLEGE 02-Feb-79, page 53

STUDENT 18

Mod	1	Normal	Pretest	05-Dec-78	10:13 AM
Mod	2	Normal	Posttest	12-Dec-78	08:50 AM
Mod	3	Extended	Pretest	12-Dec-78	09:22 AM
Mod	4	Normal	Posttest	12-Dec-78	10:18 AM

STUDENT 19

Mod	1	Normal	Pretest	05-Dec-78	10:36 AM
		Extended	Posttest	05-Dec-78	10:41 AM
Mod	2	Normal	Pretest	05-Dec-78	10:43 AM

STUDENT 20

Mod	1	Extended	Pretest	07-Dec-78	12:56 PM
		Normal	Posttest	07-Dec-78	01:06 PM
Mod	2	Normal	Posttest	08-Dec-78	01:15 PM
Mod	3	Normal	Pretest	12-Dec-78	12:39 PM
Mod	4	Normal	Pretest	12-Dec-78	12:53 PM
		Extended	Posttest	14-Dec-78	10:59 AM

STUDENT 21

Mod	1	Normal	Pretest	07-Dec-78	01:48 PM

STUDENT 22

Mod	1	Normal	Pretest	13-Dec-78	10:35 AM
Mod	2	Normal	Pretest	13-Dec-78	10:44 AM
		Normal	Posttest	13-Dec-78	12:11 PM
Mod	4	Extended	Pretest	13-Dec-78	12:20 PM
		Normal	Posttest	13-Dec-78	01:22 PM
		Normal	Posttest	15-Dec-78	11:10 AM

		Normal	Posttest	15-Dec-78	11:13 AM	D = 3 min
		Normal	Posttest	15-Dec-78	11:54 AM	
Mod	6	Extended	Pretest	19-Dec-78	08:07 AM	
		Normal	Posttest	19-Dec-78	12:05 PM	
Mod	8	Normal	Pretest	19-Dec-78	12:18 PM	
		Normal	Posttest	20-Dec-78	08:26 AM	
Mod	9	Normal	Pretest	20-Dec-78	08:31 AM	

STUDENT 23

Mod	1	Normal	Pretest	07-Dec-78	02:42 PM	
		Normal	Posttest	07-Dec-78	03:02 PM	

STUDENT 24

Mod	1	Normal	Posttest	07-Dec-78	02:55 PM	
		Extended	Posttest	07-Dec-78	03:03 PM	
Mod	2	Normal	Pretest	07-Dec-78	03:05 PM	
		Normal	Posttest	13-Dec-78	03:08 PM	
Mod	3	Normal	Pretest	13-Dec-78	03:19 PM	
Mod	4	Normal	Pretest	13-Dec-78	03:29 PM	
		Extended	Posttest	19-Dec-78	02:28 PM	
Mod	6	Normal	Pretest	19-Dec-78	02:56 PM	
		Normal	Posttest	19-Dec-78	02:59 PM	D = 3 min
Mod	7	Normal	Pretest	19-Dec-78	03:13 PM	

STUDENT 25

Mod	1	Extended	Pretest	07-Dec-78	03:54 PM	
Mod	2	Normal	Pretest	07-Dec-78	04:08 PM	
		Normal	Posttest	08-Dec-78	09:46 AM	
Mod	4	Normal	Posttest	08-Dec-78	11:19 AM	
Mod	5	Normal	Pretest	08-Dec-78	11:32 AM	

Mod 6	Extended	Pretest	08-Dec-78	02:10	PM	
	Normal	Posttest	11-Dec-78	09:35	AM	
Mod 8	Normal	Pretest	11-Dec-78	09:43	AM	
Mod 9	Normal	Pretest	14-Dec-78	02:28	PM	
	Normal	Posttest	14-Dec-78	04:21	PM	
Mod 10	Extended	Pretest	18-Dec-78	03:40	PM	
	Normal	Posttest	19-Dec-78	12:44	PM	
	Normal	Posttest	19-Dec-78	12:48	PM	D = 4 min
Mod 3	Normal	Pretest	20-Dec-78	12:16	PM	

STUDENT 26

Mod 1	Normal	Pretest	13-Dec-78	01:38	PM	
Mod 2	Normal	Pretest	13-Dec-78	01:54	PM	

STUDENT 27

Mod 1	Normal	Pretest	13-Dec-78	03:19	PM	
Mod 2	Normal	Pretest	13-Dec-78	03:32	PM	

STUDENT 28

Mod 1	Normal	Pretest	18-Dec-78	02:45	PM	
	Normal	Posttest	18-Dec-78	02:50	PM	
Mod 2	Extended	Pretest	18-Dec-78	02:51	PM	
	Normal	Posttest	18-Dec-78	07:41	PM	
Mod 3	Normal	Pretest	18-Dec-78	07:57	PM	
Mod 4	Normal	Pretest	18-Dec-78	08:20	PM	
	Normal	Posttest	18-Dec-78	08:24	PM	D = 4 min
	Extended	Posttest	19-Dec-78	12:32	PM	
	Normal	Posttest	19-Dec-78	02:08	PM	
	Normal	Posttest	19-Dec-78	02:19	PM	D = 11 min
	Normal	Posttest	19-Dec-78	08:38	PM	
Mod 9	Normal	Pretest	12-Jan-79	11:46	AM	
	Extended	Posttest	12-Jan-79	12:09	PM	D = 23 min

	Normal	Posttest	12-Jan-79	12:51 PM	
Mod 10	Normal	Pretest	16-Jan-79	11:01 AM	

STUDENT 29

Mod 1	Normal	Pretest	20-Dec-78	11:13 AM	
	Extended	Posttest	20-Dec-78	11:17 AM	
Mod 2	Normal	Pretest	20-Dec-78	11:19 AM	
Mod 3	Normal	Pretest	20-Dec-78	11:31 AM	
	Normal	Posttest	21-Dec-78	01:22 PM	
Mod 4	Normal	Posttest	21-Dec-78	01:30 PM	
Mod 5	Extended	Pretest	21-Dec-78	01:50 PM	
Mod 6	Normal	Pretest	21-Dec-78	01:55 PM	

STUDENT 30

Mod 1	Extended	Posttest	15-Jan-79	01:18 PM	
Mod 6	Normal	Pretest	15-Jan-79	02:29 PM	
	Normal	Posttest	15-Jan-79	02:57 PM	D = 28 min
Mod 8	Normal	Pretest	15-Jan-79	03:13 PM	
	Normal	Posttest	15-Jan-79	03:30 PM	D = 17 min
Mod 9	Extended	Pretest	15-Jan-79	03:48 PM	
	Normal	Posttest	15-Jan-79	03:57 PM	D = 9 min
	Normal	Posttest	16-Jan-79	12:40 PM	
Mod 10	Normal	Pretest	16-Jan-79	01:11 PM	
Mod 5	Normal	Pretest	17-Jan-79	01:01 PM	
Mod 7	Extended	Pretest	17-Jan-79	02:19 PM	
	Normal	Posttest	17-Jan-79	02:44 PM	D = 25 min
Mod 11	Normal	Pretest	17-Jan-79	02:59 PM	

Data for RHODE ISLAND JUNIOR COLLEGE 02-Feb-79, page 57

STUDENT 31

Mod 1	Normal	Posttest	15-Jan-79	01:51 PM	
Mod 8	Normal	Pretest	15-Jan-79	02:00 PM	
	Normal	Posttest	15-Jan-79	02:22 PM	D = 22 min
Mod 9	Normal	Pretest	15-Jan-79	02:40 PM	
Mod 10	Extended	Pretest	16-Jan-79	02:16 PM	

STUDENT 32

Mod 1	Normal	Posttest	17-Jan-79	10:48 AM	
Mod 2	Normal	Posttest	18-Jan-79	10:13 AM	
	Normal	Posttest	18-Jan-79	10:29 AM	D = 16 min
	Extended	Posttest	18-Jan-79	10:51 AM	D = 22 min
	Normal	Posttest	18-Jan-79	11:19 AM	D = 28 min

STUDENT 33

Mod 1	Normal	Posttest	17-Jan-79	10:56 AM
Mod 2	Normal	Posttest	18-Jan-79	10:04 AM
	Extended	Posttest	18-Jan-79	10:36 AM

STUDENT 34

STUDENT 35

Mod 1	Extended	Posttest	17-Jan-79	01:30 PM
	Normal	Posttest	17-Jan-79	01:46 PM

STUDENT 36

Mod 1	Normal	Posttest	18-Jan-79	11:36 AM
Mod 2	Normal	Posttest	18-Jan-79	11:53 AM

Data for RHODE ISLAND JUNIOR COLLEGE 02-Feb-79, page 58

SUMMARY

```
Number of  Normal  Pretests  = 131
         Extended Pretests   =  31
   Total Number of Pretests  =      162 (41.0%)

Number of  Normal  Posttests = 190
         Extended Posttests  =  43
   Total Number of Posttests =      233 (59.0%)
                                   -----
 Total Number of Tests Taken =      395


 Instances of Repeated Tests = 133
 No. of Elapsed Times < 15 m =  36 (27.1%)
 No. of Elapsed Times < 30 m =  55 (41.4%)
```

Appendix C

TEST LENGTH DATA FOR INDIVIDUAL SITES

The data in this appendix is the raw data from which Figure 14 (page 87) was created. It shows the distribution of tests for each of the possible test lengths (1 to 30 items) and summary statistics.

Data for WACHUSETT REG'L. SCHOOL DIST.   02-Feb-79, page 40

SUMMARY TEST LENGTH DATA for Modules 2 through 16

No. of Pretests =	35	No. of Extended Masters =	4
No. of Posttests =	54	No. of Extended Non-Masters =	8
Total No. of Tests =	89	Total No. of Extended Tests =	12

Test Length in No. Items	No. of Pretest Masters	No. of Pretest Non-Masters	No. of Posttest Masters	No. of Posttest Non-Masters	Total No. of Masters	Total No. of Non-Masters
1	0	5	0	4	0	9
2	0	1	0	2	0	3
3	0	1	0	2	0	3
4	0	1	0	4	0	5
5	0	0	0	3	0	3
6	0	3	0	2	0	5
7	0	0	0	1	0	1
8	0	1	0	1	0	2
9	0	1	1	2	1	3
10	0	0	4	0	4	0
11	0	1	1	0	1	1
12	0	1	0	0	0	1
13	0	1	3	1	3	2
14	1	1	1	0	2	1
15	1	1	0	0	1	1
16	0	0	0	1	0	1
17	1	0	0	0	1	0
18	0	0	1	0	1	0
19	0	1	1	0	1	1
20	0	0	1	0	1	0
21	1	1	1	2	2	3
22	1	1	0	0	1	1
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	0	0	0
26	1	0	0	0	1	0
27	0	1	0	0	0	1
28	1	0	1	0	2	0
29	0	0	0	0	0	0
30	4	2	11	3	15	5
Actual Totals:	11	24	26	28	37	52
Corrected Totals:	10	17	23	21	33	38
Statistics:						
Mean	23	11	20	7	21	9
Median	22	11	19	5	21	6

Data for RHODE ISLAND JUNIOR COLLEGE 02-Feb-79, page 75

SUMMARY TEST LENGTH DATA for Modules 2 through 16

No. of Pretests = 138 No. of Extended Masters = 29
No. of Posttests = 227 No. of Extended Non-Masters = 27
Total No. of Tests = 365 Total No. of Extended Tests = 56

Test Length in No. Items	No. of Pretest Masters	No. of Pretest Non-Masters	No. of Posttest Masters	No. of Posttest Non-Masters	Total No. of Masters	Total No. of Non-Masters
1	0	10	0	22	0	32
2	0	1	0	3	0	4
3	0	9	0	7	0	16
4	0	8	0	8	0	16
5	0	5	0	2	0	7
6	0	3	0	4	0	7
7	0	10	0	2	0	12
8	0	2	2	4	2	6
9	0	5	8	2	8	7
10	0	3	5	2	5	5
11	1	4	6	2	7	6
12	1	1	2	3	3	4
13	1	2	5	1	6	3
14	0	3	3	3	3	6
15	0	2	5	0	5	2
16	0	1	4	1	4	2
17	0	2	7	1	7	3
18	0	2	8	0	8	2
19	0	2	8	2	8	4
20	1	2	3	2	4	4
21	0	1	3	2	3	3
22	1	3	0	1	1	4
23	0	1	4	0	4	1
24	1	1	6	1	7	2
25	0	1	2	0	2	1
26	0	1	2	1	2	2
27	1	1	2	3	3	4
28	1	2	3	3	4	5
29	5	3	4	1	9	4
30	18	16	42	10	60	26
Actual Totals:	31	107	134	93	165	200
Corrected Totals:	22	81	114	61	136	142
Statistics:						
Mean	26	11	20	11	21	11
Median	29	9	19	9	20	9

Data for FALMOUTH HIGH SCHOOL 02-Feb-79, page 102

SUMMARY TEST LENGTH DATA for Modules 2 through 16

No. of Pretests	= 43	No. of Extended Masters	= 13
No. of Posttests	= 116	No. of Extended Non-Masters	= 14
Total No. of Tests	= 159	Total No. of Extended Tests	= 27

Test Length in No. Items	No. of Pretest Masters	No. of Pretest Non-Masters	No. of Posttest Masters	No. of Posttest Non-Masters	Total No. of Masters	Total No. of Non-Masters
1	0	4	0	4	0	8
2	0	0	0	2	0	2
3	0	3	0	2	0	5
4	0	1	0	6	0	7
5	0	5	0	6	0	11
6	0	2	0	3	0	5
7	0	3	0	1	0	4
8	0	1	2	2	2	3
9	0	3	0	2	0	5
10	0	0	5	2	5	2
11	1	3	3	2	4	5
12	0	0	1	1	1	1
13	0	0	2	3	2	3
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	1	0	1	0
17	0	0	0	0	0	0
18	0	1	3	2	3	3
19	0	1	0	0	0	1
20	0	1	1	2	1	3
21	0	1	1	0	1	1
22	0	0	1	1	1	1
23	0	1	0	1	0	2
24	0	0	3	1	3	1
25	0	0	1	0	1	0
26	0	0	4	0	4	0
27	0	0	1	0	1	0
28	0	4	1	1	1	5
29	0	0	1	2	1	2
30	3	5	31	8	34	13
Actual Totals:	4	39	62	54	66	93
Corrected Totals:	3	30	50	42	53	72
Statistics:						
Mean	24	11	23	10	23	11
Median	30	8	26	8	26	8

Appendix D

TEST RELIABILITY DATA FOR INDIVIDUAL SITES

This data, like that in Appendix C, is the raw data for the summary that was presented and discussed in the body of the dissertation. The test reliability data for all sites combined was presented in Table 7 on page 91 and discussed on pages 90 to 95.

DATA FOR WACHUSETT REG'L. SCHOOL DIST.

02-Feb-79

TEST TIME DATA

Total Student On-Line Time	15 hours, 41 minutes
Number of Student Logins	86
Total Student Testing Time	11 hours, 1 minutes
Number of Tests Administered	108
Average Time per Test	6.1 minutes
Average Login Overhead	42.4 %
Number of Test Items Presented	1271
Average Time per Item	31.2 seconds

Pretest Fraction = 0.273 Posttest Fraction = 0.727

TEST RELIABILITY DATA

SUMMARY: MODULES 2-16

O = Observed
E = Expected

EXTENDED DECISION	EARLY DECISION: Master	EARLY DECISION: Non-Master	
Master	O = 4 E = 3.827	O = 0 E = 0.732	4
Non-Master	O = 0 E = 0.173	O = 8 E = 7.268	8
	4	8	N = 12

% Agreement =	1.000	Avg. Item Saving =	8.8 items
Kappa =	1.000	Avg. Time Saving =	4.4 min.
Phi =	1.000	% TF and YN Items =	28.7 %
G =	1.000	% MC4 Items =	60.4 %
Chi Square =	0.986 Not Sig. at .05	% MC5 Items =	11.0 %

DATA FOR RHODE ISLAND JUNIOR COLLEGE

02-Feb-79

TEST TIME DATA

Total Student On-Line Time	124 hours, 9 minutes
Number of Student Logins	343
Total Student Testing Time	85 hours, 46 minutes
Number of Tests Administered	420
Average Time per Test	12.3 minutes
Average Login Overhead	44.7 %
Number of Test Items Presented	6585
Average Time per Item	46.9 seconds

Pretest Fraction = 0.406   Posttest Fraction = 0.594

TEST RELIABILITY DATA

SUMMARY: MODULES 2-16

O = Observed
E = Expected

EXTENDED DECISION	EARLY DECISION: Master	EARLY DECISION: Non-Master	
Master	O = 27 E = 25.924	O = 2 E = 2.474	29
Non-Master	O = 0 E = 1.076	O = 27 E = 26.526	27
	27	29	N = 56

% Agreement =	0.964	Avg. Item Saving =	11.0 items
Kappa =	0.929	Avg. Time Saving =	6.8 min.
Phi =	0.931	% TF and YN Items =	28.7 %
G =	0.929	% MC4 Items =	60.4 %
Chi Square =	1.220 Not Sig. at .05	% MC5 Items =	11.0 %

DATA FOR FALMOUTH HIGH SCHOOL

02-Feb-79

TEST TIME DATA

Total Student On-Line Time	36 hours, 4 minutes
Number of Student Logins	230
Total Student Testing Time	27 hours, 56 minutes
Number of Tests Administered	180
Average Time per Test	9.3 minutes
Average Login Overhead	29.1 %
Number of Test Items Presented	3044
Average Time per Item	33.0 seconds

Pretest Fraction = 0.267 Posttest Fraction = 0.733

TEST RELIABILITY DATA

SUMMARY: MODULES 2-16

O = Observed
E = Expected

		EARLY DECISION		
		Master	Non-Master	
EXTENDED DECISION	Master	O = 12 E = 11.480	O = 1 E = 1.376	13
	Non-Master	O = 0 E = 0.520	O = 14 E = 13.624	14
		12	15	N = 27

% Agreement =	0.963	Avg. Item Saving =	13.6 items
Kappa =	0.926	Avg. Time Saving =	5.7 min.
Phi =	0.928	% TF and YN Items =	28.7 %
G =	0.926	% MC4 Items =	60.4 %
Chi Square =	0.657 Not Sig. at .05	% MC5 Items =	11.0 %

DATA FOR ALL SITES COMBINED

02-Feb-79

TEST TIME DATA

Total Student On-Line Time	175 hours, 54 minutes
Number of Student Logins	659
Total Student Testing Time	124 hours, 43 minutes
Number of Tests Administered	708
Average Time per Test	10.6 minutes
Average Login Overhead	41.0 %
Number of Test Items Presented	10900
Average Time per Item	41.2 seconds

Pretest Fraction = 0.345 Posttest Fraction = 0.655

TEST RELIABILITY DATA

SUMMARY: MODULES 2-16

O = Observed
E = Expected

EXTENDED DECISION	EARLY DECISION: Master	EARLY DECISION: Non-Master	
Master	O = 43 E = 41.221	O = 3 E = 4.583	46
Non-Master	O = 0 E = 1.779	O = 49 E = 47.417	49
	43	52	N = 95

% Agreement =	0.968	Avg. Item Saving =	11.5 items
Kappa =	0.937	Avg. Time Saving =	6.2 min.
Phi =	0.939	% TF and YN Items =	28.7 %
G =	0.937	% MC4 Items =	60.4 %
Chi Square =	2.456 Not Sig. at .05	% MC5 Items =	11.0 %

REFERENCES CITED

ADAMS, E.N. On scoring a mastery learning control test. Journal of Computer-Based Instruction 1(2):50-58. November, 1974.

ANSFIELD, Paul J. A user oriented computing procedure for compiling and generating examinations. Educational Technology 13(3):12-13. March, 1973.

ARY, Donald, Lucy Chester Jacobs, and Asghar Razavieh. Introduction to Research in Education. Holt, Rinehart and Winston, Inc., New York. 1972.

BAKER, Eva L. Beyond objectives: domain-referenced tests for evaluation and instructional improvement. Educational Technology 14(6):10-16. June, 1974.

BAKER, Frank B. An interactive approach to test construction. Educational Technology 13(3):13-15. March, 1973.

BEJAR, Isaac I., David J. Weiss, and Kathleen A. Gailluca. An information comparison of conventional and adaptive tests in the measurement of classroom achievement. Research Report 77-7, Psychometric Methods Program, Dept. of Psychology, University of Minnesota. October, 1977.

BROWN, Williard A. Improvement of testing and course evaluation. Journal of Research in Science Teaching 5:240-243. 1967-1968.

________. A computer examination compositor for the IBM 360/40. Western Washington State College, mimeographed. 1972.

BUCHANAN, C.C. CMI at United Airlines. ADCIS SIG CBT Newsletter 1(1):1-4. July, 1978.

BUCKLEY-SHARP, M.D. A multiple choice question banking system. Educational Technology 13(3):16-18. March, 1973.

BUTZOW, John W., and Roland R. Pare. Physical science: a multi-media facilitated course. Journal of College Science Teaching 2(1):29-32. October, 1972.

CARVER, R.P. Special problems in measuring change with psychometric devices. In Evaluative Research: Statistics and Methods. American Institute for Research, Washington, D.C. 1970.

COHEN, Perrin S., and Leila R. Cohen. Computer generated tests for a student paced course. Educational Technology 13(3):18-19. March, 1973.

COX, Richard C. Evaluative aspects of criterion-referenced measures. In Popham, W. James (ed.), Criterion-Referenced Measurement, pp. 67-75. Educational Technology Publications, Englewood Cliffs, N.J. 1971.

CURLETTE, William L. Assessing the reliability of criterion-referenced tests. Paper presented at a conference on Innovative Education: Preservice thorugh Inservice, Atlanta, GA. February, 1977.

DEAN, Peter M. IBM's Field Training System. ADCIS SIG CBT Newsletter 1(2):2-3. October, 1978.

DENNEY, Cecil. There is more to a test pool than data collection. Educational Technology 13(3):19-20. March, 1973.

DIGITAL EQUIPMENT CORPORATION. The BASIC Primer. Digital Equipment Corporation, Maynard, Mass. 1979.

DUDLEY, Thomas J. How the computer assists in pacing and testing students' progress. Educational Technology 13(3): 21-22. March, 1973.

EDUCATIONAL TESTING SERVICE. Computer-assisted assembly of tests at ETS. A paper presented at a conference on Computer-Assisted Test Construction, San Diego, CA. October, 1974.

EMRICK, J.A. An evaluation model for mastery testing. Journal of Educational Measurement 8:321-326. 1971.

FERGUSON, Richard L. Computer-assisted criterion-referenced testing. Learning Research and Development Center, University of Pittsburgh, Working Paper 49. March, 1970.

________. Computer assistance for individualized measurement. Learning Research and Development Center, University of Pittsburgh. March, 1971.

FERGUSON, Richard L., and Melvin R. Novick. The implementation of a Bayesian system for decision analysis in a program of Individually Prescribed Instruction. American College Testing Program, Iowa City, Iowa, Research Report No. 60. 1973.

FRANKE, R.G., W.D. Dolphin, G.F. Covert, and C.D. Jorgensen. A computer-assisted method for teaching large enrollment lecture sessions: the biology phase achievement system. Iowa State University, mimeographed. 1972.

GARVIN, Alfred D. The applicability of criterion-referenced measurement by content area and level. In Popham, W. James (ed.), Criterion-Referenced Measurement, pp. 55-63. Educational Technology Publications, Englewood Cliffs, N.J. 1971.

GEISERT, Paul. An instructional systems approach to the development of a course in introductory biological science. The University of Wyoming, mimeographed. 1973.

GHESQUIRE, James. Computer use in teaching and research. University of Illinois, unpublished Ph.D. dissertation. 1973.

GLASER, Robert. Instructional technology and the measurement of learning outcomes: some questions. American Psychologist 18:519-521. 1963.

________. A criterion-referenced test. In Popham, W. James (ed.), Criterion-Referenced Measurement, pp. 41-51. Educational Technology Publications, Englewood Cliffs, N.J. 1971.

HAMBLETON, Ronald K., and William P. Gorth. Criterion-referenced testing: issues and applications. University of Massachusetts School of Education, Amherst, MA. September, 1971.

HAMBLETON, Ronald K., and Melvin R. Novick. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement 9:27-29. 1972.

HANSEN, Duncan N. An investigation of computer-based science teaching. In Richard C. Atkinson and H.A. Wilson (ed.), Computer-Assisted Instruction: A Book of Readings. Academic Press, New York, N.Y., pp. 209-226. 1969.

HANSEN, Duncan N., Steven Ross, and Dickie A. Harris. Flexilevel adaptive testing paradigm: hierarchical concept structures. Technical Training Division, Lowry Air Force Base, Colorado. AFHRL-TR-77-35 (II). July, 1977.

HARRIS. An interpretation of Livingston's reliability coefficient for criterion-referenced tests. Journal of Educational Measurement. Spring, 1972.

HAYNES, Judy L., and Crane Walker. Establishing criteria for objective mastery -- art or science? Paper presented at a conference of the American Education Research Association. April, 1976.

HAZLETT, C.B. MEDSIRCH: multiple choice test items. Educational Technology 13(3):24-26. March, 1973.

HEINES, Jesse M. An interactive, computer-managed model for the evaluation of audio-tutorial instruction. University of Maine at Orono, unpublished M.S. thesis. May, 1974.

________. A review of the literature on criterion-referenced and computer-assisted testing. Boston University, ERIC Document No. ED116633. November, 1975.

HIVELY, Wells. Introduction to domain-referenced testing. Educational Technology 14(6):5-10. June, 1974a.

________. Some comments on this issue. Educational Technology 14(6):60-64. June, 1974b.

HSU, Tse-Chi, and Marthena Carlson. Test construction aspects of the computer-assisted testing model. Educational Technology 13(3):26-127. March, 1973.

ISSAC, Stephen, and William B. Michael. Handbook in Research and Evaluation. Robert R. Knapp, Publisher, San Diego, California. 1971.

JENSEN, Donald D. Toward efficient, effective, and humane instruction in large classes: student scheduled involvement in films, discussions and computer generated exams. Educational Technology 13(3):28-29. March, 1973.

KELLEY, Allan C. An experiment with TIPS: a computer-aided instructional system for undergraduate education. The American Economic Review 58(2):446-457. May, 1968.

________. TIPS and technical change in classroom instruction. The American Economic Review 62(2):442-428. May, 1972.

________. How to implement TIPS. Duke University, mimeographed. 1972.

________. TIPS: a diagnostic tool to individualize instruction in the large class. Duke University, pamphlet. 1972.

LEWIS, Don, and C.J. Burke. The use and misuse of the chi-square test. Psychological Bulletin 46:433-489. 1949.

LIBAW, Frieda B. Constructing tests with the MENTREX tutorial system. Educational Technology 13(3):30-31. March, 1973.

LIPPEY, Gerald. The computer can support test construction in a variety of ways. Educational Technology 13(3):10-12. March, 1973.

LINDVALL, C.M., and Anthony J. Nitko. Criterion-referenced testing and the individualization of instruction. Paper presented at the annual meeting of the National Council on Measurement in Education. February, 1960.

LIVINGSTON, S. Criterion-referenced applications of classical test theory. Journal of Educational Measurement. Spring, 1972a.

________. A reply to Harris' "an interpretation of Livingston's reliability coefficient for criterion-referenced tests". Journal of Educational Measurement. Spring, 1972b.

________. Lecture on "statistical concepts in criterion-referenced measurement". Educational Testing Service, Princeton, N.J. February 23, 1976.

MAGER, Robert F., and Peter Pipe. Criterion-Referenced Instruction: Analysis, Design, and Implementation. Mager Associates, Inc., Los Altos Hills, CA. 1974.

MAGINN, Michael. Let's jump off the individualized instruction bandwagon. Training 14(1):23-26. January, 1977.

McLAGAN, Patricia A., Raymond E. Sandburgh, and Van O. Wright. Special report: computer-assisted instruction. Training 14(9): 48-57. September, 1977.

MENNE, John W., and Paul Lustgraaf. Computer-assisted test assembly at Iowa State University. Paper presented at a conference on Computer-Assisted Test Construction, San Diego, CA. October, 1974.

MESKAUSKAS, John A. Evaluation models for criterion-referenced testing: views regarding mastery and standard setting. Review of Educational Research 36(1):133-158. Winter, 1976.

MILLMAN, Jason. Sampling plans for domain-referenced tests. Educational Technology 14(6):17-21. June, 1974.

OLYMPIA, P.L., Jr. Computer generation of truly repeatable examinations. Educational Technology 15(6):53-55. June, 1975.

O'REILLY, Robert P., William P. Gorth, and Paul Pinsky. Computer-assisted test construction: an effort based on evaluation methodology. Educational Technology 13(3):32-34. March, 1973.

PARE, Roland R. An analysis of selected variables of student performance in an audio-tutorial course in physical science for non-science majors. University of Maine at Orono, unpublished Ed.D. dissertation. 1973.

POPHAM, W. James. Indices of adequacy for criterion-referenced test items. In Popham, W. James (ed.), Criterion-Referenced Measurement, pp. 79-98. Educational Technology Publications, Englewood Cliffs, N.J. 1971.

POPHAM, W. James, and T.R. Husek. Implications of criterion-referenced measurement. Journal of Educational Measurement 6(1): 1-9. 1969.

POSTLETHWAIT, S.N., J. Novak, and H.T. Murray, Jr. The AudioTutorial Approach to Learning. Burgess Publishing Company, Minneapolis, MN. 1972.

PROSSER, Franklin. Repeatable tests. Educational Technology 13(3):34-35. March, 1973.

REMONDINI, David J. Test Item System: a method of computer-assisted test assembly. Educational Technology 13(3):35-37. March, 1973.

REMONDINI, David J., and John E. Miller. A computerized system for preparation of tests in academic disciplines. Proceedings of a conference on Computers in the Undergraduate Curricula, pp. 7.24-7.30. The University of Iowa. June, 1970.

SALISNJACK, Julian. Computer-aided test preparation: six years of experience. Educational Technology 13(3):37-38. March, 1973.

SCHNEIDER, Edward W., and Tom McMurchie. TICCIT response recording format specification. Brigham Young University, mimeographed. May, 1973.

SCHNEIDER, Edward W., and Stephen R. Fine. Testing and grading in the TICCIT system. Institute for Computer Uses in Education, Brigham Young University, Provo, Utah. Technical Report No. 13. May, 1973.

SCHONBERGER, Richard J. Modular instruction with computer assembled repeatable exams: second generation. Educational Technology 15(2):36-38. February, 1975.

SIEGEL, Sidney. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill Book Company, New York. 1956.

SIVERTSON, Sigurd E., Richard H. Hansen, and Adeline O. Schoenenberger. Computerized test bank for clinical medicine. Educational Technology 13(3):38-39. March, 1973.

STODOLA, Quentin. Use of computer assembled tests in the California State Univerity and College System. Educational Technology 13(3):40-41. March, 1973.

SUBKOVIAK, Michael J. Estimating the reliability from a single administration of a criterion-referenced test. Paper presented at a conference of the American Education Research Association, San Francisco, CA. 1976.

________. Empirical investigation of procedures for estimating reliability of master tests. Journal of Educational Measurement 15(2):111-116. Summer, 1978.

SWAMINATHAN, H., R.K. Hambleton, and J.J. Algina. Reliability of criterion-referenced tests: a decision-theoretic formulation. Journal of Educational Measurement 11:263-267. 1974.

SWEZEY, R.W., and R.B. Pearlstein. Guidebook for Developing Criterion-Referenced Tests. US Army Research Institute for the Behavioral and Social Sciences. 1975.

TOGGENBURGER, Frank. Classroom teacher support system. Educational Technology 13(3):42-43. March, 1973.

VALE, C. David. Adaptive testing and the problem of classification. Research Report 77-1, Psychometric Methods Program, Dept. of Psychology, University of Minnesota, pp. 24-36. October, 1976.

VICKERS, F.D. Creative test generators. Educational Technology 13(3):43-44. March, 1973.

WALD, Abraham. Sequential Analysis. John Wiley and Sons, Inc. 1947.

WATERS, B.K. Empirical investigation of the stradaptive testing model for measurement of human ability. Flying Training Division, Williams Air Force Base, Arizona. AFHRL-TR-75-27, AD-A018611. October, 1975.

WEISS, David J. Computerized adaptive ability measurement. University of Minnesota, mimeographed. 1973.

ZIEKY, Michael J. Applications and adaptions of classical test theory to criterion-referenced measures. Educational Testing Service, mimeographed. 1974.

VITA

Jesse M. Heines was born to Mr. Philip Heines and Mrs. Dorothy S. Heines on April 18, 1948 in New Brunswick, New Jersey. He graduated from East Brunswick High School, East Brunswick, N.J. in 1966. In 1970 he received an S.B. in Earth Science from the Massachusetts Institute of Technology in Cambridge, Massachusetts. He received an M.S. in Science Education from the University of Maine in 1974.

From 1970 to 1972, Mr. Heines taught sixth, seventh, and eighth grade science and mathematics at the Anglo-American School in Moscow, U.S.S.R. There he met Evelyn Scott McKay, the Anglo-American School first grade teacher, and they were married on August 16, 1971 in Edinburgh, Scotland. They now have two children, Scott Douglas and Russell Gordon, and reside in Chelmsford, Massachusetts.

Mr. Heines is currently Manager of Computer-Based Course Development at Digital Equipment Corporation. This group develops CAI and CMI courses for both internal and customer training. Mr. Heines has been a member of the Association for the Development of Computer-based Instructional Systems since 1973, and he is currently the Chairperson of the Special Interest Group in Computer-Based Training.