

AD-A175 319

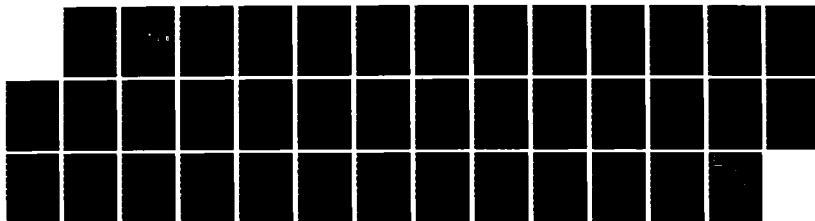
AN ANALYSIS OF MINIMUM SYSTEM REQUIREMENTS TO SUPPORT
COMPUTERIZED ADAPTIVE TESTING(U) NAVAL POSTGRADUATE
SCHOOL MONTEREY CA B J MCCORMACK SEP 86

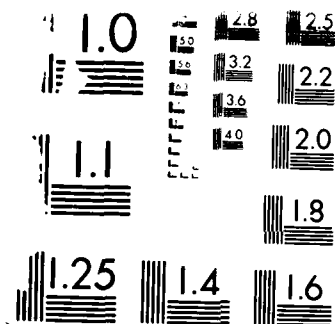
1/1

UNCLASSIFIED

F/G 5/10

NL





2

AD-A175 319

NAVAL POSTGRADUATE SCHOOL

Monterey, California



DTIC
ELECTE
DEC 29 1986
S D D

THESIS

DTIC FILE COPY

AN ANALYSIS OF MINIMUM SYSTEM REQUIREMENTS
TO SUPPORT
COMPUTERIZED ADAPTIVE TESTING

by

Brian J. McCormack

September 1986

Thesis Advisor:

Bruce Bloxom

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE				
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School		6b. OFFICE SYMBOL (If applicable) 54	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School	
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000			7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS	
			PROGRAM ELEMENT NO	PROJECT NO
			TASK NO	WORK UNIT ACCESSION NO
11. TITLE (Include Security Classification) AN ANALYSIS OF MINIMUM SYSTEM REQUIREMENTS TO SUPPORT COMPUTERIZED ADAPTIVE TESTING				
12. PERSONAL AUTHOR(S) McCormack, Brian J.				
13a. TYPE OF REPORT Master's Thesis		13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1986 September 26	15. PAGE COUNT 38
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP		
			computer adaptive test (CAT); adaptive testing	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)				
This paper discusses the minimum system requirements needed to develop a computerized adaptive test (CAT). It lists some of the benefits of adaptive testing, establishes a set of operational constraints, and reviews both software and hardware requirements based on those operational constraints. An experimental CAT system that is currently in use is reviewed in detail.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Bruce Bloxom			22b. TELEPHONE (Include Area Code) (408) 646-2905	22c. OFFICE SYMBOL 54B1

Approved for public release; distribution is unlimited.

An Analysis of Minimum System Requirements
to Support
Computerized Adaptive Testing

by

Brian J. McCormack
Lieutenant, United States Navy
B.S., United States Naval Academy, 1978

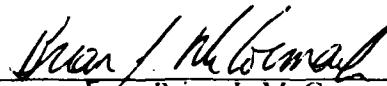
Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

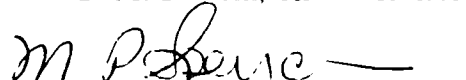
NAVAL POSTGRADUATE SCHOOL
September 1986

Author:


Brian J. McCormack

Approved by:


B. M. Bloxom, Thesis Advisor


M. P. Spencer, Second Reader


W. R. Greer, Jr., Chairman,
Department of Administrative Science


Kneale T. Marshall,
Dean of Information and Policy Sciences

ABSTRACT

This paper discusses the minimum system requirements needed to develop a computerized adaptive test (CAT). It lists some of the benefits of adaptive testing, establishes a set of operational constraints, and reviews both software and hardware requirements based on those operational constraints. An experimental CAT system that is currently in use is reviewed in detail.

Accession For	
NTIS CR&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or Special
A-1	

TABLE OF CONTENTS

I.	INTRODUCTION	6
A.	BENEFITS OF AN ADAPTIVE TEST	7
B.	RATIONALE FOR AN ADAPTIVE TEST	9
	1. Item Characteristic Curve	9
	2. Ability Estimation and Item Selection	9
C.	DEVELOPMENT OF AN ADAPTIVE TEST	9
	1. Item Pool Development	10
	2. Test Administration	10
	3. Software Alternatives	10
	4. Hardware Alternatives	11
	5. Test Evaluation	11
D.	PURPOSES OF THE THESIS	11
	1. Operational Considerations	12
	2. Specific Purposes of the Thesis	13
II.	SOFTWARE REQUIREMENTS	14
A.	TEST DEVELOPMENT AND ADMINISTRATION REQUIREMENTS	14
	1. Development of the Item Pool	14
	2. Scoring of the Test	16
	3. Selection Of The Next Item	17
	4. Graphics And Color Displays	18
	5. Summary Of Storage Requirements	18
B.	COMPILER AND OPERATING SYSTEM REQUIREMENTS	19
III.	HARDWARE REQUIREMENTS	20
A.	HARDWARE ISSUES	20
B.	SPECIFIC HARDWARE REQUIREMENTS	20
	1. Specific Functions To Be Supported	21

2. Specific Hardware Options To Choose From	21
3. Fixed Stand Alone System	21
4. Portable System	22
5. Communications	23
6. Networking	24
7. Visual Display	25
8. Microprocessor	25
9. Special Response Input Devices	26
10. Surge Protection	26
IV. EXAMPLE OF A COMPUTERIZED ADAPTIVE TESTING SYSTEM	27
A. THE EXPERIMENTAL CAT/ASVAB	27
1. Description Of The ASVAB	27
2. Minimum System Requirements	27
3. Setting of the Experimental CAT/ASVAB	28
4. Software Used by the Experimental CAT/ASVAB	29
5. Hardware Used by the Experimental CAT/ASVAB	30
6. Summary	32
V. CONCLUSIONS	33
LIST OF REFERENCES	36
INITIAL DISTRIBUTION LIST	37

I. INTRODUCTION

Testing students is an important aspect of any academic or training environment. Tests are used to measure ability, select personnel for specific programs, and to predict their future performance. They are also used to evaluate students at the end of a training exercise or classroom lesson.

The conventional way to measure a person's ability, by using pencil and paper examinations, is characterized by treating all examinees as if they required exactly the same assessment. Each examinee receives the same questions with the same levels of difficulty, and completes the test in roughly the same time block. This conventional style of testing is being considered for replacement by a new type of examination call a computerized adaptive test (CAT), which tends to include only items that are discriminating at the examinee's level of ability. Because of this increased efficiency of measurement, several CAT projects are being developed for or within the armed services. The Navy Personnel Research and Development Center (NPRDC), located in San Diego, CA, has developed an experimental form of the Armed Services Vocational Aptitude Battery (ASVAB) in a CAT form, which is based on the Apple III microcomputer. NPRDC is also has a CAT/ASVAB under development which is based on the Hewlett Packard Integral Personal Computer portable system for subsequent operational use. (Chapter IV contains a detailed description of the experimental CAT/ASVAB.) The Marine Corps is developing a CAT to measure communications electronic achievement in conjunction with ACT Corporation; and the Army is developing a CAT project in order to assist recruiters in their preliminary screening of prospective recruits. The Army project is called the Computerized Adaptive Screening Test (CAST) (NPRDC Rept. 84-17, 1984). Commercial CAT products that are being developed include Psychological Corporation's Apple based system for use with tests sold by the publisher, and Assessment System Corporation's IBM PC based system which can be used for any content area selected by the user.

This paper describes adaptive testing, lists its benefits, and discusses the minimum requirements that are needed in order to support computerized adaptive testing. As the name implies, a CAT differs from a conventional exam by being administered by computer, and by being presented adaptively. The test is administered

one item at a time, with multiple choice questions being displayed on a cathode ray tube (CRT) screen. The computer, by using a specially designed program, selects the most appropriate question from a pool of items stored in the computer, and presents it on the CRT. The examinee then answers the question, and the computer accepts the answer and grades it. (Weiss, 1982, p. 475) By presenting the test adaptively, certain unique advantages that adaptive tests have over conventional pencil and paper examinations can be realized. With adaptive testing, each individual may start the test at a different point, based on a prior estimate of that person's ability. The test difficulty is adapted to each individual by providing a more difficult question after a right answer and an easier question after a wrong answer. Each item is scored as it is administered to the examinee, producing a new estimate of ability and a measure of precision of that estimate. An item selection rule is used to select subsequent items to be asked, based on the most current estimate of the examinee's ability, and testing is terminated according to a predetermined criterion. The criterion could be a fixed number of items asked of the examinee, or a fixed level of precision of the ability estimation. (Weiss, 1982, p. 474) After the test is completed the examinee's score is given as the estimated position on the ability continuum. This is a major difference between conventional and adaptive tests. Even though each examinee's test is individual, based on the person's ability and answers, every examinee is scored on the same scale, despite his having taken different test items. This is possible because of the nature of adaptive testing. The software models that are used select and score questions based on a set of parameters that can describe each question. The computer uses the parameters of each question, and the response to each question to compute and update the estimated ability of each examinee, between questions as well as at the end of the test.

A. BENEFITS OF AN ADAPTIVE TEST

Computerized adaptive testing addresses many of the problems associated with conventional pencil and paper examinations.

Administration time is unnecessarily quite high for pencil and paper tests. In a conventional exam, each examinee must answer the same questions in the same time block, regardless of his or her individual ability. A CAT has less administrative time associated with it because for a given level of desired precision, that level is achieved with fewer items than is the case with pencil and paper exams. The shorter

administration time allows for a higher turnover rate, with more students being tested in a given time period. A CAT can also reduce administrative support time because it is administered by computer. The computer performs normal proctor duties such as timing the test and relaying instructions, allowing one proctor to administer a test to a larger number of examinees. In addition the printing, storage, and handling of test booklets and answer sheets is eliminated by using a CAT, saving administrative time and cost. (U.S. Army Research Institute Rept. 423, 1979, p. 4)

Pencil and paper tests typically provide poor differentiation among people of extreme ability, because the items are typically of only moderate difficulty. Adaptive testing allows for much better differentiation among people of extreme ability, and can even provide a constant degree of precision of measurement across a wide range of ability. Also, the test will contain few questions that are much too easy or too hard, helping to save time and ensure higher motivation and better results.

The administration of the exam by computer will result in quicker feedback for both the examinee and the proctor, and will increase overall test security. Computerized administration of an exam will result in immediate automatic scoring, reporting, and recording of test results. This results in faster feedback to the student and administrator, and reduces the chance for errors in grading that may occur when it is performed manually, or by optical scanning. Pencil and paper tests are considered vulnerable to theft and compromise, but with appropriate safeguards, a CAT can be more secure than pencil and paper exams. Test compromise can be substantially reduced by elimination of test booklets (reducing the likelihood of theft) and by the individualized adaptive test construction (thwarting the use of ordinary cheating devices). (U.S. Army Research Institute Rept. 423, 1979, p. 4)

Expensive, time consuming replacement of test questions is not a problem with a CAT. Initial development of items for a conventional test can be expensive because the test must be given to a separate, large sample of examinees and the results analyzed to ensure its reliability. Although initial development of the CAT item pool is expensive, once it is in place, new items under consideration can be tried unobtrusively in an operational setting without the need to test additional examinees. This helps to reduce the time and cost associated with testing new items and developing new types of exams.

B. RATIONALE FOR AN ADAPTIVE TEST

Most research being conducted in the field of adaptive testing is based on the three parameter item response theory, therefore this paper will consider adaptive test procedures which use item response theory (IRT) models (Green, Bock, et al, 1984, p. 348).

1. Item Characteristic Curve

In an IRT model, each item is represented by an item characteristic curve. The item characteristic curve shows the probability of a person getting an item correct, given his ability--a point on a dimension which is assumed to be common to all items in the test. The curve is an increasing function of ability and is based on three parameters: difficulty, discrimination, and guessing. Item difficulty describes how much ability a person would need in order to have a specified probability of getting the item correct; that probability is halfway between 1.0 and the guessing parameter. The discrimination parameter describes how much the item will discriminate among examinees whose ability levels are near the item's difficulty level; items with a high discrimination have a sharp inflection in the item characteristic curve near the item's difficulty level. The guessing parameter describes the probability of getting an item correctly by guessing, e.g. by a person of very low ability, thus it is the lower asymptote of the function. The interested reader can find the mathematical expressions for the item characteristic curve in Owen (1975).

2. Ability Estimation and Item Selection

Assuming that item characteristic curve parameters have already been established, an initial estimation of the examinee's ability is required. The original estimate could be based on schooling, age, the previous test performance of the examinee, or it could be the same for all examinees. The estimation of the examinee's ability is then updated after each test item is given. The new estimate is based on the original estimate of ability and previous answers given by the examinee. Based on the new estimate of ability, the computer attempts to select as the next item one that is the most discriminating among examinees near that point on the ability continuum.

C. DEVELOPMENT OF AN ADAPTIVE TEST

The procedural requirements for the development of a CAT include developing the item pool, selecting a procedure for administering the test, obtaining software and hardware to develop and administer the test, and evaluating the results of the test.

Time, manpower, and money will be required to successfully implement a CAT. Although a thorough discussion of each of these latter points is beyond the scope of this paper, the reader should realize that such factors as software development time, the cost of labor, and the purchasing of new equipment will effect the overall development of any CAT project.

1. Item Pool Development

Selection of items to constitute an adaptive test item pool is a larger undertaking than choosing items for a conventional test. Since adaptive testing involves selective administration of a small subset of a larger item pool, the item pool should be large enough to function effectively. (U.S. Army Research Institute Rept. 423, 1979, p. 28)

The development of the item pool consists of generating test items for use, and then administering the items in either a pencil and paper format or a computer format. Large numbers of test items used in conventional tests may not meet the criterion for inclusion in an adaptive test, and in many cases it may not be feasible to construct an adaptive test item pool from off-the-shelf test items. However, where large scale testing programs are already in progress, such as in military testing, current and obsolete test items should contain a sufficient number of items from which to select questions to constitute a satisfactory item pool. This will help to reduce development time and costs. (U.S. Army Research Institute Rept. 423, 1979, p. 29)

Item administration can take place by using multiple forms of the same test, and gathering the results on an operational basis, over a period of time; alternatively the item pool could be administered in a non-operational setting. Once the item pool administered, the items must be calibrated. Item calibration refers to the estimation of the parameters (difficulty, discrimination, guessing) of each item's characteristic curve (U.S. Army Research Institute Rept. 423, 1979, p. 30).

2. Test Administration

Test administration is composed of several parts. After a test item is selected for use, it is retrieved from the computer memory or storage device. The item is then placed on the screen, the examinee's response is read and scored, and the examinee's ability is re-estimated.

3. Software Alternatives

The software required to support a CAT is unique in many ways, and can be acquired through in-house development, acquired from another government agency,

and/or purchased from a vendor. The reader should be aware that in-house development costs can be very high, and that the cost of maintaining that software, in terms of both money and personnel, can also be very high (Pressman, 1982). Chapter II will present a detailed description of software requirements for a CAT.

4. Hardware Alternatives

Hardware will also be required to develop a CAT. The developers of a CAT will want to utilize any computers already in use in a command to the fullest extent possible, but it must be remembered that the hardware selected must be able to fully support the software packages being used. Chapter III will discuss hardware requirements in detail.

5. Test Evaluation

In order to ensure the reliability of the adaptive test, it must be demonstrated that the scores at one point in time correlate well with scores at another point in time, or with scores obtained using different items. Also if a CAT is going to be used to replace an old exam it must be shown to be scored on a scale equivalent to the test it is replacing. This will allow the new test to be introduced smoothly without disrupting any ongoing process, such as the flow of recruits into the service, or of students completing a training course.

A CAT can be used in several ways, such as to predict a person's performance or to assess the outcome of a training course. As an example, a CAT can be used to screen recruits for the service, and to help select them for follow-on training schools. In this respect it is being used to predict an examinee's future performance. If the CAT is replacing an old pencil and paper exam, it must be shown to predict job performance at different cut off points with a precision equivalent to the test it is replacing. A CAT can also be used to assess the outcome of training or formal schooling, such as testing students at the end of a lesson or before graduation from a school. As with pencil and paper tests, the content of the questions answered by each examinee must be shown to be representative of the full range of material to be learned.

D. PURPOSES OF THE THESIS

This paper is geared towards assisting a person who is considering the use of computerized adaptive testing in his command. It will provide an analysis of the minimum requirements of a system needed to support a CAT.

1. Operational Considerations

An essential part of a CAT system is the set of operational constraints under which it is employed. This section will specify a realistic set of assumptions about these constraints and thus form a background for requirements which will be discussed in subsequent chapters.

The system developer may choose between several options concerning the type and the number of computers for use. Micros, minis, or mainframe computers could be used with adaptive testing. This paper will consider only the use of microcomputers because they are the least expensive and the most readily available type of computer. The maximum number of examinees to be accommodated at one time is assumed to be driven by the amount of available (old plus newly acquired) hardware to be used as testing terminals.

Both the equipment and examinees will require support during a test. A dedicated space is assumed to be available to provide a secure location to store item pool data. If the equipment is not portable, a dedicated testing space is assumed. A dedicated space is not necessary if portable equipment is provided. The temperature of the testing space is assumed to be controlled within the limits of comfort to avoid distraction from the test. Lighting is assumed to be located so as not to produce eye fatigue. This can be accomplished by ensuring no strong lights are located behind the examinee terminals. The CRT display is assumed to be glare free, or adjustable by varying the angle of the screen, and the system is assumed to have an uninterrupted power source with a constant voltage in order to minimize the chance of damage to the equipment or loss of data.

The equipment and software will require maintenance. The equipment is assumed to have components that are easily replaceable if they breakdown. Copies of the software are assumed to be available in order to be used as a backup; and a proctor, with available documentation of hardware and software, is assumed to be able to make replacements and adjustments of hardware, and to specify parameters of the software as necessary, e.g., in order to prevent the examinee from taking the same test items twice.

It is also assumed that the system will be user friendly and have adequate documentation, so that special training or course work for proctors will not be required. In addition, it is assumed that a part time person with the necessary training in starting the machine, answering questions, and trouble shooting minor problems,

could be used as a proctor. This person could handle the proctor job and another job at the same time.

2. Specific Purposes of the Thesis

Chapter II will discuss software requirements of a CAT system. This will include discussion of requirements of the various components of an adaptive test problem, such as selection of items to be tested, and the scoring of the exam. Given the operational constraints specified in Chapter I, one or more software packages are needed to fulfill the system requirements.

Chapter III will discuss some of the functions that must be supported by the system's hardware. It will also review how various factors such as portability, communications, and networking can affect normal operations, protection against systems failure, and protection against possible security breaches.

Chapter IV will present a detailed look at the experimental CAT system that is currently being used by NPRDC for research on the CAT/ASVAB. This chapter will review the system and see how well the hardware and software in use meets the requirements of computerized testing as they are described here. The experimental CAT system was chosen for review because of its extensive use to collect data prior to developing the CAT/ASVAB for use in an operational environment. Also, it has sufficient documentation to allow it to be evaluated using the criteria specified in Chapters II and III.

II. SOFTWARE REQUIREMENTS

One of the most important parts of any CAT system is the software used to support the test. Before selecting any software package for use in an adaptive test, it is imperative that the developer have an understanding of the functions the software must support and how the software will support those functions. This chapter discusses some alternative approaches to developing and administering an adaptive test, the minimum requirements the software will have to support, and the storage requirements for the software.

A. TEST DEVELOPMENT AND ADMINISTRATION REQUIREMENTS

As noted earlier, most research in the adaptive testing arena has focused on the three parameter model of item response theory. For a detailed description of item response theory see Green, Bock, et al (1984, p. 348). While most of the current work being done in adaptive testing is in the three parameter IRT, there are other models in use. For example, the one parameter Rasch model has no guessing parameter; all item have equal discrimination power. The two parameter normal ogive model also has no guessing parameter. These models are less mathematically complex and result in faster computation, however, they make stronger assumptions about the item characteristic curves, and the procedures required to implement them in practice are different. Before selecting a particular model for use, the system developer should consider the appropriateness of a particular model, and plan to study the invariance of the resulting estimate of ability. (U.S. Army Research Institute Rept. 423, 1979, p. 5)

1. Development of the Item Pool

One of the primary requirements in a good CAT is a large well-developed item pool with well established item parameters (Green, Bock, et al, 1984, p. 357). Selecting the items to constitute an adaptive testing item pool is a somewhat larger undertaking than choosing items for a conventional test. The criteria for item selection and for pool construction are more rigorous than those for conventional test design, and the item pool must be substantially larger than the length of any individualized test drawn from it. For example, an experimental CAT/ASVAB subtest contains 300 questions in its item pool (NPRDC Rept. 84-33, 1984, p. A2). Since the degree to which an adaptive test realizes its potential may be limited by the size and quality of its item pool, it is

imperative that the item pool contain the necessary desirable characteristics (U.S. Army Research Institute Rept. 423, 1979, p. 8).

Once items are selected for consideration for use in the item pool, they must be evaluated empirically before they are placed in an actual test. The administration of the initial items can take place in a practice setting, or in an operational setting. The evaluation can take place in a practice setting by using the items to construct practice exams to be given to a group of students on a trial basis, without having the exam count for any grade or selection process. If the evaluation of the items takes place in an operational setting, certain questions to be evaluated could be mixed into the normal questions of an actual exam without the examinee's knowing it. These questions would not be counted towards the examinee's grade, but could then be evaluated as to their relative merit. Thus you save time and money by evaluating the items in an operational setting. The administration of the new test items can take place via a conventional pencil and paper exam or by computerized administration. Caution must be exercised, however, because the item parameters may be less accurate with pencil and paper administration of the exam.

The calibration of each test item is required before the item can be used in an adaptive test. Calibration refers to the estimation of the parameters (difficulty, discrimination, and guessing) of each of the item's characteristic curve. After sample questions are gathered for consideration, examinees must take the items before they can be used in an actual adaptive test. While there are no definitive studies to recommend a number of times to test an item, a good rule of thumb for item parameter estimation is at least 1000 examinees and at least 20 items per calibration. This will yield 20,000 data points, which are used in a computer program to estimate the parameters for each item. After the parameters for each item are assigned, the item pools can be constructed to ensure that there is a proper mix of questions with different item parameters within each item pool.

Several computer programs are available to estimate the parameters of an item's characteristic curve, for both mainframe and microcomputers. "ASCAL", developed by Vale for microcomputers and mainframes, and "LOGIST", developed by Lord for mainframes, both use the maximum likelihood estimation (U.S. Army Research Institute Rept. 423, 1979, p. 11). "BILOG", developed by Mislevy and Bock for mainframe and microcomputer use, uses the Bayesian estimation for estimating the item parameters. (Bock, Mislevy, 1982) Whichever software package is selected for

item parameter estimation, it should be able to estimate the parameters of at least 20 items with at least 1000 examinees at a time in order to provide statistically reliable estimates of the item parameters.

In addition to item pool development, there are several other important requirements that the software package must be able to meet. These include the ability to place an item on the screen without scrolling, enabling the examinee to read the entire question without moving the text up or down on the screen. Another requirement is the storage of the examinee's response without the risk of loss of the data file if the test is given in an operational setting. It is also important that the software provides rapid retrieval of items from the computer's random access memory (RAM) or disk drive. If RAM is used, the amount of RAM needed to store one item of text can be as much as 1.4 kbytes.

2. Scoring of the Test

Adaptive tests have different people taking different sets of test items. Because of this, the scoring method needs to account for not only how many items a person answers correctly, but also which items were answered and whether each was answered correctly or incorrectly. (U.S. Army Research Institute Rept. 423, 1979, p. 20)

Two alternative approaches to scoring of the items can be used, the maximum likelihood or Owen's Bayesian technique. Both the maximum likelihood and Owen's Bayesian sequential procedure are methods of estimating an examinee's location on an ability continuum. There are, however, differences in each approach.

Owen's Bayesian procedure estimates the examinee's location sequentially. It begins with an assumed normal distribution of the person's ability and updates that estimate, one item at a time, by solving equations that consider both the likelihood function of the single item score and the assumed normal distribution. The ability estimate is the final updated value after the last item score is considered. One disadvantage of Owen's Bayesian procedure is that it is order dependent, while one advantage is that it automatically includes a measure of precision of the ability estimate that may be used to decide when to terminate the test. An alternative Bayesian procedure which is not order-dependent is also available. (Bock, Mislevy, 1982)

The maximum likelihood procedure estimates the examinee's location parameters from the pattern of the examinee's right or wrong answers by solving a likelihood equation. No prior assumptions are involved regarding the examinee's

location on the ability continuum or the distribution of the attribute. One disadvantage of the maximum likelihood procedure is that it is not usable if the answers given are all correct or all incorrect, which can easily be the case early in the sequence of items.

The software estimates the examinee's location between items and also at the end of the test. It should be noted that the final score estimate does not have to be the same as the sequential estimate. For example, the maximum likelihood estimate may be used at the end for the final scoring of the exam, even though the Owen's Bayesian procedure was used during the exam.

The minimum software requirements for scoring of the test include the rapid retrieval of item parameters from RAM or disk, and the rapid computation of the score estimates during the test. The speed of the final scoring computation is less important because the examinee is not waiting for another item while his score is being computed. The storage of the current score on a disk without a loss of information in the case of a system failure is also an important requirement for the software.

3. Selection Of The Next Item

Because adaptive testing tailors each test to the individual's ability, the selection of the next item to be given to the examinee is an important part of the system's software package. There are several alternate methods that may be used for item selection. The maximum information method attempts to select the next item that will be most informative, that is the one that has the highest discrimination parameter. The Bayesian strategy will select an item which will minimize the posterior variance of the examinee's ability distribution. The minimum software requirements for the selection of the next item include the rapid computation of the selection criteria, or sufficient RAM in order to store a table of selection criteria, e.g., an information table, and then have rapid retrieval from that table.

Exposure control, which prevents each question from being seen by a large proportion of examinees, is another feature to be provided by the software. This is important for items early in the testing sequence, and it may be provided with the aid of a random number table or generator in the software.

Finally, the software package must be able to determine when to stop the exam. There are several methods that can be used, including specifying a level of precision or by using a test of fixed length. Scoring procedures not only make it possible to estimate ability levels after each item is administered and answered, but also

make it possible to determine the precision of each ability estimate. This can then be used as a criterion for termination of the test. (Weiss, 1982) Alternatively, some adaptive test strategies use a fixed test length as a stopping rule. In this case the test is terminated when the examinee has answered some fixed number of items (U.S. Army Research Institute Rept. 423, 1979, p. 6).

4. Graphics And Color Displays

Certain types of adaptive tests have the need to display graphics as a significant portion of the test. For example, the ASVAB contains three subtests, the auto and shop information, mechanical comprehension, and the electronics information test, all of which contain diagrams and graphics displays. (Green, Bock, et al, 1984, p. 7) The graphics displays require a large amount of memory in order to display the figures quickly and with high quality. Color displays, while not presently in wide use for adaptive testing, also require a larger amount of memory.

The minimum software requirements if graphics are required for the test include the rapid retrieval of pixel level array from RAM or disk, the construction of the image in RAM, and the placement of the image on the screen without scrolling.

5. Summary Of Storage Requirements

Adaptive tests will require large amounts of storage to handle the item pool and item selection algorithm. During the exam, the response time must be fast. The timely execution of the item selection algorithm will help to avoid distracting delays for the examinee (McBride, Moe, 1986, p. 21). The fastest response times can be obtained when the item pool, parameters, and information table are in RAM. As an example of the amount of RAM required for a 100 item pool of written text items, with 25 levels of ability in the information table, and with 3/4 of a 80 x 24 character screen being used:

$100 \times (3/4 \times 80 \times 24) = 144,000$ characters (bytes) for items, and

$100 \times ((3 + 25) \times 100) = 2800$ single precision numbers
with 4 bytes per number, for parameters and the information table,

which equals 155,200 bytes of storage required.

B. COMPILER AND OPERATING SYSTEM REQUIREMENTS

Software requirements will also include the need for a compiler and operating system which will support the software described in section 2 above. The compiler is the software program that converts the high level computer language, such as Pascal, to the machine level language that can be understood by the microcomputer. The operating system is the set of programs that manages the computer's memory, processor, and other resources. Some examples of CAT systems and their operating systems are noted as follows.

The experimental CAT/ASVAB described in Chapter IV uses the Apple operating system and the computer language UCSD Pascal, which is a machine independent, structured language (NPRDC Rept. 84-33, 1984, p. 1).

The Hewlett Packard Integral personal computer, which is being tested for use in a portable CAT system, uses the Unix operating system and the C computer language.

The prototype CAT system described by McBride for use in elementary schools uses the Apple II microcomputer with the Apple operating system (McBride, Moe, 1986, p. 1).

ACT Corporation's adaptive testing system uses an IBM PC with the DOS 3.1 operating system. ACT is currently gathering research on different compilers, and is considering both Fortran and C computer languages for use in the system.

Assessment Systems Corporation's adaptive testing system uses an IBM PC with the DOS 2.1 operating system and the Pascal computer language.

III. HARDWARE REQUIREMENTS

Once a software package that will support adaptive testing has been selected or written, a hardware system must be assembled that will adequately support the software. The proliferation of hardware vendors, the ability to select components from a variety of sources, and the frequently decreasing costs of hardware all combine to make the selection of the best hardware system even more difficult. This chapter explains some of the options that can be considered before making a decision to utilize a particular piece of hardware.

A. HARDWARE ISSUES

Rather than purchase new equipment specifically for use in a CAT, most commands will want to make the best use of equipment already on board in order to reduce costs and to allow personnel to work with equipment that they are already familiar with. By conducting a detailed inventory of all hardware already being used, the command will be able to identify the type of software the hardware will be able to support, and be able to estimate the amount and type of new hardware purchases that will be required.

One point to keep in mind when deciding how many testing stations to install is that operational considerations during peak demand will drive how many stations will be required. The number of stations will effect the total cost, storage requirements, and the software development.

The system developer should also remember that peak demand could be lowered by staggering the scheduling of tests. This fits in well with CAT's because each examinee is not required to begin and finish testing at the same time as other examinees, due to the individualized nature of the adaptive test.

B. SPECIFIC HARDWARE REQUIREMENTS

The key requirement for any hardware that is selected for use in an adaptive system is that the hardware must be able to support the software requirements of available, useful software packages.

1. Specific Functions To Be Supported

Specific functions to be supported by the hardware include block testing for the development of the item pool, and adaptive testing using the item pool.

Block testing occurs during development of the item pool, and is used to establish item parameters. The hardware must be able to support administering the item pool in a conventional mode in order to check item parameters if the items have been previously calibrated on a pencil and paper examination, or in order to establish and calibrate the item pool parameters if the items to be used are new ones.

The hardware must also be able to support the system when it is used for testing in the adaptive mode. As noted in Chapter II, and as indicated below, this will require adequate storage and may require graphics capability.

2. Specific Hardware Options To Choose From

Once a suitable software package has been selected, there are several hardware options that can be considered for use in the adaptive test system. These options include a fixed standalone machine, a portable system, the use of communications equipment to link a testing station a remote site, and the use of networking to connect several testing stations together. This paper will assume that the minimum system requirement is a single microcomputer capable of handling the necessary software. However, the hardware alternatives noted above will also be discussed briefly, in terms of their differences in memory requirements during testing.

3. Fixed Stand Alone System

a. Procedures Necessary For Normal Operations

Normal operations of a fixed, stand alone microcomputer configured for a CAT would require a hard or floppy disk to retain data, scores, item pool, and the scoring and item selection software. The disk used must have sufficient storage capacity to be able to retain all applicable information at the end of the testing sequence. Volatile memory requirements for normal operations include the need to store the compiled program, load module, and test data files, which include the item parameters and the information table.

b. Procedures Necessary For Protection Against System Failure

The disk should retain data, such as a list of items that have been administered and the responses to them, during testing to preclude the loss of test data if the system fails during an examination. In addition, the proctor must be able to re-load and re-boot the system if necessary.

c. Procedures Necessary For Protection Against Security Breach

Individual floppy disks which contain test items must be strictly accounted for in order to prevent a compromise of the examination. Also, any terminal keys not used during the exam should be locked out in order to prevent tampering with the test by the examinee; or, if funding allows, a special examinee input device could be used. (See Chapter 4.A.5)

d. Memory Requirements

A fixed, stand alone system being used for adaptive testing requires that the compiled procedures for item selection, item administration, and scoring be stored in volatile memory during testing. The item parameters and information table are stored in volatile memory to enable the rapid scoring and item selection between test items; the item contents may be stored in volatile memory or on a hard disk during testing. The record of the items administered, the responses to those items, and the final score of the test may be written to a floppy or hard disk during the testing and not be retained in RAM. Also, the source of the load module may also be read from or written to floppy or hard disks during testing and not be retained in RAM.

4. Portable System

a. Procedures Necessary For Normal Operations

Portable equipment must be quite rugged in order to withstand the rigors of constant moving about. For this reason hard disks should not be used because they are not generally durable under conditions involving frequent movement. Care should also be taken to ensure that the equipment selected is truly portable and easy to move. Physical size, weight, and durability of the equipment must also be considered before any hardware selection is made.

The requirements for volatile memory and floppy disk capacity would be the same as for the fixed system previously described, except that the functions served by a hard disk in the fixed system need to be served by the volatile memory or the floppy disks of a portable system. Also, the item bank needs to be in volatile memory in order to provide for rapid access.

b. Procedures Necessary For Protection Against System Failure

The system must be able to store all test data on the floppy disks in order to prevent the loss of data in the case of system failure. It should also be able to be re-loaded and re-booted by the proctor after a failure.

c. Procedures Necessary For Protection Against Security Breach

Because of the portable nature of both the equipment and the software, the proctor must ensure that adequate security is provided in order to prevent theft or compromise. This includes ensuring security of the system as well as of the floppy disks.

d. Memory Requirements

A portable system being used for adaptive testing requires that the compiled procedures for item selection, item administration, and scoring be stored in volatile memory during testing. Also, the item parameters, information table, and item contents are stored in volatile memory during testing. The items administered, the responses to the items, the final scores, and the source of the load module are stored on floppy disks.

5. Communications

a. Procedures Necessary For Normal Operations

Testing may take place at a remote site, with test results being forwarded after the exam. One example of this is the Marine Corps project noted in Chapter I. Examinee's take a CAT at the Marine base in Twenty Nine Palms, CA, and the results are transmitted via phone lines to the ACT offices in Iowa City, Iowa, where the results are analyzed. ACT personnel can also run the system from Iowa City in order to trouble-shoot any problems that develop.

A functional advantage of communication is the ability to upload and download software, item pool, and test data both before and after an examination. This lessens the requirements for non-volatile memory, while having the same RAM requirement. However, hard or floppy disks are still needed to prevent the loss of test data in the case of a system failure, and testing time may be too long if items and responses are transmitted during the test because of transmission delays between the remote site and the testing center.

Additional software is required to support the communications function, thus somewhat more volatile memory is required.

b. Procedures Necessary For Protection Against System Failure

If a failure occurs, a floppy or hard disks is required at each testing station in order to prevent the loss of test data. Also, the proctor must be able to re-boot the system, reloading can be accomplished via the communications link.

c. Procedures Necessary For Protection Against Security Breach

The necessary procedures are the same as for the fixed system previously described, except that items need not be on the disk at the examinee's station.

d. Memory Requirements

An adaptive testing system that uses communications requires that the compiled procedures for item selection, item administration, and scoring be stored in volatile memory during testing. The item parameters and information table are stored in volatile memory, while the item contents may be stored in volatile memory or can be read from hard disk during testing. The items to be administered and the response to the items are read from and written to floppy disks, while the final scores and the source of the load module may be sent to or received from a remote site.

6. Networking

Another option for a testing configuration is a network. A computer network is established when two or more computers are interconnected via a communications link (Stallings, 1985). Several terminals may be linked to one computer which controls the examination and store the test data. Chapter IV describes the experimental CAT ASVAB system, which uses a partial network configuration.

a. Procedures Necessary For Normal Operations

Networking may be cost efficient in situations with a large number students being tested in a fixed location on a regular basis. A network system may include two or more testing stations linked together by telephone lines or hard wired, and connected to a master station or terminal, from which the proctor can run the examination. Before testing begins, each individual station is loaded with the test data by the proctor, and all stations are given a systems check to test for component failure and to synchronize internal clocks. It is necessary to periodically save test data by transferring data from the individual station to the proctor's station both during the test and at its conclusion.

b. Procedures Necessary For Protection Against System Failure

A failure in the proctor's station could effect all examinee's stations, therefore the hardware should be configured to allow for storage of data from all examinee stations at the proctor's station in case of systems failure.

c. Procedures Necessary For Protection Against Security Breach

These requirements are the same ones as required by the communications system.

a. Memory Requirements

An adaptive test system that uses networking requires that the compiled procedures for item selection, item administration, and scoring be stored in volatile memory during testing. The item parameters and information table are stored in volatile memory, while the item contents may be sent to or from a remote station during testing. The items that are administered and the responses to the items may be read from or written to floppy disks during testing, or they could be sent to or from a remote site. The final scores and the source of the load module may be sent to or from a remote site or the proctors station.

7. Visual Display

In order to reduce fatigue and prevent distractions, quality equipment is needed to present the test material in a clear, high resolution image. Pixel size is a characteristic to consider when selecting CRT screens for use in a testing environment. The number of pixels per square inch will determine the quality of the screen picture. This will be an important factor when a test involves graphic displays, such as the ASVAB, because the graphics display will require high resolution and quality in order to be accurate.

Other factors that should be considered are the numbers of lines and columns on the screen for presenting text and the overall legibility of the screen. The CRT display should be adjustable so that students are able to tilt the screen in order to reduce glare and eye fatigue; also, the use of glare shields on the screens will help to make the examination more readable.

8. Microprocessor

The microprocessor chip is the heart of any microcomputer, and the type of chip used will determine the speed and storage capacity of the computer. The microprocessor also is responsible for memory management capacity and graphics management. It is important that the chip have a memory management capacity sufficient to support the volatile memory requirements that were noted earlier. When graphics are used, the microprocessor must rapidly build up the graphics image in memory, then move the image to the screen for display. Examples of microprocessors used in CAT systems are: The IBM PC microcomputer used in ACT Corporation's CAT project uses the Intel 8088 processor, with an Intel 8087 numeric coprocessor, and has 256 kbytes of RAM with one 360 kbyte disk drive; The Hewlett Packard Integral Personal Computer in the operational CAT ASVAB project, which uses the Motorola 68000 microprocessor chip, with a 16 mbyte capacity.

9. Special Response Input Devices

The use of special response input devices can help to simplify the examinee's task, and are suitable for multiple choice questions. McBride's description of a CAT prototype system which, used the Apple II microcomputer, noted the beneficial use of two prominent labels with "yes" and "no" printed on them. In this particular system, the examinee moved an arrow by answering yes or no to the question "is this the correct answer?" until the arrow pointed to the answer the examinee thought was correct. (McBride, Moe, 1986, p.4) The CAT/ASVAB system described in Chapter IV also uses a special response input device in the form of a keyboard cover. The cover allows only certain keys to be used by the examinee, which prevents unauthorized tampering with the system and makes it easier for the examinee to enter answers into the system. This makes the system more user friendly, which is particularly helpful for those students with no experience in using computers.

10. Surge Protection

Equipment must be protected from variation in the power supply in order to prevent damage to hardware and loss of software and test data. Filters can be used to level out fluctuations in power that may harm equipment or cause the computer to lose data. They are considered a mandatory piece of equipment whenever an unstable or portable power source is used.

IV. EXAMPLE OF A COMPUTERIZED ADAPTIVE TESTING SYSTEM

A. THE EXPERIMENTAL CAT/ASVAB

1. Description Of The ASVAB

The ASVAB is a test given to all potential recruits before they are selected for enlistment in the armed forces. It consists of ten separate subtests, which test the examinee's knowledge in areas such as general science, paragraph comprehension, and arithmetic reasoning. It is of critical importance to the services because the scores help to determine who will be selected for enlistment, and to determine the enlisted specialties, follow-on training, and advanced schooling the enlistees will receive. In the conventional pencil and paper format, the ASVAB consists of 350 questions, and takes about four hours to complete.

Currently, a joint-service project is underway to develop a CAT system to support the mission of the ASVAB. The Department of the Navy is the lead service for the project, and NPRDC is the lead laboratory. If this experimental test project, which is described and discussed in this chapter, proves to be successful, an operational adaptive test could be used to replace the conventional pencil and paper ASVAB. (NPRDC Rept. 84-32, 1984, p. 1)

2. Minimum System Requirements

In addition to satisfying the operational constraints that were discussed in Chapter I, the experimental CAT/ASVAB system has been developed to support the following minimum requirements.

The Apple microcomputer system used for the test is connected to a Corvus hard disk and can be configured to give up to 20 subtests in any order, and each subtest can contain up to 20 questions.

The test is self-instructional and friendly. All examinees are presented with a familiarization session which they must pass before they proceed with the examination. If they can not pass the self-instruction session, the proctor is automatically called.

The system keeps track of how much time the examinee has spent on the familiarization session, the subtest instructions, each subtest, and in the entire session. It also tracks the number of times the proctor is called to assist an examinee for each subtest, and for the entire test.

Scoring results can be provided to the examinee at the end of a question, subtest, or at the end of the session.

Minimal data loss occurs if there is a loss of power to the system. If it does occur, power loss will result in all examinees losing data for the current subtest they are taking. When power is restored, examinees must log in again, and repeat the subtest they were taking when the power failure occurred. Files are updated at the conclusion of each subtest in order to save test data on the Corvus hard disk.

Up to 21 item pools can be created. Each item pool is a subtest that may contain up to 300 questions. All items in the subtest, including instructions, samples, and questions, can be modified or deleted as necessary.

System security and graphics support are also included. Security is provided by preventing the examinee from logging on until the proctor has authorized the log on, while graphics capability is provided by incorporating a graphics editor into the system.

Overall system friendliness is provided by using simple menus and providing instructional prompting from the computer. (NPRDC Rept. 84-33, 1984, pp. A1-A3)

3. Setting of the Experimental CAT/ASVAB

The proctor is a key part of the experimental CAT/ASVAB system, and is responsible for setting up the equipment and administering the exam. A user's manual is provided which provides step by step instructions to guide the proctor in use of the equipment and testing procedures, and no formal training is necessary.

The test site is a dedicated room that contains all the necessary test equipment. Seven testing stations are aligned in a row and are connected by cables to the Corvus hard disk. The test area is physically separated from the equipment area, which contains the Corvus disk, multiplexer, and printer, in order to reduce noise and distractions for the examinees. A sound screen is used as wall to separate the two areas.

Once the equipment is set up following the instructions in the user's manual, testing may begin. The proctor follows the startup procedures, which includes checking the status of the Apple computers, setting the Apple's internal clocks, initializing the system, loading the operating system, and logging personal data on the examinees, such as name, social security number, and date of birth. After the examinees are admitted to the test room, they are given a brief introduction to the system. When the introduction is complete the examinees begin the test, and the proctor remains

available to answer questions or handle any problems that arise. When testing is complete for the day the proctor follows the procedures in the user's manual to secure the equipment.

4. Software Used by the Experimental CAT/ASVAB

The experimental CAT/ASVAB system uses several of the theoretical options that were discussed earlier in Chapters I and II. The three parameter item response theory model is used for the item characteristic curve; however, the parameters can not be estimated by the system. The system is not intended for use in developing an item pool. At the very least, the development of an item pool requires the use of other software and hardware assets to estimate parameters of the items. The maximum information rule is used for item selection; and Owen's Bayesian procedures are used to calculate the examinee's ability between items and at the end of testing, and to provide a measure of the variance of the ability estimates. Exam termination can be determined by the minimum variance rule: when the standard error of the examinee's ability decreases to a predetermined level, the test can be terminated. Another option available for ending the test is when the examinee completes a fixed number of questions, which can range as high as 20.

As noted earlier, the CAT/ASVAB system uses the UCSD version of the Pascal computer language. The software system is composed of seven programs which handle all aspects of the test, from administration to diagnosis. The complete listing of the Pascal program is available for reference (NPRDC Supp. to Rept. 84-33, 1984).

The test administration program gives the examinee a practice session to familiarize the student with the computer system, presents general instructions and log in procedures, and administers the test. After administering a question, the program updates the examinee's ability level. The files containing examinee test scores are updated after every subtest, so if a power loss or crash of the system occurs, the examinee will have to log on again and repeat only the subtest that was being used when the interruption occurred. This program also terminates the exam by using the minimum variance rule, or by giving only the specified number of items in the case of a fixed length examination.

The configure test parameters program is run at the beginning of the testing day, and allows the testing parameters to be set up. These parameters include the ability for any combination of subtests to be selected in any order, and allows up to 20 questions per subtest to be given. It also allows for a delay after each subtest if it is needed, and for establishment of the feedback parameters that will be used.

The test manager program maintains a data base of up to 21 subtests and their item pools. The program can create, list, and delete subtests from the data base, and can also transfer a subtest from the Corvus disk to a floppy disk. The program can also insert, modify, or delete questions and instructions from the subtest.

The examinee data manager program maintains and provides access to information on up to 50 examinees. It includes the ability to enter the examinee's personal data, log the examinee on to the system, and can list the status of each examinee as to whether they are partially complete or finished with the subtest.

The strategy data manager program provides for maintenance and access to the information tables that support the adaptive test. For a given level of ability, the information table lists the items by identification numbers in order starting with the most discriminating item.

The graphics editor program allows for the construction of graphics to support subtests that require them. After specifying the subtest to be modified, the program will allow certain options to be selected in order to modify the graphics package.

The diagnostic program allows the proctor to verify system stability by checking information tables, graphic questions, and subtest questions. If any errors are found, the location of the error is noted and an error listing can be printed for reference. (NPRDC Rept. 84-33, 1984, pp. A11-A16)

5. Hardware Used by the Experimental CAT/ASVAB

The experimental CAT/ASVAB is configured in a network system, with seven testing stations linked together. The system is also designed to be sufficiently portable so that it may be moved between military bases every few months; however due to the number of individual components and the time needed to assemble and disassemble the system, it is not considered truly portable.

Commercially available hardware was selected for the experimental CAT/ASVAB. It consists of an Apple III computer with Sanyo video screen, a Corvus disk drive, a Corvus constellation multiplexer, a Panasonic videotape recorder, and two Topaz voltage regulators. Additional equipment used to support the system includes floppy disks, videocassette tapes, glare screens, power strips, extension cords, ribbon cables, video recorder cables, video output cables, power cords, and disk transfer containers.

The Apple III computer used for the CAT/ASVAB is programmed to to present the adaptive test on the video screen, receive the responses from the examinee, and calculate the test scores. The Apple III must have at least 256 kbytes of memory available, and be equipped with a Thunderclock Plus timer. The Thunderclock is a commercial product used by the Apple computer to track item response time. The keyboard in use has been designed for CAT administration, and each keyboard has a temporary cover on it that permits the examinee to press only certain designated keys.

The Sanyo video screen is placed on top of the computer, and displays questions, presents instructions, and lists test results. Each video screen has a glare shield attached to it, which helps to reduce eye strain and fatigue. The video screen can also be adjusted for brightness and contrast in order to present the test as clearly as possible.

The Corvus disk drive is programmed to collect and store the information obtained from the computers, and differs from the disk drive located in each computer. The Corvus drive contains the program source and data files necessary to administer the test for all the testing stations, while the disk drive in the Apple computer is used to check the status of the computer, run internal trouble shooting checks, initialize the computer's internal clock, and load the operating system. The location of the source and data files on the Corvus disk also contributes to system security because the files are not accessible to examinees and are not easily down-loaded to the individual Apples due to the presense of keyboard covers and the proctor. The Corvus disk must have a minimum of 10 mbytes of storage, and can be linked to as many as eight Apple microcomputers. The Corvus Constellation multiplexer coordinates communications between the Corvus disk drive and the Apple computers. It determines the order in which the computers will communicate with the Corvus disk.

The Panasonic video recorder is used as an auxiliary backup in the event of a power loss or other system failures. It can record and store information as instructed and therefore be used to transfer the data files to other computers if the original system goes down. (NPRDC Rept. 84-32, 1984, pp. 3-8)

The Topaz voltage regulators are used to stabilize the electric current, which may come from an external line or an internal generator. They are used to protect the computer system in case of a surge or an overload of the current.

6. Summary

The experimental CAT/ASVAB, which consists of a specially designed software package and commercially available hardware, is currently being evaluated at NPRDC. Preliminary results of the evaluation are encouraging, and an operational version of the CAT/ASVAB may be in use soon.

V. CONCLUSIONS

The computerized administration of adaptive tests has a promising future. Although continued research is necessary to fully develop the potential of adaptive testing, several projects, in particular the experimental CAT/ASVAB, have shown adaptive testing to be both technically feasible and practical. Also, adaptive testing has many benefits associated with it that make it an attractive alternative to conventional pencil and paper testing. These benefits include reduced administrative time, better differentiation among students of extreme ability, and the immediate scoring, reporting, and recording of test results. Additional benefits are that adaptive tests allow easier and less expensive replacement of examinations, require less time for the examinee to take, and are more secure due to the elimination of test booklets and due to the individualized construction of each exam.

However, before deciding to implement an adaptive test, it is important to understand the technical issues in the merging of software and hardware components into an operationally workable, efficient system. One key to an effective CAT system is a software program that fulfills the necessary system requirements. To select among software alternatives, the systems developer should understand the available approaches to developing a CAT item pool and to administer and evaluating an adaptive test. The hardware that is selected for use must be able to support the software that has been obtained. When a microcomputer based system is used for adaptive testing, the system developer can choose from several hardware options, including the use of a fixed stand alone system, portable hardware, a system that includes communications options, and series of testing stations connected via a microcomputer network.

To use adaptive testing to effectively take advantage of its benefits, it is important to understand the operational requirements of implementing a CAT in a command because they constrain the selection of the software and hardware. Lack of computer experience for test proctors and examinees requires good documentation of the software, and also user friendly software. The lack of a dedicated and secure space for testing limits the hardware choice to systems which can be easily assembled and disassembled, which in turn can limit use of software which calls for extensive networking or communications.

It is important to understand the software requirements of adaptive testing, because they constrain operational options and selection of hardware. The necessity of developing and securing a large item pool requires secure storage of the equipment and disks if a dedicated testing space is not available. Also, the necessity of being able to rapidly access item parameters and item from a large item pool can require the use of a hard disk or a substantial volatile memory.

It is important to understand the hardware requirements of CAT system because they can constrain operational and software options. If budgetary limitations or a judgement of hardware alternatives lead to the using of previously acquired hardware, an operational constraint may be imposed because of the necessity of using a dedicated space and only testing a limited number of examinees at one time due to a limited number of testing stations. Also, budgetary constraints on the hardware may limit software options to what has been or can be developed for that system, although if funding allows, the addition of more volatile memory or a hard disk may expand the hardware capability.

If the software required by the CAT is not already developed and available from a Department of Defense laboratory or a vendor, then designing and maintaining the software necessary to support the adaptive test will be the biggest challenge to the system developer. The question that must be answered first is whether to develop the software in-house or purchase the services of an outside contractor. Due to the shortage of skilled programmers and high costs associated with in-house development, many large scale software development efforts rely on outside contractors. Once the decision is made on how to develop the software, the reader must remember that it is extremely difficult to accurately predict development time and cost of large software development projects. As an example of the time needed to develop an adaptive test, a feasibility study was made in 1978 to see if the ASVAB could take advantage of the growing adaptive testing technology. An interservice coordinating committee was formed to plan for the development and implementation of a CAT version of the ASVAB, and preliminary evaluation of the CAT/ASVAB began in 1982. (NPRDC Technical Note 85-1, 1984, p. 6) The goal of the CAT/ASVAB project is to eventually test all potential recruits for the armed forces using the CAT/ASVAB system. Future CAT projects will be able to benefit from research and lessons learned from the CAT/ASVAB; these lessons may reduce the cost and amount of time needed to develop a CAT for other types of testing projects.

In summary, this paper has described some of the benefits that adaptive testing can provide for a command, while at the same time indicating areas that may cause problems for the development and implementation of a large scale testing system such as this. Knowledge gained from the analysis of the design and implementation of the experimental CAT/ASVAB project described in this paper provides useful guidelines for investigating the replacement of a pencil and paper examination with a CAT, as well as for the implementation of a CAT per se. Although adaptive testing is in its infancy, it has a bright future, and offers many benefits for its users over conventional pencil and paper examinations.

LIST OF REFERENCES

- Bock, R.D. and Mislevy, R.J., "Adaptive EAP Estimation of Ability in a Microcomputer Environment," *Applied Psychological Measurement*, v. 6, Fall 1982.
- Green, B.F., Bock, R.D., Humphrey, L.G., Linn, R.L., and Reckase, M.D., "Technical Guidelines For Assessing Computerized Adaptive Tests," *Journal Of Educational Measurement*, v. 21, Winter 1984.
- McBride, J.R. and Moe, K.C., *Computerized Adaptive Testing: A Prototype*, paper presented at the meeting of the National Council on Measurement in Education, San Francisco, California, April 1986.
- Navy Personnel Research and Development Center Report 84-17, *Computerized Adaptive Screening Test (CAST): Development For Use In Military Recruiting Stations*, by H.G. Baker, B.A. Rafacz, and W.A. Sands, January 1984.
- Navy Personnel Research and Development Center Report 84-32, *Computerized Adaptive Testing (CAT): A User Manual*, by S. Hardwick, R. Copper, L. Eastman, and F.L. Vicino, March 1984.
- Navy Personnel Research and Development Center Report 84-33, *Microcomputer Network For Computerized Adaptive Testing (CAT)*, by B. Quan, T.A. Park, G. Sandahl, and J.H. Wolfe, March 1984.
- Navy Personnel Research and Development Center Supplement to Report 84-33, *Microcomputer Network For Computerized Adaptive Testing (CAT): Program Listing*, by B. Quan, T.A. Park, G. Sandahl, and J.H. Wolf, March 1984.
- Navy Personnel Research and Development Center's Manpower and Personnel Laboratory Technical Note 85-1, *Evaluation Plan for the Computerized Adaptive Testing Vocational Aptitude Battery*, by B.F. Green, R.D. Bock, L.G. Humphreys, R.L. Linn, and M.D. Reckase, November 1984.
- Owen, R.J., "A Bayesian Sequential Procedure For Quantal Response In The Context Of Adaptive Mental Testing," *Journal Of The American Statistical Association*, v. 70, pp.351-375, 1975.
- Pressman, R.S., *Software Engineering: A Practitioner's Approach*, p. 322, McGraw-Hill, 1982.
- Stallings, W., *Data And Computer Communications*, p. 10, McMillan, 1985.
- U.S. Army Research Institute For The Behavioral And Social Sciences Report 423, *Adaptive Mental Testing: The State Of The Art*, J.R. McBride, November 1979.
- Weiss, D.J., "Improving Measurement Quality And Efficiency With Computerized Adaptive Testing," *Applied Psychological Measurement* v. 6, Fall 1982.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142-2 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Chairman, Department of Administrative Sciences, Code 54 Naval Postgraduate School Monterey, California 93943-5000	1
4.	Professor Bruce M. Bloxom, Code 54BL Naval Postgraduate School Monterey, California 93943-5000	1
5.	Professor Michael P. Spencer 5853 Cohasset Way San Jose, California 95123	1
6.	I.T. Brian J. McCormack 3925 Fulton Avenue Seaford, New York 11783	2

END

2-87-

DTIC