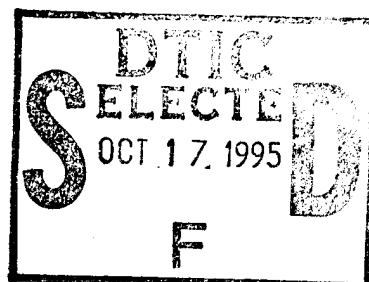




Computerized Adaptive Testing Version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB): Computer System Development



Bernard A. Rafacz

19951012 075

DTIC QUALITY INSPECTED 5

**Computerized Adaptive Testing Version of the Armed Services
Vocational Aptitude Battery (CAT-ASVAB):
Computer System Development**

Bernard A. Rafacz

Reviewed, approved, and released by
Kathleen E. Moreno
Director, Personnel and Organizational Assessment

Approved for public release;
distribution is unlimited.

Navy Personnel Research and Development Center
San Diego, CA 92152-7250

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 1995		3. REPORT TYPE AND DATE COVERED Summary—FY85-FY91	
4. TITLE AND SUBTITLE Computerized Adaptive Testing Version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB): Computer System Development				5. FUNDING NUMBERS Program Element: Reimbursable O&M,A Work Unit: 88WRE5093	
6. AUTHOR(S) Bernard A. Rafacz					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Navy Personnel Research and Development Center San Diego, CA 92152-7250				8. PERFORMING ORGANIZATION REPORT NUMBER NPRDC-TN-95-8	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Chief of Naval Operations Navy Department Washington, DC 20350-2000				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES Functional Area: Personnel Product Line: Computerized Testing Effort: Computerized Adaptive Testing					
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				12b. DISTRIBUTION CODE A	
13. ABSTRACT (Maximum 200 words) In 1979, the Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) program was initiated by the Department of Defense. One objective of the (Joint-Service) CAT-ASVAB program was to develop a distributed processing computer system capable of deploying CAT at aptitude testing sites of the United States Military Entrance Processing Command (USMEPCOM). In 1985, the CAT-ASVAB program was redirected with the initiation of the Accelerated CAT-ASVAB Project (ACAP). The purpose of ACAP was to test the feasibility of using CAT at USMEPCOM testing sites prior to initiation of another full scale development effort. To accomplish this purpose, ACAP had to develop a computer system with all of the essential Joint-Service specifications for deploying CAT. The UNIX-based computer system developed under ACAP meets this goal.					
14. SUBJECT TERMS Computer-based testing, computerized adaptive testing, personnel selection and classification, local area network, distributed processing, UNIX, CAT, CBT				15. NUMBER OF PAGES 57	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED		20. LIMITATION OF ABSTRACT UNLIMITED	

Foreword

This report describes a computer system developed in support of research efforts for the Computerized Adaptive Testing version of Armed Services Vocational Aptitude Battery (CAT-ASVAB) program, specifically, the Joint-Service Accelerated CAT-ASVAB Project (ACAP) (Program Element Reimbursable O&M,A, Work Unit 88WRE5093). The objective of ACAP was to accelerate the operational use of the CAT technology at testing sites of the United States Military Entrance Processing Command (USMEPCOM). ACAP includes development of all the computer hardware and software needed to support CAT at USMEPCOM testing sites in accordance with the functional specifications of CAT-ASVAB, as dictated by the Joint Services in a Stage 2 Request for Proposal, dated 4 June 1984. Parallel efforts, not detailed in this report, include the research to develop the psychometric portion of the CAT-ASVAB tests before installation in the ACAP computer system.

Work on the development of the computer system to support ACAP proceeded along two avenues: The selection, procurement, and installation of the microcomputers needed to automate the CAT-ASVAB tests and the design, development, and testing of the computer software needed to support the CAT-ASVAB functions on the microcomputers. Most of this work was accomplished in-house over a 6-year period beginning in 1985 and resulted in a computer system which was used in the Score Equating Development and Verification data collection phases of the program. Beginning in 1991, the software system was modified in a few critical areas so that it could also be used to support a 1992 start-up of the Operational Test and Evaluation (OT&E) phase of the program; the OT&E phase is not discussed in this report.

The ACAP computer system can support all of the critical functions for automating a CAT technology within USMEPCOM. It is being used to give "scores-of-record" at selected USMEPCOM testing sites. In addition, ACAP has served as a test bed for the practical—and theoretical—concerns of implementing a CAT test. The ACAP computers are also being used in field tests of psychomotor and other computerized tests to increase the predictive validity of the CAT-ASVAB. This effort is known as the Enhanced Computer Administered Tests.

KATHLEEN E. MORENO
Director, Personnel and
Organizational Assessment

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

Summary

Problem

Since 1979, under the Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) program, the Joint Services have been developing a computer system to support the implementation of the CAT strategy at testing sites of the United States Military Entrance Processing Command (USMEPCOM). In 1984, a Full Scale Development (FSD) contracting effort was initiated with the expectation of using extensive contractor support to design and manufacture a unique computer system that could be used at USMEPCOM. In 1985, the FSD effort was terminated and the Accelerated CAT-ASVAB Project (ACAP) was initiated.

Purpose

The purpose of the ACAP was to accelerate the operational use of the CAT technology at testing sites of the USMEPCOM.

This report documents the design, development, and testing of the hardware and software for the ACAP computer system, which was developed with commercially available equipment to serve as the delivery vehicle for the CAT technology within the CAT-ASVAB program.

Approach

To accelerate the software development efforts, the computer hardware, operating system, and programming language were selected early in ACAP. The computer selected for a Local CAT Network (LCN) at USMEPCOM testing sites operates in accordance with Joint-Service specifications. Each LCN consists of a test administrator (TA) station and a sufficient number of examinee testing (ET) stations to accommodate CAT-ASVAB testing at a testing site. Large random access memory (RAM) at the ET stations permits the ET stations to store all of the software and data during the entire testing session.

To have appropriate software development and testing tools, a multitasking operating system and a programming language that can support good programming and documentation standards and provide the tools for effective testing were specified. Different software systems were required for the ET station, the TA station, and the Data Handling Computer. The design process for each system began with the development of functional specifications to establish concurrence with the user and to isolate critical design issues. The software was developed with the use of a main driver program supported by routines that were called into execution as necessary by the driver. Use of the "C" programming language greatly facilitated structuring the software and providing effective testing tools.

Results, Discussion, and Conclusions

The ACAP system effectively meets the requirements for a CAT-ASVAB system of computers to support the CAT-ASVAB technology.

The ACAP system is being used as the delivery vehicle for CAT-ASVAB. In some areas, the ACAP system exceeds the original requirements (e.g., the build time for the display of items

at an ET station). In a few areas, the ACAP system falls short of the requirements, (e.g., providing an automated examinee "help" function).

The most important criterion for evaluating the ACAP system is its psychometric acceptability (i.e., the capability of the system to support the psychometric requirements for implementing the CAT technology into the USMEPCOM testing environment). From 1988 to 1992, the ACAP computer system was used to collect Score Equating Development and Verification data at the testing sites of six administrative segments of USMEPCOM. Over 12,000 examinees were tested using the ACAP computers. None of the data collected seem to have been significantly compromised as a result of collection on the ACAP system.

In summary, the ACAP computer system is a viable delivery vehicle for CAT-ASVAB. However, its use as the instrument for deploying CAT-ASVAB on a nationwide scale is constrained by several considerations. The computer used to support the ACAP system has not been manufactured by the vendor since July 1989. In any event, new and improved testing procedures may require a new hardware system to support future requirements. Also, the CAT-ASVAB program emphasizes the development of psychomotor and other types of computerized tests to improve the predictive validity of the CAT-ASVAB. This will result in new functional specifications and in the eventual selection of a new hardware system to support the new requirements.

Future Efforts

Any future attempts at developing a CAT-ASVAB system should include the supporting studies and documentation necessary for a full-scale development Request for Proposal. ACAP can greatly assist in this requirement. The Joint Services must select a concept of operation for the future CAT-ASVAB system.

The CAT-ASVAB program must decide on a procurement strategy for a future implementation effort. The ACAP strategy of procuring off-the-shelf computer systems should be examined. This strategy avoids the enormous cost, time, and complexity associated with the Government building a CAT-unique hardware system and makes the equipment immediately available for software development.

Any further effort to procure a replacement ACAP computer system should consider the merits of using a large RAM-based design for ET/TA stations. In addition, future software design and development efforts should parallel those used within ACAP, as much as is possible, to make maximum use of software already developed and tested within ACAP.

Contents

	Page
1.0 Introduction	1
1.1 Problem	1
1.2 Purpose.....	1
1.3 Background	1
2.0 Approach and Results.....	2
2.1 Concept of Operation.....	2
2.2 Functional Specifications	3
2.3 Operational Scenario.....	3
2.3.1 MET Site Testing	4
2.3.2 MEPS Site Testing	5
2.4 ACAP Hardware Selection	5
2.4.1 Selection Factors.....	5
2.4.2 Generic Designs	7
2.4.3 Hardware Selected	8
2.5 ACAP Software Development.....	12
2.5.1 TA Station Software.....	14
2.5.1.1 TA Station Operation	16
2.5.1.2 Failure Recovery Operations	21
2.5.2 ET Station Software.....	22
2.5.2.1 Software Design Specifications	22
2.5.2.2 Implicit and Explicit Help Calls	26
2.5.3 Data Handling Computer (DHC).....	27
2.6 ACAP System Documentation	31
2.7 ACAP System Testing Procedures.....	32
3.0 Discussion and Conclusions.....	32
4.0 Future Efforts.....	38
References.....	39
Appendix—An Overview of Computerized Adaptive Testing.....	A-0
Distribution List	

List of Figures

1.	ET station ready for examinee testing.....	10
2.	TA station ready for use	10
3.	Local CAT-ASVAB Network (LCN).....	11
4.	ACAP data communications network.....	13
5.	Flowchart of the TA station functions.....	17
6.	Menu to process stations, sessions, or RASP.....	18
7.	Main menu for administering a testing session.....	19
8.	Sample assignment report	19
9.	Sample screen for collecting examinee testing data.....	20
10.	Examinee Input Device (EID).....	24
11.	Typical CAT-ASVAB test question	25
12.	Main menu of DHC functions.....	29
13.	DHC auxiliary menu	29

1.0 Introduction

1.1 Problem

Since 1979, under the Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) program, the Joint Services have been developing a computer system to support the implementation of the CAT strategy at testing sites of the United States Military Entrance Processing Command (USMEPCOM). In 1984, a Full Scale Development (FSD) contracting effort was initiated with the expectation of using extensive contractor support to design and manufacture a unique computer system that could be used at USMEPCOM. In 1985, the FSD effort was terminated and the Accelerated CAT-ASVAB Project (ACAP) was initiated.

The advent of powerful microcomputer systems on the commercial market encouraged the CAT-ASVAB (Joint-Service) program to pursue the use of such microcomputers to test the feasibility of deploying a computer system to support the functional specifications for CAT-ASVAB.

The implementation issues focused primarily on the psychometric requirements of the CAT-ASVAB system; specifically, the equating of CAT-generated aptitude scores to the paper-and-pencil ASVAB (P&P-ASVAB) (production testing) aptitude scores. Essentially, the Joint Services decided that all the computer support components should be in place so that the psychometric research could be conducted without confounding by factors other than those affecting operational use of such a system. Therefore, the problem was to develop a computer system capable of supporting all of the functional specifications of CAT-ASVAB in a time frame consistent with continued support of the program.

1.2 Purpose

The purpose of the ACAP was to accelerate the operational use of the CAT technology at testing sites of the USMEPCOM.

This report documents the design, development, and testing of the hardware and software for the ACAP computer system, which was developed with commercially available equipment to serve as the delivery vehicle for the CAT technology within the CAT-ASVAB program.

1.3 Background

The original computer system and related computer components for the CAT-ASVAB program were intended to be developed under a major competitive procurement contracting effort consisting of three stages. The system design and prototype development stage extended from December 1981 through October 1983. During this period, three contractors built three prototype CAT-ASVAB computer delivery systems. These systems included the hardware and software needed to demonstrate some of the capabilities for implementing CAT-ASVAB. The result of this effort was a demonstration of the CAT-ASVAB prototype systems for at least some of the CAT functional specifications.

The operational test and evaluation stage was intended to build production quality systems capable of supporting the requirements and specifications in the Request for Proposal document N00123-84-R-0637 (Stage 2 RFP) issued by the Naval Regional Contracting Center (1984). The

Stage 2 RFP was not awarded partly because powerful microcomputers became commercially available. However, the Stage 2 RFP continued to define CAT-ASVAB program requirements. The final stage in CAT-ASVAB system development was concerned with the actual deployment of the CAT-ASVAB system to field locations. The U.S. Army was to be responsible for implementing the CAT-ASVAB system; USMEPCOM would be the consumer.

With the demise of the Stage 2 RFP effort, the emphasis of the CAT-ASVAB program shifted to the development of the computer system for the ACAP. ACAP was initiated by the Assistant Secretary of Defense for Manpower, Installations, and Logistics in 1985 to accelerate progress toward operational use of CAT-ASVAB.

ACAP was tasked to develop a CAT-ASVAB computer system to refine the operational requirements for the eventual system and to complete the psychometric research efforts for the equating of CAT scores with those of the P&P-ASVAB. Therefore, ACAP tried to identify and address these requirements as much as possible in an operational environment. This was accomplished by using commercially available computer hardware in a field-test of CAT-ASVAB functions at selected USMEPCOM sites. At those sites, CAT-ASVAB testing occurred in accordance with the specifications declared in the Stage 2 RFP and subsequent specifications. The new specifications arose from psychometric developments during the course of ACAP development.

2.0 Approach and Results

The design and development of the computer system to support CAT-ASVAB progressed along two obvious interrelated dimensions: computer hardware and software. The hardware needed for the CAT-ASVAB system had to be selected before the operating system and programming language could be identified.

Section 2.1 discusses the concept of operation; Section 2.2, the functional specifications for CAT-ASVAB; Section 2.3, the manner in which CAT-ASVAB must operate when deployed at the USMEPCOM testing sites; Section 2.4, the approach to the selection of the computer hardware; Section 2.5, the development of three software systems for the ACAP; Section 2.6, ACAP system documentation; and Section 2.7, system testing procedures.

2.1 Concept of Operation

To paraphrase the Stage 2 RFP, the CAT-ASVAB system must replace the operational P&P-ASVAB used for selection and classification of enlisted personnel. In 1985, P&P-ASVAB testing occurred at 68 Military Entrance Processing Stations (MEPSs), 2 substations (Anchorage and Guam), and approximately 900 field locations, identified as Mobile Examining Team (MET) sites. The MEPS and MET sites are under the control and administrative responsibility of USMEPCOM. The CAT-ASVAB system must be capable of being used at all of the testing sites as the P&P-ASVAB as well as provide automated, on-line test delivery and score reporting using adaptive, conventional, and timed psychometric tests. Item response theory (Lord, 1980) constitutes the theoretical foundation for CAT-ASVAB adaptive testing.

The CAT-ASVAB system must be capable of administering a battery of instruments equivalent to the present P&P-ASVAB. The P&P-ASVAB consists of eight power tests and two

speeded tests. After implementation, the CAT-ASVAB system must also be capable of administering other cognitive and noncognitive operational and experimental instruments, as determined by Department of Defense (DOD) policy.

Once fully deployed, the CAT-ASVAB system must be capable of accommodating a normal testing volume of approximately 1,000,000 examinations per year during normal working hours. In the event of mobilization, this requirement could be as high as 2,200,000 for a 6-month period. In general, 20% of the P&P-ASVAB testing occurs at MEPS; the remaining 80%, at MET test sites. Historical data on frequency and volume of testing, broken out by region, MEPS, and MET sites, are presented in *U.S. MEPCOM* (1983).

MEPS P&P-ASVAB testing is only part of the processing of applicants for enlistment at fixed-site locations in relatively controlled environments. Typically, military personnel conduct the testing. In contrast, Office of Personnel Management (OPM) employees working under a service agreement with DOD often conduct MET site testing at borrowed facilities. USMEPCOM does not have permanent control over the MET facilities or the authority to modify them. Despite these limitations, the Stage 2 RFP required the CAT-ASVAB system to be used by non-USMEPCOM examiners who must transport the CAT-ASVAB testing equipment to and from non-USMEPCOM facilities as required by testing schedules.

ACAP assumed the concept of operation for CAT-ASVAB, except that the scope of the ACAP, which is basically a research study, is more limited. Only six MEPS with their associated MET sites are involved in ACAP. In addition, CAT-ASVAB testing at any one site occurs only for a few months. However, even though limited in deployment, the ACAP was still required to have full CAT-ASVAB functional and operational capability, as specified by the Stage 2 RFP (Naval Regional Contracting Center, 1984).

2.2 Functional Specifications

To develop a system concept/design for the computer system supporting the equating and field-testing of CAT-ASVAB within the context of ACAP required specifications and standards for the performance of the system when implemented. In 1985, at the initiation of the ACAP, the performance standards were documented in the Stage 2 RFP. However, as the ACAP progressed, some of these specifications were modified to meet the scope of ACAP and to accommodate CAT in an operational environment.

2.3 Operational Scenario

This section presents an abbreviated concept of the manner in which the CAT-ASVAB system (as specified in the Stage 2 RFP) should function once it is implemented.

Specifically, a Local CAT-ASVAB Network (LCN) of interconnected computers administers CAT-ASVAB to applicants for enlisted military service at the MEPS or MET testing sites within USMEPCOM. In addition, a Data Handling Computer (DHC) at each MEPS handles communication of information between the LCN units and a CAT central research facility. The DHC also stores examinee testing and equipment utilization data for 6 months, as required.

2.3.1 MET Site Testing

The hardware configuration for the transportable computer systems at the testing sites is based on the concept of a "generic" LCN. A generic LCN consists of six examinee testing (ET) stations monitored by a single test administrator (TA) station and peripheral support equipment (e.g., mass storage devices and printers). A single TA station must monitor up to 24 ET stations (i.e., administer the CAT-ASVAB to 24 examinees simultaneously). The CAT-ASVAB portability requirements state that each generic LCN consist of up to eight components weighing a total of no more than 120 pounds and each component weighing no more than 23 pounds. Environmental requirements for operating temperature, humidity, and altitude are also specified. The TA and ET stations must be interchangeable so that each TA and ET station can serve as the backup for any other station in the LCN.

The LCN computer hardware specifications are as follows. Each ET station consists of a response device, a screen display, and access to sufficient random access memory (RAM) and/or data storage for administration of any CAT-ASVAB test; the amount of RAM required depends on the specific application software and networking design used. The ET stations are tied to a TA station by networking cables; Each TA station is essentially an ET station with a mass storage device and full-size keyboard. The failure of one station must not affect the performance of any other unit in the LCN. Each TA station has a very portable printer and modem. All components operate on ordinary 110 VAC line current. Battery packs are not used because they add weight and require additional logistic support.

The LCN operational requirements are as follows: Each LCN administers the CAT-ASVAB to military applicants scheduled for testing at the MET site. Initially, an OPM examiner would pick up the LCN equipment at a staging area (*U.S. MEPCOM*, 1983), transport it to the test site (sometimes a hotel room), carry it from the vehicle to the test site, and configure it for testing. When the system is ready for testing (i.e., "booting" and loading of source code/data files are completed), the examiner solicits personal data (name, Social Security Account Number [SSAN], etc.) from each examinee and enters this information into the system at the examiner's TA station. Then, the examiner instructs each examinee to sit at a specified ET station and start psychometric testing, without further examiner assistance. Examinee item response information is stored on a nonvolatile medium (e.g., microfloppy disk) to allow the test to continue at another ET station, in the event the original ET station fails during a testing session.

When an examinee is seated at each available ET station, the examiner enters the remaining personal data (e.g., USMEPCOM Form 714-A information) into an LCN Examinee Data File. When an examinee has completed testing, he or she is asked to review the personal data for accuracy. Incorrect information is flagged and corrected by the examiner before the examinee is excused. Finally, the personal and test item response data for each examinee are merged and saved in a nonvolatile storage medium, accessible to the TA station for later transmission to the MEPS DHC. The examiner is expected to monitor the various testing activities at the ET stations (e.g., CAT-ASVAB testing progress status and use of a "help" function).

After all examinees at a MET testing site have completed testing, the examiner sends the entire Examinee Data File (consisting of the personal data, item level responses, test scores, and composite scores) to the DHC unit at the associated MEPS using a modem and dial-up

telephone line, if available. If this was not possible (e.g., no telephone line at the test site), the data are transferred when the equipment is returned to the staging area.

Finally, the examiner packs up and returns all equipment to the staging area.

2.3.2 MEPS Site Testing

MEPS equipment is stationary, but otherwise is identical to MET site equipment. In contrast to most MET sites, each TA at a MEPS testing site can monitor 24 ET stations simultaneously. In addition, on start-up, the TA obtains the latest software and testing data from the DHC unit at the MEPS via either a hard-wired connection or a transportable medium. At the end of testing, testing data are sent to the DHC using the same medium. An LCN at the MEPS would not use dial-up telephone lines.

The MEPS site implementation of CAT-ASVAB also includes a DHC unit to collect data daily from each LCN in the associated MEPS administrative segment, including any LCNs at MET sites. These data are to be compiled and organized on the DHC for:

1. Daily transmission of an extract of examinee data collected that day to the USMEPCOM minicomputer located at the MEPS. These data are equivalent to the examinee data sent for P&P-ASVAB examinees.
2. Periodic transmission of all examinee data to the Navy Personnel Research and Development Center (NAVPERSRANDCEN) (perhaps via the USMEPCOM telecommunications network).
3. Archiving of all examinee and equipment utilization data at the MEPS for at least 6 months.

The MEPS DHC also must be capable of receiving new software, test item bank updates, and instructions from NAVPERSRANDCEN and telecommunicating this information to field LCN units.

2.4 ACAP Hardware Selection

Computer hardware selection initially involved selection of an LCN to administer the CAT-ASVAB. The DHC is considered to be a data management support device, not directly related to psychometric functionality.

2.4.1 Selection Factors

Psychometric acceptability, portability, reliability, environmental conditions, ease of use, and security were the most important factors used to select the computer system design for an LCN to support CAT-ASVAB testing.

Psychometric Acceptability. The computer hardware must meet the CAT-ASVAB psychometric requirement. Jones-James (1986) and the appendix detail the computerized testing algorithms and data needed to administer the CAT-ASVAB tests. The factors that most influenced the hardware selection process are: on-line data storage and display response time for test items.

The on-line data storage requirement dramatically influenced the selection of computer hardware to support the ACAP system as well as the design of the software for the ET station, TA station, and DHC units. Several types of data must be stored: (1) calibrated item banks (two forms or banks are used in ACAP), (2) uncalibrated or seeded item files, (3) supporting data files, (4) software (executable code segments), (5) operating system usage, and (6) growth and expansion storage. Depending on the hardware design and its capabilities, each of these data storage dimensions affects a candidate LCN's ability to perform CAT-ASVAB functions. Table 1 lists the data storage requirements of the ACAP data files for the ET station computer hardware (at the time of hardware selection, these requirements were estimated to be larger).

Table 1

**Data Storage Requirements for ET Station During
Score Equating Verification (SEV) Data Collection**

Description	Data Storage (Kilobytes)
Bootup software and data	380
Test administration software	270
Calibrated test items/form	365
Uncalibrated test items	25
Supporting data files	40
Operating system	300
Total	1380

Note. ET = examinee testing.

Computerized test administration demanded that the computer hardware (and software) generate and display the next test item quickly. The rate at which test items may be displayed to an examinee must be virtually constant so that the examinee cannot easily detect the display variability. The Stage 2 RFP required initiation of the display of the next test item within 2 seconds of the examinee's response to the current item and completing the display within 3 seconds under maximum load conditions (i.e., 24 examinees being tested in an LCN). These requirements would be a problem for a hardware system relying on telecommunication lines to a central computer.

Portability. The portability factor required that the OPM examiner be able to carry the computing equipment from one test site to another in various environmental conditions. Weight and size specifications limit any one portable component to 23 pounds. In addition, a generic LCN could consist of no more than eight components (six ET stations and a single TA station, and peripheral support equipment) with a combined weight of no more than 120 pounds.

Reliability. It must be virtually impossible to be unable to complete a CAT-ASVAB test once initiated, if electrical power is still available in the testing room. If any system component fails (ET station, TA station, the supporting electronic network, etc.), testing at (or near) the point of failure must resume within 5 minutes without loss of critical examinee testing data. In addition, failure of an ET or TA station must not affect the performance of any other station.

Environmental Conditions. The CAT-ASVAB computer hardware must use standard 115 volt/15 ampere AC power and operate in temperatures ranging from 60° to 100°F with relative humidity ranging from 50% $\pm$ 10% to 94% $\pm$ 4% (without condensation). Nonoperating temperature range is from 10° to 150°F.

Ease of Use. Examinees must be able to view the test items displayed on the screen of the ET station with ease. The computer display screen must be as independent of ambient light as possible given current technology. USMEPCOM testing sites have heterogeneous room lighting. The TAs must be able to assemble and prepare the equipment for operation from the transportable packages in 30 minutes.

Security. The Stage 2 RFP focused on the security of personal information (because of the Privacy Act of 1974) being telecommunicated among various components of the CAT-ASVAB system. However, these concerns are not relevant for ACAP, which does not use telecommunications for data transmission; all information exchange occurs using microfloppy disks. ACAP test item data must be secured at all times, preferably by some type of encoding mechanism. (The Stage 2 RFP did not require coding of test item data.) It was desirable for the test item data to be encoded onto the testing site storage media and for these media to be physically secured at all times.

2.4.2 Generic Designs

Tiggle and Rafacz (1985) discussed the three generic computer system designs being considered for use as the local computer network for the CAT-ASVAB program. The three designs differed in how they stored and provided access to test items during test administration. Storing test items on removable media (e.g., 3.5 inch microfloppy disks) or a central file server (e.g., a hard disk) had disadvantages with security, media updating, ease of use, maintenance, reliability, and response time.

The design selected emphasizes the use of RAM. Each TA and ET station requires at least 1.5 megabytes (MB) of internal RAM, which can accommodate all the software and data needed to administer the CAT-ASVAB tests (Table 1). In case of LCN failure, each ET station can operate independently of any other station in the network. The ET station needs one microfloppy disk drive and an electroluminescent or LCD technology display screen. In addition, the TA station can perform the functions of an "electronic" file server. The TA station could have a large amount of total RAM available, which provides great flexibility in the total number of alternate forms available during any one test session.

Several removable media or disks are required to boot the systems. However, regardless of the number of ET stations in an LCN, one microfloppy disk can store the test item banks for one of the two CAT-ASVAB forms. Supporting data files and software reside on a third disk. Each ET station also requires one microfloppy disk installed as a bootup and archive disk. Normally (after initial booting at the beginning of a test session), the TA would not move the microfloppy disks (i.e., the network would move all data).

This design offers many advantages including a large degree of flexibility with respect to design options. The ET stations can operate as stand-alone devices (i.e., without the use of the TA station). As such, it would be virtually impossible for an examinee's test session to fail to be completed; each ET station would be a backup station for every other station in the LCN. This

design is very reliable because it minimizes use of mechanical devices. Finally, it provides a very high level of security because volatile RAM is erased when the power to the computer is turned off.

LCN monitoring and the system response time requirements are not functionally related. The computer hardware can be configured so that the data storage requirements (Table 1) (for any one CAT-ASVAB form) reside at the ET station. Therefore, the response time display of test items can be independent of the LCN. The item display process takes place at RAM speed, resulting in a maximum response time on the order of 1 second, which is well within CAT-ASVAB specifications.

The major disadvantage of this design is that it cost more than some of the other alternatives.

2.4.3 Hardware Selected

Additional specifications (in addition to those in the Stage 2 RFP) were required to ensure the successful completion of the software to support the CAT-ASVAB psychometric requirements. These specifications stated that:

1. The computer hardware must support a UNIX¹ operating system. Operating systems for microcomputers (e.g., MS-DOS, Xenix) indicate a trend towards a multitasking environment, which was already available as part of UNIX. With UNIX, several development tasks (e.g., edit, compile, and execute programs) can occur simultaneously on the same computer.
2. The computer hardware must be based on a 8 MHz central processing unit (CPU) to meet the CAT-ASVAB system response times and the networking capability needed to support CAT-ASVAB. In 1985, the Motorola 68000, 8 MHz CPU was one of the better options.
3. Each LCN requires a printer to produce operator and recruiter reports for examinees processed. Ink-jet printers were needed to meet the noise level specifications at testing sites.
4. Each station requires an electroluminescent screen. The electroluminescent screen is virtually independent of ambient light and involves minimal weight, both essential qualities for testing at USMEPCOM testing sites.
5. A 3.5 inch microfloppy disk drive was specified to serve as a nonvolatile device to store the history file of examinee testing data for failure recovery purposes. The 3.5 inch disk is smaller, much less destructible, and more reliable than is the 5.25 inch disk.

The ACAP specified "C" as the programming language because it is native to the UNIX operating system and has the following characteristics that greatly aid software development, performance, and testing: (1) support of structured programming, (2) portability, (3) execution speed, (4) concise definitions and fast access to data structures, and (5) real-time system programming.

¹Identification of specific equipment and software is for documentation only and does not imply endorsement.

The hardware procurement was negotiated by the Navy Supply Center, San Diego using a brand name or equivalent procurement strategy. This resulted in the selection of Hewlett Packard Integral Personal Computer (HP-IPC). Each ET station consists of the following components in a single compact and transportable (25 pound) package:

1. One 8 MHz 68000 CPU with 1.5 MB of internal RAM with an internal data transfer rate (RAM to RAM) of 175 KB/second.
2. One read-only memory (ROM) chip with 256 KB of available memory containing a kernel of the UNIX operating system.
3. One microfloppy disk drive (710 KB capacity) with data transfer rate (disk to RAM) of 9.42 KB/second.
4. One adjustable electroluminescent display with a resolution of 512 (horizontal) by 255 (vertical) pixels (screen size 9 inches measured diagonally; 8 inches wide by 4 inches high).
5. One custom-built examinee input device (essentially a modification of the standard HP-IPC keyboard).
6. One Hewlett Packard Interface Loop (HP-IL) networking card.
7. One integrated ink-jet printer for use when the ET station must serve as a backup to the TA station.

Figure 1 shows an ET station ready for examinee testing.

Each TA station is configured identically to the ET station, but includes 2.5 MB of internal RAM and a full-size ASCII keyboard. Figure 2 shows a TA station, as configured from its transportable package.

In summary, each generic LCN (i.e., 6 ET stations tied to a single TA station) consists of seven transportable components weighing approximately 175 pounds. Figure 3 depicts an expanded LCN with up to 30 ET stations connected to a single TA station and the counterclockwise flow of data throughout the LCN. Using the HP-IL networking card and special network driver software achieves a network data transfer rate of approximately 9KB/ per second.

The data handing computer (DHC) system, also based on the HP-IPC, consists of the following components:

1. One ET station with a full-size keyboard.
2. Two 55 MB hard disk drives (primary and backup data archive units).
3. One cartridge tape drive unit; periodically, a cartridge tape of examinee testing data is to be sent to NAVPERSRANDCEN.
4. Telecommunications hardware to communicate with the MEPS minicomputer.

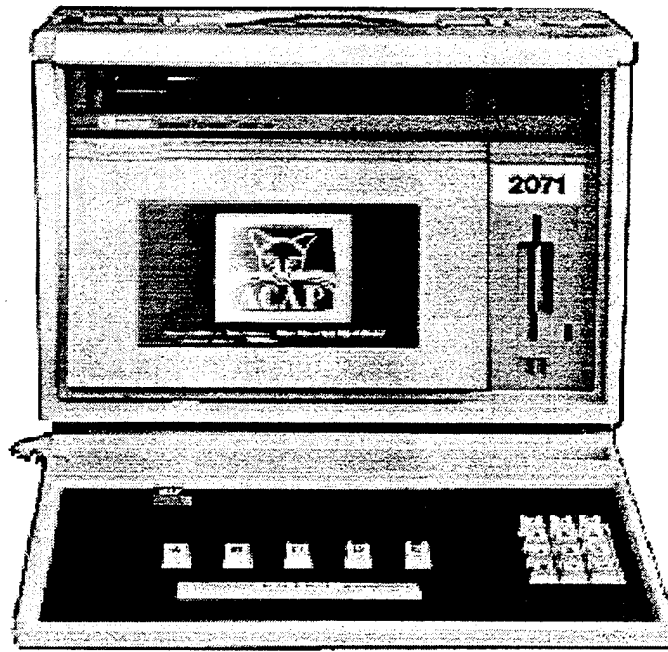


Figure 1. ET station ready for examinee testing.

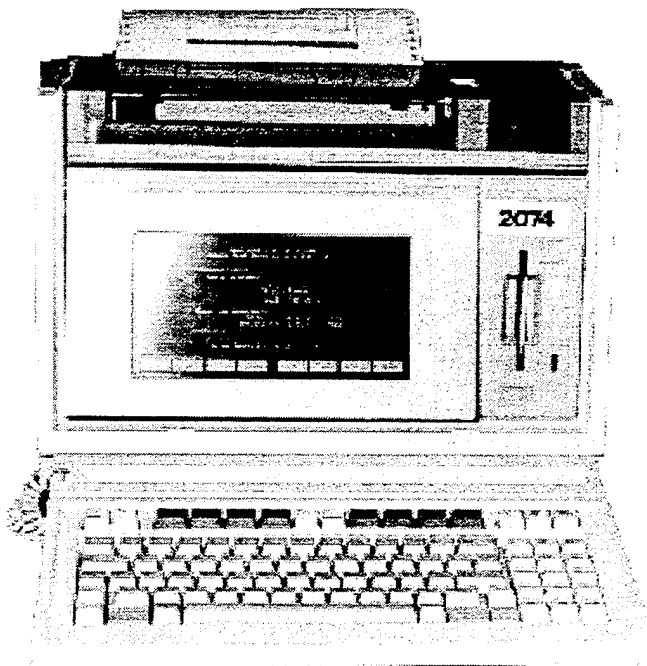
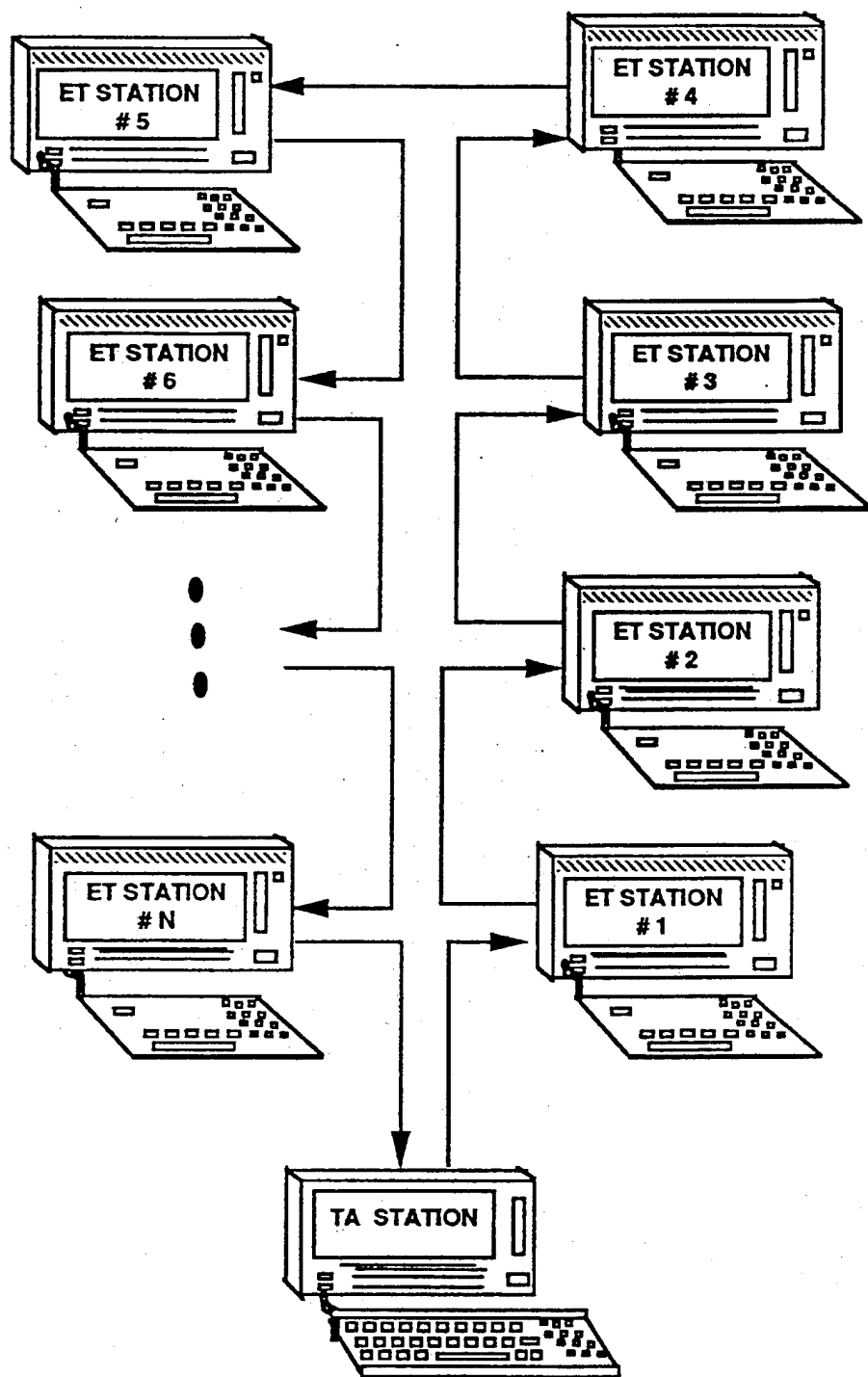


Figure 2. TA station ready for use.



(N <= 30)

Figure 3. Local CAT-ASVAB Network (LCN).

2.5 ACAP Software Development

The approach to software development progressed along two avenues: managerial and technical. Plans and progress of the CAT-ASVAB program have been reported annually (Sands, 1985; 1986; 1987; 1988; 1989; 1990).

Technically, the approach to the software development efforts proceeded along traditional lines; that is, a top-down structured design approach was used, consistent with current military standards for software development (e.g., DOD-STD-2167A). The functional requirements for each of the three software packages—TA station, ET station, and DHC—were identified and developed to assist in the development of a macro-level design for each package; that is, how the software is going to work from the standpoint of the user/operator. These requirements also served as the basis for developing detailed computer programming logic to support the main functions within the macro-level design. A thorough study of this logic permitted the identification of the primitive routines and procedures that were required (e.g., a routine was required to confirm the correct insertion of a disk into the disk drive, and to solicit and confirm the entry of ET station identification numbers). Then, using the primitive routines, main stream (logic) drivers were developed to link the primitives into a working system that mirrors the functional requirements of the macro-level design. The software was then tested, errors identified and corrected, and retested until all portions of the software worked together as required. Occasionally, the software design required modification as the impact of the interaction among various routines became more complicated and/or specifications were more clearly defined.

In addition to an overall plan for software development, the following guidelines for software development were devised to minimize disruption to existing USMEPCOM procedures:

1. Operational USMEPCOM procedures, the processing of examinees, and the movement of data throughout the ACAP system should follow the current P&P-ASVAB system as much as possible. Figure 4 portrays the ACAP data communications network, slightly modified to accommodate research data collection. Examinee testing data are collected at a MET site on a Data Disk,² which is sent by registered mail to the parent MEPS for further processing by the DHC operator; MEPS testing site TAs manually carry the Data Disk to the DHC operator.

At the DHC, an extract of each examinee's data record (as represented on a Data Disk) is compiled and sent to the USMEPCOM MEPS minicomputer (Sperry System 80). Each examinee extract consists of 153 bytes of a total of approximately 3500 bytes. The format of the extracted data is identical to that of the data sent to the MEPS computer by the P&P-ASVAB program. Finally, the DHC periodically compiles all of the testing data from the testing sites and communicates it by means of a cartridge tape to NAVPERSRANDCEN. This guideline avoids the movement of data via telecommunications (as specified in the CAT Stage 2 RFP).

2. The system should be designed for ease of use by the TA rather than the ease of programming a particular function. For example, in designing the interactive screen dialogs, serious concern was given to including such features as: (a) menu driven logic paths of minimal length, (b) invalid key lockout or audible/visual feedback when using invalid keys, (c) error-

²This report does not use ACAP procedure, which follows the manufacturer's practice of using the name "disc" when referring to a "disk."

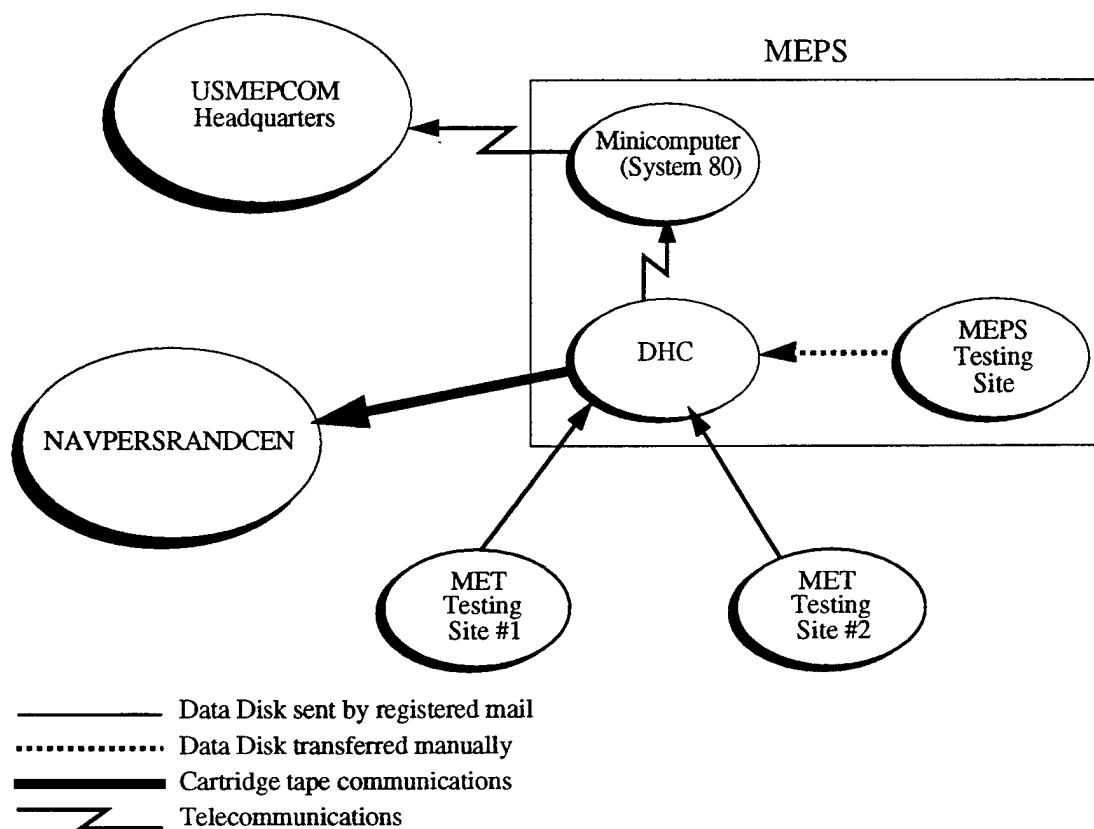


Figure 4. ACAP data communications network.

trapping when incorrect logic paths are requested, (d) the use of keywords to describe available functions, (e) explicit header messages on all screens (so the operator can always find the location within the overall logic), (f) confirmation of data before updating on nonvolatile media, (j) minimal use of removable media, and (h) the ability to generate screen or hard-copy reports of certain testing session information. In addition, the software was designed to accommodate the use of only a minimum number of keys on the keyboards of the ET and TA stations. Finally, the software was designed to include procedures that would permit the operator to easily recover in the event of a hardware failure. Additional information on these and related features of the software is presented later in this report.

3. System security procedures should be an integral part of system design. Once the use of telecommunications was eliminated, the security issue was simplified to addressing how to protect the test item files. This was approached by encoding the test item files onto the removable media (microfloppy disks) and by minimizing the media needed to support a testing session. The following five types of disks are needed to conduct an ACAP testing session:

- **ACAP system disks**, identified as A, B, C, or D, contain encoded test item files for the two CAT-ASVAB forms used in ACAP, supplementary data, and certain test administration software.

- **TA disk** is used by the TA station and maintains history files of information on testing sessions in progress and completed.
- **ET disk** is used by an ET station to boot that station and to maintain a history file of examinee testing activity for recovery in case of failure.
- **Data Disk** records CAT-ASVAB testing performance for all examinees tested during a testing session.
- **RASP disk** is used to allocate examinees to certain testing rooms for a research data collection effort during the initial phases of ACAP.

In actual use, ACAP system disks A, B, C, and D are used only when test item files and test administration software is read into RAM of the TA station. These disks would be secured while the TA station is broadcasting the encoded item files to the ET stations in its network. At the ET station, the item files are decoded in preparation for display during test administration. (Since these data are stored in [volatile] RAM, they are erased when electrical power is turned off at the conclusion of the testing session.)

4. The use of mechanical devices should be avoided to increase the reliability of the system. This guideline has reduced use of the printer and the 3.5 inch microfloppy disk drive. For the microfloppy disk drive, the software development effort was primarily concerned with the number of disk accesses, particularly to support failure recovery operations. For the ET station, failure recovery was considered to be recovery to the first uncompleted test, therefore necessitating disk access only at the end of each of the 11 CAT-ASVAB tests,³ instead of at the end of each item administered in each test as originally considered.

5. An LCN should have two modes of operation: networking and stand-alone. The networking mode requires full CAT-ASVAB functional capability and an active direct communications link between the TA station and the ET stations in the network. For this reason, the networking mode would be the predominant mode of operation. The stand-alone mode would be used primarily for failure recovery operations and would essentially consist of the ET stations operating without benefit of electronic communications with the TA station. In either mode, the full complement of functions supporting the CAT technology must be available, in accordance with the specifications for CAT-ASVAB testing.

The following sections describe the macro-level design of the TA station, the ET station, and the DHC.

2.5.1 TA Station Software

To design the software for the TA station, the functions to be supported by the TA station were compiled. The duties of the TA during an ACAP CAT-ASVAB testing session were identified (Rafacz, 1986). Concurrently, a test administration user's manual (Rafacz, 1990) was developed. The user's manual describes CAT-ASVAB testing at a testing site and how the TA uses the ACAP equipment.

³The P&P-ASVAB consists of 10 tests (see the appendix).

The TA station software must support the following functions:

1. The TA station prepares and communicates all software and data necessary for CAT-ASVAB test administration to ET stations in the LCN. Appropriate verification, error checking, and backup procedures must be in place in the event a transmission of software/data should fail.

2. The TA must be able to identify examinees by means of a unique identifier (e.g., SSAN) and to record (in a retrievable file) other examinee personal data. These personal data include: name (last, first, and middle initial), test type (initial or retest), service and component processing for, gender, education level and degree code, and race/population group. In addition, it should be easy for the TA to add or modify any of the personal data.

3. The software for the TA station must randomly assign (transparent to the TA) an examinee taking CAT-ASVAB to one of the two CAT-ASVAB forms used. This assignment is subject to the condition that examinees who have previously been administered a CAT-ASVAB form must be retested on the alternate CAT-ASVAB form. A printed report of the examinee/ET station assignments should be available to the TA upon demand. In addition, the software must maintain an accounting of examinee assignments and be prepared to develop new assignments in the event of failure of any station in the LCN.

4. During examinee testing (in the networking mode of operation), the TA must be able to receive a status report on the progress of examinees upon demand. This report should include: examinee's SSAN, ET station assignment, accumulated time spent on the CAT-ASVAB test being administered (also, the accumulated administration time on that test), and the expected completion time for the entire battery of tests.

5. The TA must be able to move the completed testing data recorded from an ET station to the TA station for additional processing. At this point, the TA station generates number-correct scores (from score equating tables) for the examinee as a function of the scores obtained on each of the CAT-ASVAB tests (see the appendix). The TA must also be able to easily generate a report for the examinee's military recruiter; the Score Equating Verification (SEV) version of this report includes an Armed Forces Qualification Test (AFQT) percentile score and personal identifying information. A backup mechanism is essential in the event the networking procedures for an LCN fail.

6. The TA must be able to store the testing data for all examinees who have gone through the TA station collection process in a nonvolatile medium (i.e., a Data Disk) for later communication to the parent MEPS. The TA retains a duplicate Data Disk at the testing site in case the disk sent to the MEPS is lost in transit.

7. Finally, it must be almost impossible for an examinee's testing session not to be completed with complete security for examinee testing information. If an examinee's assigned ET station should fail, that examinee must be reassigned to another available station and continue testing at the beginning of the first uncompleted CAT-ASVAB test. Likewise, if the TA station should fail, the LCN should fail, or electrical power is interrupted, the TA must be able to recover and continue the testing session promptly.

2.5.1.1 TA Station Operation. Figure 5 presents a flowchart of the TA station functions, which are discussed below.

1. Turning on the power to TA station starts bootup operations, which include performing the following tasks: (a) verifying that the TA station computer clock is set to the correct date and time, (b) soliciting the name and SSAN of the TA, (c) requesting the TA station computer identification number, (d) asking for the testing site identification number, and (e) performing certain file maintenance activities on the TA disk such as deletion of unneeded history files from previous testing session. The bootup operations terminate with the reading and execution of the main driver TA station software from ACAP system disk D.

2. Normally, the TA selects the networking mode of operation for the current testing session; the stand-alone mode, which is a failure recovery procedure, will be discussed in Section 2.5.1.2. Selection of the networking mode offers the TA the option of performing several network diagnostic tests: (a) correct loading test of the network driver software, (b) network loop integrity test, and (c) a communications test. Normally the tests are performed and, if errors are found, the software suggests one or more remedial actions.

3. If the TA decides to transmit testing data to the ET stations in the LCN, then instructions are provided for loading the data from the three system disks A, B, and C. These disks contain test administration software, item level data files (encoded), and supporting data (seeded test items, information tables, and item exposure control values, which are explained in the appendix). After these data and software are loaded into RAM of the TA station, the system disks are secured. The ET station randomly identifies a CAT-ASVAB test form with each ET station so that approximately 50% of the ET stations receive one of the two CAT-ASVAB forms. The TA station then proceeds to broadcast the test administration software and data files (one at a time, alternately) to the ET stations requiring a given form, then to the remaining stations. Therefore, while one set of stations (identified with one of the two forms) is receiving one file of test items, the remaining stations are storing the test items received into RAM.

If a data transmission error occurs during this process, the TA can take remedial actions (as prompted by the TA station) or resort to the failure recovery procedure (i.e., shift to the stand-alone mode of operation).

The operator would elect not to transmit testing data to the ET stations in the network in certain failure recovery situations. Specifically, if the TA station were to fail after all testing data files were communicated to the ET stations in the network, but before examinee testing had begun, it is possible for the TA station to recover in the networking mode. In this case, it is not necessary that the testing data files be recommunicated to the ET stations.

4. Before the ET stations are ready to administer the CAT-ASVAB, some additional processing is necessary. Figure 6 displays the menu of options available to the TA. Use of the Recruit Allocation System Program (RASP) results in a procedure for identifying examinees to be tested using CAT-ASVAB for the current testing session; see Huynh (1990). The RASP is a temporary interface to the TA station computer that permits the allocation of all examinees either to a testing room administering the CAT-ASVAB or to a testing room administering a P&P-ASVAB test. Another computer would be executing the RASP software whereby the operator would identify

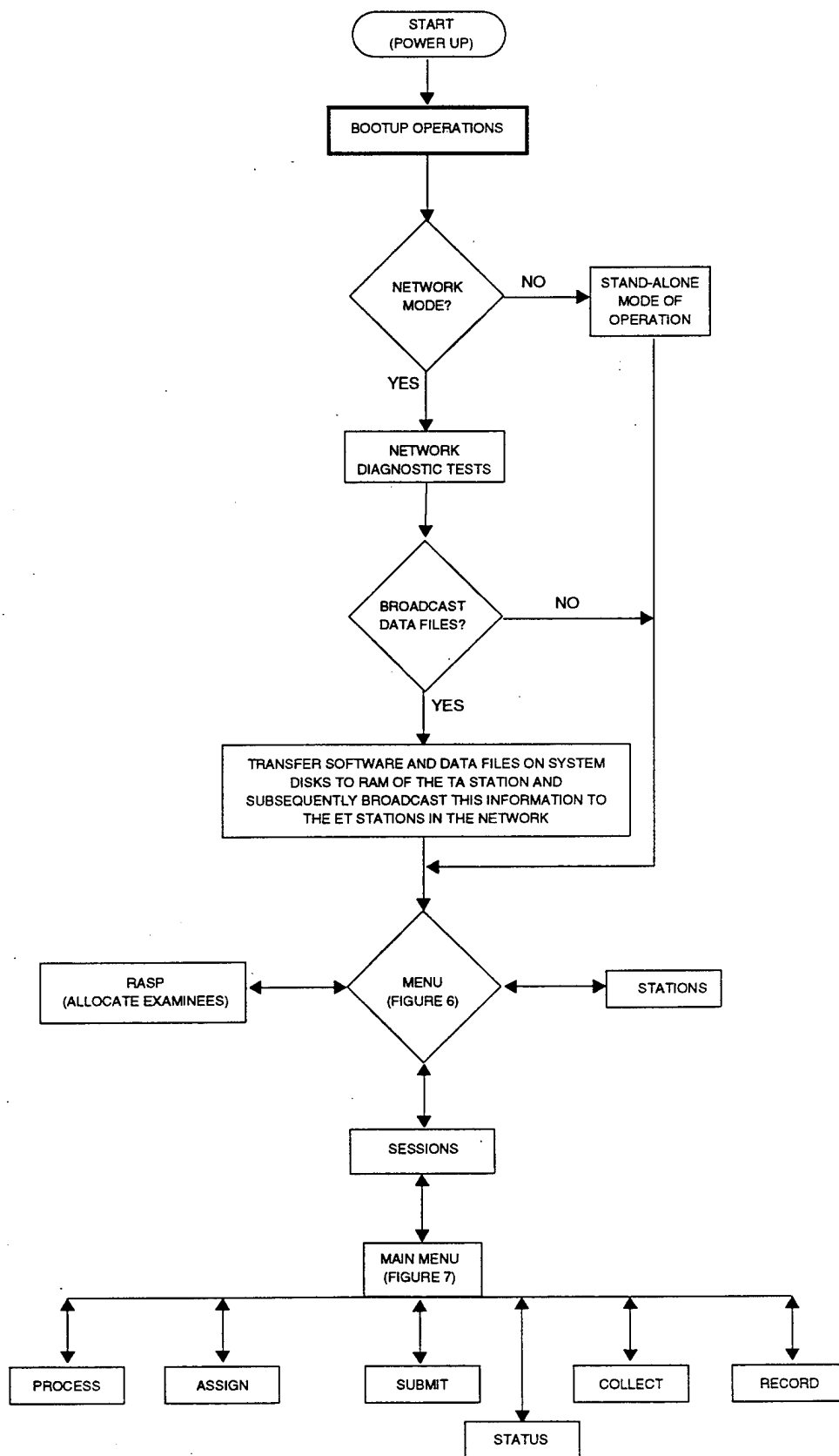


Figure 5. Flowchart of the TA station functions.

ACAP COMPUTERIZED TESTING	
Select a Test Administration Function	
f1: Process Examinee Testing	STATIONS
f2: Process ACAP Computerized Testing	SESSIONS
f3: Transfer Examinees Allocated on	RASP disk to TA disk

STATIONS	SESSIONS	RASP				HELP
-----------------	-----------------	-------------	--	--	--	-------------

Figure 6. Menu to process stations, sessions, or RASP.

all examinees arriving for testing in terms of their name, SSAN, and test type. Subsequently, these examinees are allocated on that computer and various reports produced. The examinees allocated to be CAT-ASVAB tested are also recorded on the microfloppy disk associated with the RASP computer; called the RASP disk. Choosing the RASP option in Figure 6 results in a request to insert the RASP disk into the TA station disk drive whereby those examinees are read by the TA station software and subsequently recorded onto the TA station disk.

5. When RASP processing is completed, the menu in Figure 6 reappears. In the stand-alone mode of operation or in certain failure recovery operations, the STATIONS option is selected (discussed in Section 2.5.1.2). Use of this option allows the TA the option of creating a new list of stations, adding (or deleting) stations for testing, and receiving a screen and/or printed list of stations.

6. The TA selects the SESSIONS option (from the menu in Figure 6) to work with the upcoming testing session. After the TA enters the date and approximate starting time for the session, the Main Menu (Figure 7) is displayed. The Main Menu displays the primary functions performed during a testing session. PROCESS is a means for the TA to identify examinees to be tested. When CAT-ASVAB is deployed, the PROCESS option will be used exclusively and the use of RASP will be phased out. The PROCESS option includes procedures for editing an examinee's name, SSAN, and test type information. Other options available under PROCESS include: creating a new list of examinees for testing, adding (or deleting) an examinee for testing, and providing a screen and/or printed list of examinees for testing.

7. The ASSIGN option (Figure 7) assigns a set examinees scheduled for CAT-ASVAB testing by randomly assigning each examinee to one of the two CAT-ASVAB test item bank forms and to one of the ET stations in the LCN. Initially, the screen of the TA station displays each examinee's name, SSAN, test type, and the number of the assigned ET station. The examinee assignments are recorded on the TA disk at the TA station, printed at the TA station (see Figure 8), and then broadcast to the ET stations in the LCN. When an ET station receives the list, it searches the list for its number. If found, it displays the assigned examinee's SSAN and waits for additional keyboard input. Stations that do not find their number assume that they have not been assigned an examinee for testing and wait to serve as a failure recovery station.

PROCESS ACAP COMPUTERIZED TESTING SESSIONS		
Select a Test Administration Function for		Date: 02/12/90 Time: 14:00
		Current Time 14:21:59
f1:	PROCESS	Examinees for this Session.
f2:	ASSIGN	Examinees to ET Stations for this Session.
f3:	SUBMIT	Examinee Personal Data for this Session.
f4:	COLLECT	Examinee Testing Data for Session.
f5:	RECORD	All Examinee Data for this Session.
PROCESS ASSIGN SUBMIT COLLECT RECORD STATUS HELP MENU		

Figure 7. Main menu for administering a testing session.

ASSIGNMENT REPORT				
Site Location = <u>111222</u>		Session Date = <u>02/12/90</u> Start time = <u>1400</u>		
TA Station Number = <u>2111</u>		Test Administrator: <u>Smith Harry O.</u>		
Name	SSAN	Test Type	Original Form	Station Assignment
Curie, Marie	333-33-3333	Initial		1526
Gable, Clark G	444-44-4444	Retest	02C	2007
Martinez, Pepito J	222-22-2222	Initial		2053
Selleck, Thomas H	666-66-6666	Initial		2089
Monroe, Marilyn B	777-77-7777	Retest	01C	2120
House, Nora D	555-55-5555	Initial		2190
Bonita, Juanita L	111-11-1111	Initial		2335
Valentino, Ramon M	888-88-8888	Retest	15A	2360
			(Unassigned)	2032
			(Unassigned)	5820
8 Examinees have been assigned				

Figure 8. Sample assignment report.

8. At this point, the examinees to be tested are directed either to a P&P-ASVAB testing room or to the CAT-ASVAB testing room, according to the decision of the RASP. CAT-ASVAB examinees are directed to sit at the seats corresponding to their assigned ET station (see Figure 8). They receive general instructions that start CAT-ASVAB test administration. At the TA station, the Main Menu (Figure 7) is displayed and the TA can use the STATUS option for a screen report on the progress of examinees during testing. This report includes the examinee's name, SSAN, total time accumulated since the CAT-ASVAB test began, the test being administered, the accumulated time on that test, and the expected completion time for the entire battery of CAT-ASVAB tests. The examinee's recruiter uses the expected completion time to assist in scheduling.

During testing, the TA submits various personal information on the examinees being tested. This information includes service and component for which examinee is being processed; gender; education level and degree code; and race/population group. Choosing the SUBMIT option in the Main Menu enables the TA to enter into a menu-driven dialogue with the TA station that enters this information from the examinee's USMEPCOM Form 714-A.

9. At the conclusion of an examinee's CAT-ASVAB test, the STATUS option (Figure 7) displays the word DONE in the status screen report for the examinee. The TA uses the COLLECT option (Figure 7) to retrieve the examinee's testing data from the examinee's assigned ET station. The subsequent COLLECT screen (Figure 9) lists the examinees being tested and their testing progress in terms of having been processed through the SUBMIT (Sub), COLLECT (Col), and RECORD (Rec) options of the Main Menu (Figure 7). From Figure 9, once the TA selects an examinee (by the use of cursor control keys), the TA station automatically collects the testing data from the assigned ET station and updates the screen display.

PROCESS ACAP COMPUTERIZED TESTING SESSIONS																														
Collect Examinee Testing Data for		Date: 02 / 12 / 90 Time: 14: 00																												
MODE: NETWORK / LCN COLLECT																														
Use the Up/Down Arrow Keys to Select an Examinee to COLLECT.																														
Name: Martinez, Pepito J. SSAN: 222-22-2222																														
Press RETURN to Accept.																														
		<table border="1"> <thead> <tr> <th>Last Name</th> <th>Station</th> <th>Status</th> </tr> </thead> <tbody> <tr> <td>Bonita</td> <td>2335</td> <td>Sub ...</td> </tr> <tr> <td>Curie</td> <td>1526</td> <td>Sub Col Rec</td> </tr> <tr> <td>Gable</td> <td>2007</td> <td>Sub Col Rec</td> </tr> <tr> <td>House</td> <td>2190</td> <td>Sub ...</td> </tr> <tr> <td>Martinez</td> <td>2053</td> <td>Sub Col ...</td> </tr> <tr> <td>Monroe</td> <td>2120</td> <td>Sub Col ...</td> </tr> <tr> <td>Selleck</td> <td>2089</td> <td>... ...</td> </tr> <tr> <td>Valentino</td> <td>2360</td> <td>Sub ...</td> </tr> </tbody> </table>		Last Name	Station	Status	Bonita	2335	Sub ...	Curie	1526	Sub Col Rec	Gable	2007	Sub Col Rec	House	2190	Sub ...	Martinez	2053	Sub Col ...	Monroe	2120	Sub Col ...	Selleck	2089		Valentino	2360	Sub ...
Last Name	Station	Status																												
Bonita	2335	Sub ...																												
Curie	1526	Sub Col Rec																												
Gable	2007	Sub Col Rec																												
House	2190	Sub ...																												
Martinez	2053	Sub Col ...																												
Monroe	2120	Sub Col ...																												
Selleck	2089																													
Valentino	2360	Sub ...																												
AUTO	REPORT	DISK COL	MENU																											

Figure 9. Sample screen for collecting examinee testing data.

The AUTO option (Figure 9) requires the TA station to query the ET stations constantly to determine whether an examinee has completed testing. Testing data for examinees who have completed testing are automatically collected and the screen display (Figure 9) updated.

If in the course of collecting examinee testing data, a network error is detected, the TA selects the DISK COL option to transfer the testing data manually to the TA station; that is, he or she retrieves the ET disk from the ET station and inserts it into the TA station disk drive when requested. The TA station then reads the examinee's testing data directly from the ET disk and collects the data. Similar mechanisms are available throughout the TA station software to assist the TA in failure recovery situations.

At the end of the data collection process for an examinee, the TA station software produces 10 (equated) number-right scores (described in the appendix). The number-right scores are interchangeable with the P&P-ASVAB scores and can be used for selection and classification purposes. The TA station software uses these equated scores to produce an AFQT percentile score (Segall, 1989). This AFQT score with personal identifying information is automatically reported on an Unverified ACAP Test Score Report and given to the examinee's military recruiter. Use of the REPORT option (Figure 9) allows the operator to reproduce this report as often as necessary.

10. Finally, at any point during the testing session, the TA can record the examinee testing data on a set of microfloppy disks for transfer to the MEPS. Selecting the RECORD option (Figure 7) instructs the TA to first insert a MASTER Data Disk and then a BACKUP Data Disk. The MASTER Data Disk is sent to the parent MEPS for processing, while the BACKUP Data Disk remains secured at the testing site and is sent to the MEPS if needed.

2.5.1.2 Failure Recovery Operations. Special software routines that support common failure recovery situations are discussed below.

1. ET station failure. As soon as the TA realizes that an ET station has failed, he or she removes the failed station from the LCN and from the list of available ET stations (select STATIONS in Figure 6 and then the DELETE option in the subsequent menu). The ASSIGN option (Figure 7) generates a new ET station assignment for the examinee. If a station is not available, the examinee must wait until another examinee completes testing and the testing data are collected. In all cases, the TA station software attempts to assign an examinee to an available ET station with the test item files of the examinee's original station already loaded at the new station. If this is not possible, the ET station software will detect this inconsistency and require reloading of the correct form of item level data from the ACAP system disks. This reloading of test items is a rare occurrence.

2. TA station failure. If the TA station fails, recovery is possible by using any available ET station. However, as the network synchronization is interrupted with the failure of the TA station, the TA would normally use the stand-alone mode of operation to continue the testing session. Fortunately, the only significant loss of capability is that the TA must collect examinee testing data by moving the ET disk from each ET station with an assigned examinee over to the TA station at the conclusion of testing (DISK COL option in Figure 9). In all other respects, the testing session may continue normally.

If the TA disk fails during a testing session, special routines in the TA station enable the TA to use a BACKUP TA disk. Once the TA station is powered up using a BACKUP TA disk, the TA is required to insert the original (now defective) TA disk into the disk drive of the TA station. If possible, a history file (containing information related to testing sessions) on the original disk is read and moved over to the BACKUP TA disk. This approach makes it possible to synchronize the

TA station software to the point of failure of the TA disk and continue the testing session without compromising the test administration process and any of the examinee testing data being recorded.

3. LCN failure. If the LCN fails while the TA is using it, an error message appears at the TA station. Although on-line diagnosis of LCN errors is very difficult, the TA station software can tell when the error occurred and provide suggestions for isolation of the fault and for remedial action. If the TA cannot isolate and correct the fault, he or she must change the testing session to the stand-alone mode of operation.

4. Catastrophic failure. A total electrical power failure or a similar catastrophe requires a combination of recovery procedures, since essentially all stations have failed. Briefly, the TA should shift to the stand-alone mode of operation and recover each examinee's testing data at the examinee's original station; reassignment is needed only if some of the ET stations were damaged as a result of the abrupt removal of power. This process takes about one-half hour for an LCN with about 10 ET stations.

2.5.2 ET Station Software

The design of the software for the ET station was based on the psychometric requirements for CAT supplemented by specifications associated with the computer administration of any test, improved psychometric procedures, and requirements unique for military testing. CAT-ASVAB and ACAP specifications are discussed in the appendix.

2.5.2.1 Software Design Specifications. The choice of a large RAM-based computer for the LCN computer hardware permits the ET station test administration process to be conducted independently of computing activities at the TA station. An ET station is only required to communicate with the TA station in the following (noncompromising) situations:

1. Prior to testing, when the TA station is broadcasting software and data to the ET stations in the network.
2. Prior to testing, when the TA station is broadcasting examinee assignments to the ET stations.
3. During a testing session, but at the end of administration of each item (and before the next item is displayed), for the purpose of providing status information to the TA station (Option 6 in Figure 7).
4. At the end of an examinee's battery of CAT-ASVAB tests, to communicate testing results to the TA station for subsequent compilation onto a Data Disk.

A network error during a testing session that occurs with situation 3 (above) will not result in an interrupt or error condition at the ET station. The other situations interrupt processing at the ET station, but do not impact examinee test administration. At all other times, the ET station is "on its own" and completely controls the test administration process.

In addition to the purely psychometric functions supporting the use of the CAT technology, the software design considers the functions supporting computer operations at the ET station. During examinee test administration, two operations are of concern: failure recovery at the ET station (discussed in Section 2.5.1.2) and examinee implicit and explicit requests for help (discussed in Section 2.5.2.2). The ET station software design with respect to all functions supported is discussed below:

1. Placing an ET disk in the disk drive of the ET station, initiates the following bootup operations: hardware verification procedures (screen, disk drive, and keyboard), soliciting the mode of operation for the computer (networking or stand-alone), requesting the ET station computer identification number, and verifying that the ET station computer clock has been set to the correct date and time. Finally, the bootup process is completed with the ET station performing file maintenance activities on the ET disk, such as deleting examinee testing data files from previous testing sessions that are no longer needed. These processes must be completed at each ET station prior to the start of the testing session.

Normally the networking mode of operation is selected. If the stand-alone mode is selected, broadcasting of software and data files (paragraph 2) is not required. In that case, the ET station reads the necessary testing data and software directly from the ACAP system disks. In addition, ET station assignments, dictated by the TA station, and test type (initial or retest) are entered manually at each ET station. Finally, examinee testing information recorded on the ET disk is collected manually by moving the ET disk to the TA station at the conclusion of examinee testing.

2. Now, the ET station is ready to receive test item data files and software from the TA station. The first file received describes all the files the ET station is to receive over the network, including which of the two forms of test items the ET station is to receive. The first file is the actual test administration software which, once received, is executed, terminates the bootup program, and then monitors receipt of the following data files (from the TA station) to support examinee test administration: power and speeded test item text, graphic, and item parameter files; information table files; and exposure control parameters for power test items. Each power test item file is stored in the ET station RAM, which is designed to support subsequent random retrieval (according to the information table associated with each power test; see the appendix). As the ET stations assigned one of the two power test forms are storing the item file in RAM, the remaining ET stations are receiving the alternate form item file. This procedure reduces overall LCN start-up time.

3. After an ET station has received all of the required data files, it is ready to receive the examinee assignment list from the TA station. Once this list is received, the ET station searches for its identification number. If its identification number is not on the list, no examinee has been assigned to this station for testing, and it prepares to support Failure Recovery operations or the testing of an examinee who arrived after the testing session has begun (i.e., a NEW examinee). If its number is on the list, the ET station prepares for test administration of the examinee with the SSAN associated with its number. This requires confirming that the correct form of test items has been loaded for the assigned examinee. If not, the ACAP system disks are requested and the correct testing data files are loaded into RAM; this rarely happens.

4. Now that the ET station is ready to administer the CAT-ASVAB test to the assigned examinee, the TA must give the examinees verbal instructions and direct each examinee to the

assigned ET station. The TA verifies the displayed SSAN with the examinee and modifies it, if required. The examinee presses the ENTER key on the keyboard of the ET station when requested to begin CAT-ASVAB test administration in accordance with the interactive dialogues specified by Rafacz and Moreno (1987). The dialogue for the remainder of examinee test administration is between the ET station (software) and the examinee; neither the TA nor the TA station is involved.

At the beginning of the dialogue with the examinee, the ET station prepares an examinee data file in the ET station RAM and on the ET disk for later retrieval of information collected during the test administration process. This information includes a comprehensive accounting of the performance of the examinee being administered the CAT-ASVAB tests. A detailed description of this information is presented in Wilson, Bebb, Higgins, Boston, and Agree (1989).

5. Initially, the computer screen presents the examinee with information on how to use the ET station keyboard, also called the Examinee Input Device (EID); see Figure 10. The examinee learns how to use all of the keys labeled ENTER, A, B, C, D, E, and HELP; the examinee does not use the keys on the right side of the EID.

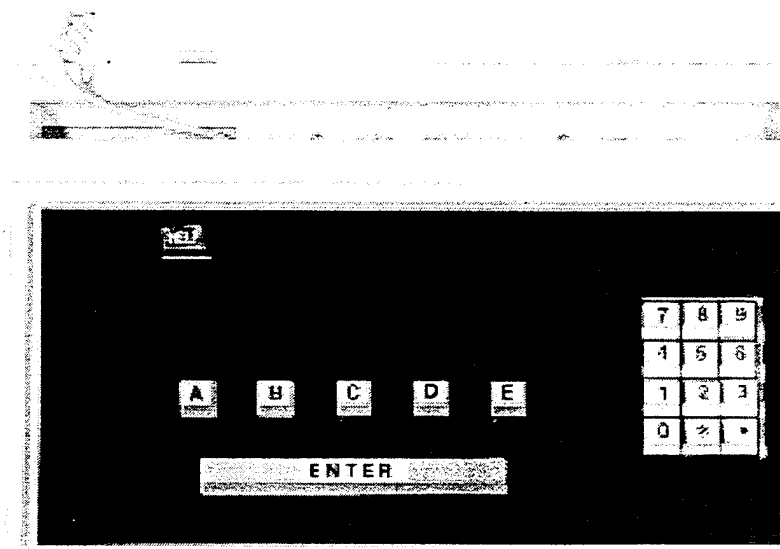


Figure 10. Examinee Input Device (EID).

The keys labeled 0, 1, . . . 9, “.” and “*” are used only by the TA; if the examinee attempts to use these keys an error message will result. The TA uses the numeric keys to enter the examinee’s SSAN at the ET station. The asterisk key (*) is used to remove the last character entered and is equivalent to the backspace key found on a typewriter.

6. Next, the examinee is trained on how to answer the power test items. Training on how to respond to the speeded test items is given just before these tests are administered. The examinee can request to repeat the training on how to use the keyboard and answer test items. On the second

request, the ET station halts the interactive dialogue with the examinee so that the TA can be called to enter a pass code for the interactive dialogue to continue. The ET station software describes the current situation, and then requests that the TA monitor the examinee's progress briefly before continuing with normal duties.

7. At this point, four power tests (General Science [GS], Arithmetic Reasoning [AR], Word Knowledge [WK], and Paragraph Comprehension [PC]) are administered. For each test, the examinee is initially presented with a practice item. The examinee is given an indication that the answer is correct or incorrect and the opportunity to request repeating the practice item. The second request initiates a call to the TA, who must enter a pass code to repeat the practice item. Finally, the examinee is ready to be administered the actual test items.

Figure 11 shows a typical GS test item displayed in the upper left section of the screen, with the item stem first followed by alternatives (A, B, C, D, and E). The examinee answers the test item by pressing the key corresponding to the alternative selected and then confirms the answer by pressing the ENTER key. Any other answer can be selected before ENTER is pressed. Selection of a valid response alternative highlights that alternative on the screen until another alternative is selected. Pressing an invalid key results in an error message being briefly displayed. As each item is displayed on the computer screen, the lower right corner of the screen presents the number of the item being administered, relative to the total number of items and the number of minutes remaining in the test.

A rose is

- A. an animal.
- B. a flower.
- C. a bird.
- D. a fish.
- E. an insect.

TYPE YOUR ANSWER (A, B, C, D, or E)

Question 10 of 16

5 Minutes TO GO

Figure 11. Typical CAT-ASVAB test question.

While the examinee studies the test item, his or her performance is recorded by the software monitoring the EID for a keypress. Overall, if the examinee does not confirm a valid response within the maximum item time limit (see Table A-1), the test is halted and a TA implicit help call is initiated (Section 2.5.2.2). In addition, if the examinee fails to complete the specified number of test items in the allotted maximum time limit for the *entire test* (see Table A-1), the test is automatically terminated (without a TA call) and the examinee continues with the next CAT-ASVAB test. If the examinee presses an invalid key (i.e., "*" key, the "." key, or any of the 0, 1, . . . 9 keys), an error message is briefly displayed. Three invalid keypresses result in an implicit

help call. Pressing the HELP key initiates the explicit help call sequence (Section 2.5.2.2). For a speeded test, a valid key response (A, B, C, D, or E) at this point results in the immediate display of the next test item. For power tests, a valid key response (A, B, C, D, or E) must be followed by the confirmation key (ENTER) in order to generate the display of the next item.

8. The test continues until the number of items administered (including one seeded item for a power test) equals the required test length or the maximum test time limit has been reached. As soon as the examinee completes the test, certain examinee test administration information is recorded in the ET station RAM and on the ET disk. For each item administered, this information includes: the item identification code, the examinee-selected response alternative, the time required to select the response (but not confirm), the new estimate of ability based on the selected response, and any implicit or explicit help call. In addition, the Bayesian modal estimate for the test (see the appendix) is recorded, as well as information pertaining to the performance of the examinee in completing the practice screens for the test. This information is also recorded on the ET disk (a nonvolatile medium) as a backup if the ET station should fail during testing.

9. The Numerical Operation (NO) and Coding Speed (CS) speeded tests are administered after the first four power tests. As with the power tests, practice test items are administered first. The examinee can repeat the practice items up to three times before a TA call is initiated. Examinee test administration of the speeded tests and the power tests differs. The speeded test items are administered in the sequence in which they appear in the item file, without using any adaptive testing strategy (see the appendix). In addition, the examinee does not confirm an answer by pressing the ENTER key. Rather, the ET station selects the first valid keypress (A, B, C, D, or E) as the examinee's answer. The display format of the CS test items is also different in that seven items are displayed on the same computer screen; NO and the power tests display only one item per screen. Rate-scores (see the appendix) are recorded as the examinee's final power test score. In all other respects, speeded test administration (including the availability of implicit and explicit help calls, the recording of examinee performance information) is identical to that of the power tests.

10. Once the speeded tests are completed, the examinee is administered the remaining five power tests (Auto Information [AI], Shop Information [SI], Mathematics Knowledge [MK], Mechanical Comprehension [MC], and Electronics Information [EI]). The procedure for administering these tests is identical to that for the original four power tests. Once the EI test is completed, the examinee's testing performance is concatenated in the ET station RAM and on the ET disk into a single file identified by examinee SSAN. This SSAN file is collected by the TA station for subsequent compilation onto a Data Disk. The examinee is instructed by the ET station to return to the TA station for further instructions and, then, is excused. The ET station is now available for testing some other examinee whose assigned station has failed during the testing session.

2.5.2.2 Implicit and Explicit Help Calls

During examinee test administration, normal administration activities can be interrupted to accommodate situations involving an examinee's need for assistance. These situations are either implicit help requests where the software of the ET station infers that the examinee needs assistance

or explicit help examinee requests where the examinee presses the red HELP key on the EID (see Figure 10). These types of help requests are discussed below.

In an implicit help call, the test administration software concludes that the examinee needs assistance because the examinee either is taking too long to answer (and confirm a response to) an item or does not seem to be pressing valid keys. The ET station computer checks the maximum allowable time for responding to an item (Table A-1), halts test administration as soon as this time limit is reached, and freezes all test administration timers (for total test time, total time on an item, etc.) for that examinee. Likewise, if the implicit help call is initiated because the examinee pressed too many invalid keys, the test administration timers are frozen and test administration is halted. In either case, the ET station software records certain implicit help call information for subsequent recording within the examinee SSAN file (discussed in Section 2.5.2.1). The first implicit help screen informs the examinee to call the TA, who submits a pass code to permit the dialogue to continue. Once the correct pass code is submitted, the ET station continues a dialogue with the TA, providing information on the cause for the call and suggested remedial action. Test administration continues with the examinee repeating the test practice item and the test item during which the help call was initiated. As soon as the original item is displayed, the test administration timers resume recording response latency information and the test continues normally.

In an explicit examinee help call, similar logic halts test administration, freezes the timers for the examinee, and solicits the assistance of the TA. A menu of possible causes for examinee's help request is displayed to the TA, who selects the most appropriate option from the list (e.g., the examinee did not understand the instructions, the examinee was ill, the examinee believed the computer was failing, the examinee wished to terminate testing). Based on the response of the TA, the test administration software selects one of the following options: (1) continues test administration after a repeat of the practice question (e.g., the examinee did not understand how to take the test), (2) continues test administration with the item underway when the HELP key was pressed (e.g., the examinee accidentally pressed the HELP key), (3) terminates test administration for the examinee (e.g., the examinee refuses to continue the test), or (4) continues testing at another ET station (e.g., the ET station seems to be failing). If testing is continued at another ET station, testing resumes at the beginning of the first uncompleted test. In either case, as soon as the correct item is displayed, the test administration timers continue with the recording of response latency information and the test continues normally.

2.5.3 Data Handling Computer (DHC)

Software development was less critical for the DHC than for the ET and TA stations because the DHC serves primarily as a manager of examinee testing data *after* test administration. The DHC has two primary functions:

- 1. Data compilation.** The DHC compiles and organizes examinee testing data recorded on the Data Disks from the testing sites. Data recorded on a Data Disk must be removed and stored on a nonvolatile medium for subsequent communication to users of the CAT-ASVAB system. Appropriate backup mechanisms must be in place before data are purged from a Data Disk; once purged of its data, the Data Disk is returned to a testing site for reuse.

- 2. Data distribution.** The DHC must be able to communicate the examinee testing data to users of the system. Specifically, an extract of each examinee's testing record must be

communicated to the USMEPCOM (System 80) minicomputer at the parent MEPS (Figure 4). In addition, all of the examinee testing data must be communicated to NAVPERSRANDCEN for software quality assurance processing and communicating the data to other users of the CAT-ASVAB system.

DHC software must also ensure that the DHC collects each examinee's testing data only once and distributes each compiled data set only once to each user. An override mechanism must be available to send the information again if the original information is lost in transit. Finally, it must be possible for the DHC to recover from a hardware failure.

In order to accommodate the functions that the DHC must support and yet provide sufficient backups for various components of the system, the Hewlett Packard Integral Personal Computer (HP-IPC) was again used as the computer upon which to develop a DHC. The minimal hardware configuration needed at each MEPS to support DHC operations is outlined below:

1. One HP-IPC with 1.5 MB of available RAM with a standard keyboard (for backup purposes, an ET station using the spare standard keyboard is available with each LCN).
2. Two 55 MB hard disks (labeled primary and backup) with only one connected to the HP-IPC at a time.
3. One tape drive capable of processing 67 MB tape cartridges.
4. One serial interface board for asynchronous communications with a Sperry microcomputer (or equivalent). Subsequently, the Sperry microcomputer is used to communicate with the MEPS System 80 minicomputer (Figure 4).
5. One "mouse," which the DHC operator can use in place of the standard keyboard.

This equipment uses the HP-IPCs already at the MEPS. If the DHC fails, it is possible to borrow an ET station from the CAT-ASVAB testing room at the MEPS to perform DHC functions. In addition, one 55 MB disk serves as the mass storage device for data received from the testing sites (via Data Disks) until communicated to NAVPERSRANDCEN. The second 55 MB disk serves as the backup storage device. The tape drive is used to generate tape cartridges containing examinee testing data. The DHC operator periodically downloads examinee testing data from the 55 MB disk onto the tape cartridge, which is sent to NAVPERSRANDCEN. The serial interface board is used to transfer an extract of each examinee's testing data record via a Sperry microcomputer to the MEPS minicomputer. Finally, a mouse is available (but not required) for use by the DHC operator to serve as an alternative to the keyboard.

Initially, functional requirements documents (Folchi, 1986; Folchi & Rafacz, 1986) were developed to isolate the critical logic flow paths in the software and to establish concurrence with the DHC user—an individual at the MEPS.

As with the ET and TA station software packages, the actual software development involved isolating code segments for certain critical functions and then calling these routines as

needed from a main driver. From the standpoint of the operator, this reduces the complexity of the functions that were required. Figure 12 shows the DHC main menu; Figure 13 shows the DHC auxiliary menu displayed (the AUXIL option in Figure 12). The more critical DHC functions and their impact upon DHC operations are discussed below.

DHC PROGRAM MAIN MENU		
f1:	PROCESS	Process Data Disks.
f2:	DISTRIB	Generate Database Distribution Tapes.
f3:	TRANSFER	Transfer Examinee Extracts to Sperry PC.
f4:	BACKUP	Generate DHC BACKUP Tape.
f5:	SELECT	Select Extracts to re-send to Sperry PC.
f6:	AUXIL	DHC Auxiliary Menu.
f8:	QUIT	Quit the DHC Program.

PROCESS	DISTRIB	TRANSFER	BACKUP	SELECT	AUXIL		QUIT
----------------	----------------	-----------------	---------------	---------------	--------------	--	-------------

Figure 12. Main menu of DHC functions.

DHC PROGRAM Auxiliary Menu		
f1:	PURGE	Erase Data Disks that have been processed.
f2:	DELETE	Delete Backup Database Distribution files that have been verified by NPRDC.
f3:	PRINT	Print the DHC LOG file.
f4:	RESET	Configure the DHC for new MEPS locale.
f5:	REPORT	Print REPORT about Examinee Extracts on DHC.

PURGE	DELETE	PRINT	RESET	REPORT			MAIN MENU
--------------	---------------	--------------	--------------	---------------	--	--	------------------

Figure 13. DHC auxiliary menu.

1. Normally, the operator uses the menu options in the order shown in Figure 12. PROCESS permits the operator to process the Data Disks received from the testing sites; it instructs the operator to insert a Data Disk for processing. The software confirms that a Data Disk was inserted and that this Data Disk had not previously been processed (preventing the accidental transfer of duplicate data). A single session file on the Data Disk (identified by the date and starting time of the session for some MEPS/MET site) includes the examinee testing data records for all examinees tested during the session. An examinee's testing data record is read from the session file and then some portion of that data is appended to the following three files on the 55 MB disk: an Examinee Extract file, an Examinee Extract Archive (EEA) file, and a Database Distribution file. The Examinee Extract file contains a 153 byte portion of the total (approximately 3500 byte) examinee testing record found in a session file on the Data Disk. This extract includes information specifically required by USMEPCOM to complete the enlistment processing of the examinee at the MEPS. Periodically, the Examinee Extract file is sent to the MEPS minicomputer using the TRANSFER option described below. The EEA file is essentially identical to the Examinee Extract file except that each examinee record includes the date the testing data extract was sent to the MEPS minicomputer. Finally, the Database Distribution file includes the full complement of examinee testing data as recorded on a Data Disk. As examinee data recorded on Data Disks are processed by the PROCESS option, these data are appended to the Database Distribution file; consequently the file grows in size.

2. The DISTRIB option (see Figure 12) permits the operator to perform one of two functions. The first function involves downloading the examinee testing data in the Database Distribution file onto a cartridge tape using the DHC 67 MB tape drive. If this function is selected, the operator sends (via registered mail) the cartridge tape to NAVPERSRANDCEN. Once the information is successfully copied to the cartridge tape, the Database Distribution file on the 55 MB disk is renamed using the current date. This procedure permits the retention of the information in the Database Distribution file on the primary disk but also prevents new examinee data from being confounded with examinee data already sent to NAVPERSRANDCEN.

If the cartridge tape sent to NAVPERSRANDCEN is lost in transit or is defective, the operator can invoke the second function of the DISTRIB option (i.e., re-create the original cartridge tape of information). The Database Distribution file was renamed to the current date when the original cartridge tape was created. The second function of the DISTRIB option enables the operator to download the information created on that date to a cartridge tape. This procedure provides a reliable backup mechanism for examinee testing data stored on the 55 MB disk.

3. The TRANSFER option (Figure 12) permits the DHC operator to transfer examinee extract data, recorded in the Examinee Extract file, to the MEPS minicomputer. This option requires the operator to follow a set of procedures that move the Examinee Extract file first to a Sperry microcomputer and then to the MEPS minicomputer (Folchi, 1990a). Once these processes are completed, the DHC software attaches each examinee extract record with the current date and then appends the current Examinee Extract file onto the EEA file. The Examinee Extract file is then purged, ready for new examinee extract data. Therefore, the EEA file is essentially a backup of examinee extract information that the DHC operator can use to select the examinee records in the EEA file that need to be re-sent to the MEPS computer.

4. To protect the information on the primary 55 MB disk in case of failure, the DHC operator must periodically dump the entire contents of the primary 55 MB disk onto a cartridge tape. After the DHC operator uses the BACKUP option (Figure 12) to create the backup cartridge tape,

the DHC hardware can easily be reconfigured to recover the information on the primary disk and continue DHC processing at the point of failure.

The DHC software also creates a chronological record of operator requests in a Log file starting with the operator's last use of the BACKUP option. Therefore, once the backup 55 MB disk is installed with the data from the cartridge tape of information originally on the primary 55 MB disk, the operator can interrogate the Log file to recover to the point of failure of the primary disk by duplicating the operations recorded in the Log file.

5. The SELECT option (Figure 12) enables the DHC operator to select certain examinee extract records from the EEA file for communication to the MEPS minicomputer. Even though the examinee records in the EEA file are already communicated to the minicomputer, it could be necessary to re-send any information that was lost or corrupted in transit. The DHC operator can select information from the EEA file on the basis of either the examinee's SSAN or the approximate time interval during which the information was originally sent to the MEPS minicomputer. The selected examinee extract data are written to the Examinee Extract file and re-sent to the MEPS minicomputer during the next request of the TRANSFER option.

6. Selection of the AUXIL option (Figure 12) displays the auxiliary menu screen (Figure 13), a collection of subordinate functions that support the main menu functions (Figure 12). The PURGE option permits the operator to remove all of the examinee testing data on a Data Disk prior to it being reused at a field testing site. In practice, the PURGE option is used only on Data Disks processed prior to the generation of a backup cartridge tape, ensuring that no examinee testing information is lost in the purging process. The DHC operator uses the DELETE option to remove files from the 55 MB disk containing examinee testing data that have already been sent to NAVPERSRANDCEN. These files are identified by the date the cartridge tape was created with the examinee testing information. In practice, the subject file is removed only when NAVPERSRANDCEN verifies the receipt and integrity of the information recorded on the cartridge tape for the date in question. The PRINT option provides the DHC operator with a hard-copy listing of the contents of the Log file for purposes of failure recovery. The RESET option allows the operator to configure the DHC hardware for a new MEPS CAT-ASVAB testing site with identifying information, if the equipment must be moved to a new MEPS. Finally, the REPORT option permits the DHC operator to prepare reports on: session files transferred to the MEPS minicomputer in the last 7 days, session files waiting to be transferred to the minicomputer, examinee extracts sent to the minicomputer in the last 7 days, and/or examinee extracts waiting to be transferred to the minicomputer.

2.6 ACAP System Documentation

For each of the three software systems (TA station, ET station, and the DHC), user/operator manuals, programmers reference manuals, and system test plans were developed and kept up to date.

To support the use of the ACAP system at selected MEPS sites in an operational mode (and provide examinee scores of record), the following documentation was developed in accordance with DOD-STD-7935 (*DOD Standard: Automated data systems [ADS] documentation*, 1983): (1) for the ACAP system, functional description, system/subsystem specifications, data requirements, and data element dictionary (four documents); (2) for the TA station, ET station, and DHC software

systems, a programmer's maintenance manual and a system test plan (six documents); (3) a user's manual to accommodate all of the ACAP software systems (one document); and (4) an operations manual for the TA station and the DHC (two documents).

2.7 ACAP System Testing Procedures

To minimize software errors, "C" was chosen as the programming language for ACAP. "C" language supports structured programming (including concise definitions), permits fast access to data structures, and includes a repertoire of debugging aids. It is also reasonably portable from one system to another while permitting fast execution time of the compiled code.

The programming standards and practices were also selected to minimize the chances of creating errors. For example, the software was designed as modular units with minimal interaction among the units. The modules are executed by a main "driver" program that controls the sequence of executions and verifies the results, if necessary. Strict adherence to appropriate software development standards (e.g., DOD-STD-2167A) also minimized errors.

The software was tested to locate the errors, errors were corrected, and the software was retested until no errors were found. To address the concern that not all the errors were found, the ACAP specified built-in-test (BIT) software. The BIT procedures for the ET station (the most logically complex package) include adding software that can read examinee responses directly from a separate (scenario) file. This scenario file also includes predetermined response latencies for test items as well as various testing times for the tests. By using such scenarios, a multitude of different logic flow paths and testing configurations can be evaluated without involving real examinees. Once a scenario is completed, the system tester can survey the output data to confirm that the recorded information matches that specified in the scenario. For the most part, any differences in information can be attributed to software errors, which are quickly located and corrected. Using BIT techniques minimizes the amount of time required to test many logic paths within the software.

The system testing procedures are documented in Folchi (1990b), Huynh (1990), Jones-James (1990), and Rafacz and Winfred (1988).

3.0 Discussion and Conclusions

The ACAP system effectively meets the requirements for a CAT-ASVAB system of computers to support the CAT-ASVAB technology.

The ACAP system is currently being used as the delivery vehicle for CAT-ASVAB (Naval Regional Contracting Center, 1984). In some areas, the ACAP system exceeds the original requirements (e.g., the build time for the display of items at an ET station). In a few areas, the ACAP system falls short of the requirements, (e.g., original telecommunications requirement). The extent to which the ACAP system meets the requirements of the nine factors listed in Appendix C of the Stage 2 RFP (Naval Regional Contracting Center, 1984) and the ACAP system requirements is discussed below.

1. Performance. Performance includes the functions, equipment, and software to provide an automated, adaptive, and psychometrically acceptable replacement for the P&P-ASVAB production testing program.

- As a result of recent psychometric requirements, CAT-ASVAB needs the entire 1.5 MB of RAM available at each ET station. However, it is possible to expand the available RAM to 2.5 or 4.5 MB without modifying the size, weight, or networking capabilities of an ET station to meet additional capacity requirements.
- Microfloppy disk transfer between the TA station and DHC is both efficient and responsive to the data transfer requirements of the CAT-ASVAB program, and hence, the telecommunications requirements of the Stage 2 RFP were relaxed for ACAP.
- When an examinee presses the HELP key, he or she is instructed to "raise his or her hand" for assistance. This procedure has been effective at ACAP testing sites. The automated HELP call required by the Stage 2 RFP would have been ineffective and complicated to implement.
- Table 2 estimates the maximum system "build" time observed among three generic scenarios of examinee response alternatives for the nine CAT-ASVAB power tests; the speeded test item build time was (for all practical purposes) almost instantaneous.

Table 2
Maximum System Build Time for
CAT-ASVAB Power Tests

CAT-ASVAB Test	MAX Build Time (seconds)
General Science (GS)	0.73
Arithmetic Reasoning (AR)	0.70
Word Knowledge (WK)	0.55
Paragraph Comprehension (PC)	1.17
Auto Information (AI)	0.80
Shop Information (SI)	0.83
Mathematics Knowledge (MK)	0.85
Mechanical Comprehension (MC)	0.98
Electronics Information (EI)	0.76

The build time is the elapsed time between the last valid examinee keypress for the last item displayed and the complete display of the next item. The build time includes the time required to: (a) score the previous item, (b) perform the appropriate mathematical routines to determine the new ability estimate, (c) search the appropriate information table for the next item to be administered, (d) retrieve

the item level data from RAM and build the screen for the new item, and (e) display the new item. Step d is very demanding for the MC test because every item includes a graphic that is decompressed from the information recorded in RAM before the new item can be built for display. However, the PC test is most demanding because the large amounts of text that must be displayed on the screen require more portions of the bit-map image of the screen to be accessed than do the other tests. The build time estimates (Table 2) are well within the Stage 2 RFP requirements (maximum of 3 seconds) and are independent of the number of examinees being tested during a testing session, as required.

2. Suitability. Suitability includes the aspects of the CAT-ASVAB system that make it suitable for use by USMEPCOM. These aspects include environmental specifications and TA "start-up" time.

- The HP-IPC effectively meets all of the critical environmental requirements. The most critical environmental specifications relate to the humidity requirements. The Stage 2 RFP specifies that the operating relative humidity range from 50% $\pm$ 10% to 94% $\pm$ 4%. The HP-IPC meets this requirement; however, commercially available microfloppy disks require a relative humidity range of approximately 8% to 80% for successful operation. Therefore, the ACAP system (regardless of the microcomputer used) must operate within the range of relative humidity required by the microfloppy disks.
- The Stage 2 RFP specified that operators would need no more than 4 hours of training to operate the system. TAs require at least 1 week of training to use the system effectively (RGI, Inc., 1988). Even though the ACAP system includes many user-friendly features (e.g., error trapping, on-line help messages), potential TAs need a number of practice sessions before attempting to use the system in an operational mode. In addition, TAs must learn to perform the management duties associated with military testing (e.g., security procedures for testing materials, controlling examinee movements, conducting the necessary liaison activities). All of these requirements are time-consuming.

3. Reliability. Reliability describes the specifications that are concerned with consistent use of the CAT-ASVAB system in the MEPS/MET environment.

- The ACAP system meets most of the Stage 2 RFP specifications for reliability. During the Score Equating Development (SED) data collection effort, testing data for 18 examinees (of over 5500 examinees completing the entire CAT-ASVAB test) were corrupted and, therefore, unusable. Most of these 18 instances were attributed to certain software errors, which were corrected during later phases of the data collection.
- It is almost impossible for an examinee's CAT-ASVAB test not to be completed once the examinee starts the CAT-ASVAB, assuming the availability of electrical power. This high estimate of reliability is attributed to the failure recovery mechanisms available to ACAP system TAs. Every ET station is a backup for every other ET station, and an ET station may be used as a TA station. In addition, the failure of any one station does not affect the performance of any other station.

Easy-to-use recovery procedures have been built into the software to permit recovery within 5 minutes (assuming the availability of a station). Likewise, the DHC includes sufficient backup devices at the MEPS for an easy recovery due to component failure.

4. Maintainability. Maintainability refers to specifications related to maintaining the system when in use at CAT-ASVAB testing sites throughout USMEPCOM.

As the ACAP system was intended for limited—not nationwide—deployment at several MEPS/MET sites, many of the Stage 2 RFP specifications—particularly those related to maintainability—do not apply to ACAP. However, ACAP includes many of the desired maintainability features. In particular, no special technical skills are required to operate the system or provide the routine preventative maintenance for computers equipped with microfloppy disk drives. The TA has the option of employing self-diagnostic hardware tests on bootup of the ET station. The ACAP system includes several comprehensive self-diagnostic tests for the electronic network supporting the LCN. In addition, clear directions are provided to the TA to resolve some system faults and continue the testing session. Finally, the ET station interactive screen dialogues (Rafacz & Moreno, 1987) illustrate many of the software controls used to monitor the examinee testing process and to accomplish failure recovery.

5. Ease of Use. Ease of use refers to ease of use of the CAT-ASVAB system by TAs with little or no computer experience and by examinees.

- The ACAP system can monitor the status of up to 30 ET stations (each administering the CAT-ASVAB test to an examinee) simultaneously, exceeding the specification of 24 ET stations.
- Each transportable package for an ET or TA station weighs 25 pounds and each generic LCN weighs about 190 pounds in eight transportable packages, including supporting testing materials and supplies. This did not meet the specified weights of 23 and 120 pounds, respectively.
- The HP-IPC displays alphanumeric characters in a variety of dot matrix dimensions or font sizes. The font size approved by the Joint-Services for ACAP includes dot matrix dimensions of 5 X 8 dots embedded in a 7 X 11 field. At this resolution, it is possible to display 23 lines with at least 73 characters per line on the HP-IPC screen. The HP-IPC screen includes 512 picture elements horizontally by 255 picture elements vertically. In addition, the screen on the HP-IPC uses electroluminescent technology, thus making it independent of ambient light within the class of flat panel displays. Therefore, the screen on the HP-IPC is relatively easy to read in the wide range of lighting conditions at the testing sites used by USMEPCOM.

6. Security. Security refers to issues relevant to the processing of sensitive defense information and personal information protected by Privacy Act of 1974.

- The ACAP system employs an extensive system of security procedures (RGI, Inc., 1988) to address the requirements of the Privacy Act of 1974, of the Freedom of Information Act, and for the control of military testing materials. At

the beginning and end of an ACAP testing session, the TA is required to sign a log sheet for the receipt of microfloppy disks (identified by serial number) and related testing materials. The transfer of materials is always verified by a second person. During the testing session, all materials are under close physical scrutiny by the TA; at all other times, the testing materials are secured in a double-lock safe. Examinee testing information, recorded on the Data Disks, is sent by registered mail to the parent MEPS for further processing by the DHC. Cartridge tapes prepared by the DHC for subsequent communication are also sent by registered mail. In addition, *every* disk used by ACAP is registered with a unique serial number and controlled by appropriate logs and labels identifying the disk as containing personal information, if appropriate. ACAP uses registered mail to transfer disks within the computer system.

- To protect against unauthorized access of the test item data stored on an ACAP system disk, all three of the following events must occur for access: The individual must know the key (i.e., the password), the individual must have a copy of the ACAP system disk, and the individual must have exact information on the ACAP encoding algorithm, which is not public information (in contrast to the Data Encryption Standard [DES] algorithm [U.S. Department of Commerce, 1988], which is public information; under DES, an individual need only gain access to the key and the data). An encoding algorithm enables project managers to provide a high level of security for the test items without the cost of having a "key operator" responsible for key security at each testing site. In summary, the ACAP system provides greater security than required by Stage 2 RFP.

7. Affordability. Affordability refers to cost effectiveness. The objective is for the CAT-ASVAB system costs to compare favorably to P&P-ASVAB costs over the projected CAT-ASVAB system life cycle.

Affordability is really not an issue due to the limited scope and purpose of the ACAP. The ACAP system was developed with the intent of refining the operational requirements for deploying such a system and providing a vehicle for the collection of data in support of the psychometric research for equating CAT-generated scores with those of the ASVAB. The ACAP uses commercially-available computer systems: About 460 HP-IPCs were procured. However, thousands of computers would be required for a nationwide implementation. In addition, while the HP-IPC has enjoyed a relatively long life on the commercial market (about 5 years), it is no longer being marketed by Hewlett Packard.

8. Expandability/Flexibility. Expandability/flexibility describes the characteristics of the CAT-ASVAB system that allow rapid and effective growth to meet changing USMEPCOM requirements over the system's operational life cycle.

- In terms of maximum supported RAM, the HP-IPC can be expanded (in increments of 1 MB, 2 MB or 4 MB) to a total of 7.5 MB of addressable RAM. Three fourths of the available RAM can be configured as a high-speed, electronic RAM-disk drive, in addition to the standard 710 KB microfloppy disk.

- In terms of expandability or peripherals that can be added to the HP-IPC, several interfaces are available. Two Human Interface Loop (identified as HP-HIL) interfaces at the front of the HP-IPC allow the addition of serial devices such as a keyboard, track-ball, mouse, control dial module, graphics tablet, bar code reader, and digitizer.
- In terms of flexibility, the ACAP computer hardware and software system can easily accommodate unpredicted needs at USMEPCOM testing sites based on an analysis of the ways in which the software design uses the hardware resources of the HP-IPC. For example, the RAM of the TA station can be used as a file server, eliminating the need for a mechanical device to serve this purpose. In addition, the ACAP LCN operates in either a networking or a stand-alone mode of operation. In either mode of operation, the ET station operates independently of any other station in the LCN during examinee CAT-ASVAB testing. In a station failure situation, the ACAP LCN is very flexible, as any station can serve as a backup for any other station at any time during the testing session. These and other features of the software supporting the ACAP LCN, developed for the HP-IPC, point to the ability of the system to accommodate most USMEPCOM testing situations, with minimal resources needed for moving among those situations.

9. Psychometric Acceptability. Psychometric acceptability refers to the capability of the CAT-ASVAB system to accommodate the psychometric requirements for implementing the CAT-technology and all of its associated specifications.

- The psychometric requirements for installing the CAT-technology in the ACAP computer delivery system have evolved and grown since the beginning of the ACAP (i.e., since the end of the Stage 2 RFP). The most important specifications are discussed in the appendix. To summarize, the ACAP system is responsive to the requirement for psychometric acceptability. This is evidenced by the performance of the ACAP system in supporting the SED and SEV data collection efforts (February through December 1988 and September 1990 through April 1992, respectively). The SED data collection effort included CAT-ASVAB testing of over 5500 examinees and the SEV effort included testing of almost 7000 examinees. Very little of the information collected was compromised by the hardware or software of the ACAP system. The data compromises seem to have been software related; these types of errors did not reappear in later phases of the data collection effort under revised versions of the software. Additional SED results were documented by RGI, Inc., (1989).
- The ACAP also supported the data collection efforts for the SEV phase of the project. Examinee "scores of record" were collected at the same sites that were visited during SED. This was made possible by the psychometric analysis of SED data and the production of equating tables to transform the CAT-ASVAB generated scores into a number-right score for classification and assignment (Segall, 1990). The HP-IPC has accommodated SEV psychometric requirements because of enhancements made to the SED software. With these SEV enhancements to the ACAP system, the functional requirements specified by the Joint Services (during and after the Stage 2 RFP) have been effectively realized in the ACAP computer system.

4.0 Future Efforts

Any future attempts at developing a CAT-ASVAB system should include the supporting studies and documentation necessary for a full-scale development RFP. ACAP can greatly assist in this requirement. The Joint Services must select a concept of operation for the future CAT-ASVAB system.

The CAT-ASVAB program must decide on a procurement strategy for a future implementation effort. The ACAP strategy of procuring off-the-shelf computer systems should be examined. This strategy avoids the enormous cost, time, and complexity associated with the Government building a CAT-unique hardware system and makes the equipment immediately available for software development.

Any further effort to procure a replacement ACAP computer system should consider the merits of using a large RAM-based design for ET/TA stations. In addition, future software design and development efforts should parallel those used within ACAP, as much as is possible, to make maximum use of software already developed and tested within ACAP.

References

- Department of Defense. (1983, February). *DOD Standard: Automated data systems (ADS) documentation standards* (DOD-STD-7935). Washington, DC: Author.
- Department of Defense. (1988, February). *Military Standard: Defense system software development* (DOD-STD-2167A). Washington, DC: Author.
- Folchi, J. S. (1986). Communication of computerized adaptive testing results in support of ACAP. *Proceedings of the 28th Annual Conference of the Military Testing Association*, 28, 618-623.
- Folchi, J. S. (1990a). *An operator's manual for the data handling computer during score equating verification*. San Diego: Navy Personnel Research and Development Center.
- Folchi, J. S. (1990b). *A system testing plan for the data handling computer*. San Diego: Navy Personnel Research and Development Center.
- Folchi, J. S., & Rafacz, B. A. (1986). *Functional requirements for the data handling computer (ACAP)*. San Diego: Navy Personnel Research and Development Center.
- Huynh, C. (1990). *System testing plan for the recruit allocation system program (RASP)*. San Diego: Navy Personnel Research and Development Center.
- Jones-James, G. (1986). Design and development of the ACAP test administration software. *Proceedings of the 28th Annual Conference of the Military Testing Association*, 28, 612-617.
- Jones-James, G. (1990). *System testing plan for the examinee testing station*. San Diego: Navy Personnel Research and Development Center.
- Lord, F. M. (1980). *Applications of Item Response Theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), *New horizons in testing* (pp. 223-236). New York: Academic Press.⁴
- Moreno, K. E. (1987). *Military applicant testing: Replacing paper-and-pencil with computerized adaptive tests*. Paper presented at the 1987 American Psychological Association Conference.⁴
- Naval Regional Contracting Center. (6 June 1984). *CAT-ASVAB Stage 2 Request for Proposal (RFP)* (Document number N00123-84-R-0637). Long Beach, CA: Author.
- Owen, R. J. (1969). *A Bayesian approach to tailored testing* (Res. Rep. 69-92). Princeton, NJ: Educational Testing Service.⁴
- Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of American Statistical Association*, 70, 351-356.⁴

⁴Cited in the appendix.

- Quan, B., Park, T. A., Sandahl, G., & Wolfe, J. H. (1984). *Microcomputer networks for computerized adaptive testing (CAT)* (NPRDC-TR-84-33). San Diego: Navy Personnel Research and Development Center. (AD-140 256)⁴
- Rafacz, B. A. (1986). Development of the test administrator's station in support of ACAP. *Proceedings of the 28th Annual Conference of the Military Testing Association*, 28, 606-611.
- Rafacz, B. A. (1990). *A test administration user's manual for the ACAP System*. San Diego: Navy Personnel Research and Development Center.
- Rafacz, B. A., & Moreno, K. (1987). *Interactive screen dialogues for the examinee testing (ET) station*. San Diego: Navy Personnel Research and Development Center.
- Rafacz, B. A., & Winfred, G. G. (1988). *Test administrator station system testing plan*. San Diego, CA: Navy Personnel Research and Development Center.
- RGI, Incorporated. (1988). *Administering the CAT-ASVAB, reference guide* (Deliverable under Contract Number N66001-87-D-0010). San Diego: Author.
- RGI, Incorporated. (1989). *Data collection for ACAP score equating* (Deliverable under Contract Number N66001-87-D-0010). San Diego: Author.
- Sands, W. A. (1985). An overview of the accelerated CAT-ASVAB program. *Proceedings of the 27th Annual Conference of the Military Testing Association*, 27, 19-22.
- Sands, W. A. (1986). Computerized Adaptive Testing hardware/software development for the U.S. military: Introduction and overview. *Proceedings of the 28th Annual Conference of the Military Testing Association*, 28, 598-600.
- Sands, W. A. (1987). The Joint-Service CAT-ASVAB program: Progress and plans. *Proceedings of the 29th Annual Conference of the Military Testing Association*, 29, 335-338.
- Sands, W. A. (1988). Computerized adaptive testing: The CAT-ASVAB program. *Proceedings of the 30th Annual Conference of the Military Testing Association*, 30, 458-462.
- Sands, W. A. (1989). Computerized Armed Services Vocational Aptitude Battery: Progress and plans. *Proceedings of the 31st Annual Conference of the Military Testing Association*, 31, 541-546.
- Sands, W. A. (1990). Joint-Service computerized aptitude testing. *Proceedings of the 32nd Annual Conference of the Military Testing Association*, 32, 245-250.
- Segall, D. O. (1987). *ACAP item pools: Analysis and recommendations*. San Diego: Navy Personnel Research and Development Center.⁴
- Segall, D. O. (1989). *Specifications for scoring the CAT-ASVAB*. San Diego: Navy Personnel Research and Development Center.

⁴Cited in the appendix.

- Segall, D. O. (1990). *Score equating development analyses of the CAT-ASVAB*. San Diego: Navy Personnel Research and Development Center.
- Sympson, J. B., & Hetter, R. D. (1985, October). *Controlling item exposure rates in computer adaptive testing*. Paper presented at the 27th Annual Conference of the Military Testing Association, San Diego.⁴
- Tiggle, R. B., & Rafacz, B. A. (1985). Evaluation of three local CAT-ASVAB network designs. *Proceedings of the 27th Annual Conference of the Military Testing Association*, 27, 23-28.
- U.S. Department of Commerce. (1988, January). *Data encryption standard* (FIPS PUB 46-1). Gaithersburg, MD: Author
- U.S. MEPCOM mobile examining team site requirements for computerized adaptive testing (1983). Director of Testing, USMEPCOM Letter Report dated 22 December 1983. Vols. I, II, and III.
- Wilson, L. D., Bebb, L. E. D., Higgins, J. W., Boston, G. D., & Agee, G. (1989). *Data base documentation for score equating* (deliverable under Contract Number N66001-87-D-0010). San Diego: RGI Incorporated.

⁴Cited in the appendix.

Appendix
An Overview of Computerized Adaptive Testing

An Overview of Computerized Adaptive Testing

The purpose of this appendix is to provide an overview of the computerized adaptive testing (CAT) technology as used for the accelerated CAT version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) Project (ACAP). In contrast to the paper-and-pencil (P&P) methodology for test administration, CAT is based on the strategy that any item presented to an examinee should be the *best* item for that examinee. The item is best if it provides maximum information about the examinee's ability. Lord (1980) provides the theoretical foundation (within the framework of item response theory) for the amount of information given by an item at a particular ability level. In general, the information is a function of the difficulty, discriminating power, and guessing parameters associated with the items in the pool, and conditional on the current estimate of examinee ability. For each interval of examinee ability (using the midpoint of the interval as the target ability estimate), it is possible to develop a set of information values for each item in the pool, ranked from the highest to the lowest. When considered over the entire range of ability for examinees, it is possible to develop an information table about examinee ability for items in the pool. This table is called the information table for an item pool on a certain dimension of ability (e.g., verbal, mechanical, mathematical). The i,j th entry of this table contains the i th-ranked item from the pool in terms of information value, while the j th column refers to an interval of examinee ability within the range of -2.25 to +2.25. ACAP employs 37 intervals, each of width 0.125, and zero (average ability) centered on the 19th interval.

To clarify this concept with a case in point, all examinees receive the same items for a test in P&P testing. Therefore, examinees of low ability encounter very difficult items that they cannot answer, contributing to their frustration and anxiety and wasting their time. Little information is derived from their attempt at the difficult items, as nearly all low ability individuals would get the items incorrect. That is to say, difficult items offer little information in terms of discriminating among low ability individuals. Conversely, high ability examinees would get almost all easy items correct; therefore, little information about their ability is obtained from administering easy items. Easy items do not contribute to discriminating among high ability examinees. In summary, it would seem that a more effective testing strategy is to selectively administer items to examinees as a function of the current estimate of their ability. This is the basis for CAT—it selects items from a large pool of items in such a way that only those items appropriate for the examinee are presented, which can contribute to a better understanding (estimate) of the examinee's ability. In this way, the test is "adaptive," as the strategy results in the generation of items for the examinee that are somewhat unique for the examinee. Therefore, any two examinees would (normally) receive a different set of items for a test.

To further illustrate the CAT concept of test administration, consider Figure A-1. This figure illustrates the logic inherent to the CAT process as used within ACAP. Initially, an ability estimate of "average" (zero on a scale of from -3 to +3) is associated with the examinee, as the actual ability is unknown and average ability is the initial best estimate. As such, an item of average difficulty is first administered to the examinee. (In the information table for the test, the "best" average-difficulty item to administer is that item with maximal information in the ability interval containing zero.) If the examinee gets this item correct, then a new ability estimate is computed to be higher than zero, the average score. ACAP uses the Owen (1969, 1975) theoretical framework for developing a new ability estimate as a function of the examinee's score on the current item and

preceding items. Note that the *accuracy* of this new estimate would increase as some information on the examinee's ability is available; namely, the response to the first item. Using the new ability estimate, the procedure is to return to the information table and seek that item with maximal information with respect to the interval associated with the new estimate of ability. In general, a correct response results in a more difficult item being chosen and, when answered, an improved estimate of the examinee's ability. Suppose that the examinee gets the new item incorrect. In that case, the ability estimate would drop slightly; however the *accuracy* of the estimate would still increase. In general, the accuracy of the ability estimate increases because more information is gained about the examinee as more and more items are administered. This process continues, as illustrated in Figure A-1, with items answered correctly increasing the ability estimate (resulting in a slightly more difficult item being administered), and items answered incorrectly decreasing the ability estimate (resulting in a slightly less difficult item being administered). In all cases, the *accuracy* of the ability estimate *increases*, or equivalently, uncertainty about the current estimate of ability decreases.

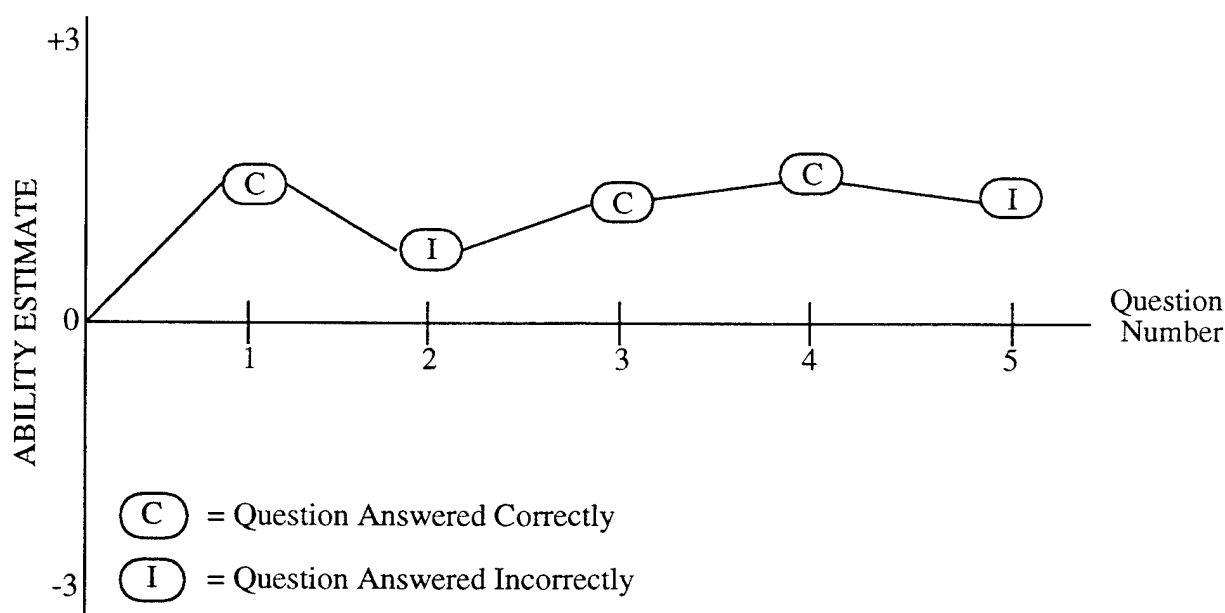


Figure A-1. CAT concept of test administration.

Note. This process may result in the administration of duplicate items. For ACAP, if the new ability estimate is in an interval where the item of maximal information is identical to some previously administered item, then the item of next highest information in the current interval is considered for administration. This procedure continues until an item is found that was not previously administered.

This administration process continues until some "stopping rule" is satisfied (e.g., the accuracy of the resultant ability estimate is sufficiently high, a fixed number of items have been administered, or some hybrid stopping criterion is met). In variable-length testing, the test stops when there is sufficient information. It is interesting to note that a high degree of accuracy can be achieved by administering about 10 items (McBride & Martin, 1983), far fewer than the number of items administered in a typical P&P test. ACAP employs a fixed-length stopping rule, based on the number of items administered.

CAT Testing Within ACAP

The use of the CAT testing technology within the CAT-ASVAB Program (and hence ACAP) is the process of applying the CAT item selection and ability estimation strategy to the current set of P&P-ASVAB tests. In order for the CAT technology to be applied to ACAP, and the CAT-ASVAB program in general, it was necessary to supplement this process with psychometric requirements related to computerized test administration within the military, and the development of improved psychometric procedures in using CAT (Moreno, 1987; Segall, 1987, 1990). A list of the more important set of expanded CAT specifications is provided below:

1. Power and speeded tests. The current P&P-ASVAB test battery consists of eight "power" tests and two "speeded" tests over two pools (or forms) of items. However, as CAT is appropriate only for a power test, the speeded tests (Numerical Operations [NO] and Coding Speed [CS]) have just been automated within ACAP for purposes of computer administration. Table A-1 represents the complete set of ability tests administered in the CAT-ASVAB testing program.

Table A-1

Test Administration Parameters for ACAP

CAT-ASVAB Test	Test Length	Max Item Time (seconds)	Max Test Time (minutes)
General Science (GS)	16	120	8
Arithmetic Reasoning (AR)	16	380	39
Word Knowledge (WK)	16	100	8
Paragraph Comprehension (PC)	11	390	22
Numerical Operations (NO)	50	30	3
Coding Speed (CS)	84	120	7
Auto Information (AI)	11	120	6
Shop Information (SI)	11	110	5
Mathematics Knowledge (MK)	16	220	18
Mechanical Comprehension (MC)	16	240	20
Electronics Information (EI)	16	120	8

The psychometric requirements for CAT-ASVAB demand that all adaptive tests be unidimensional. Hence, there was a need to develop test items for two unidimensional tests (subsequently called the Auto Information [AI], and Shop Information [SI], content areas) from test items previously used for the Auto and Shop Information (AS) P&P-ASVAB test. Therefore, there are nine power tests in the CAT-ASVAB battery.

Table A-1 represents the order in which each test in the CAT-ASVAB is administered during ACAP. Note that the two speeded tests (NO and CS) are administered in the middle of the power tests. A total of 11 tests are administered. In addition, 16 items are administered for each power test, except for Paragraph Comprehension (PC), which require only 11 items to be administered.

2. Power test dimensionality. All of the ACAP power tests (except for General Science [GS]) are treated as unidimensional tests (i.e., only one ability trait is measured by the items). The GS test is divided into three dimensions of ability: life science, physical science, and chemical science. Essentially this means that three information tables must be used when selecting items; one table for each dimension. Items are selected by considering the life science information table for the odd-numbered items (in presentation position sequence), the physical science information table for the even-numbered items, and the chemical science information table for the last item to be presented.

3. Seeded items. In addition to those items that actually contribute to an estimate of examinee ability, experimental or "seeded" power test items are also administered in ACAP. One of the items administered in each power test is a seeded item and is administered randomly as either the 2nd, 3rd, or 4th item in the test. Seeded items are administered without regard to the information table and Owen ability estimation procedures for the sole purpose of gaining response information on these experimental items. Therefore, as an example (see Table A-1), the Arithmetic Reasoning (AR) test includes 15 items that contribute to a measure of examinee ability and one seeded item.

4. Test time limits. Normally all examinees would attempt all power test items in a reasonable time per item. However, for slow-responding examinees, a maximum time limit to respond to the displayed item with a valid (and confirmed) response has been instituted. Table A-1 lists the maximum time limit to respond to an item. For example, if an examinee does not confirm a response to an SI item within 110 seconds of the display of the item, an implicit Test Administrator (TA) help call is initiated (see section 2.5.2.2). Once the TA has determined the nature of the problem (i.e., the examinee is simply responding too slowly), the examinee is readministered the practice question and continues testing on the item that produced the time-out. Likewise, a maximum time limit to complete all the items in each of the tests is imposed (see Table A-1). For example, if an examinee does not complete the PC test within 22 minutes, that test will be stopped automatically and the examinee will immediately continue with the NO test.

5. Interactive dialogs. The wide variety and scope of reading abilities in the military applicant pool required the development and implementation of a unique set of interactive dialogs to use in training and testing military examinees on the ACAP computer equipment. Rafacz and Moreno (1987) documented the subject dialogs. The dialogs were developed to accommodate a sixth-grade reading level; due consideration was given for the "lessons-learned" during a pretest of the ACAP system in October 1987 at the San Diego Military Entrance Processing Station. In

addition, the dialogs are similar to those used in the Navy Personnel Research and Development Center's Apple III Experimental System (Quan, Park, Sandahl, & Wolfe, 1984).

6. Exposure control. In order to avoid overexposure of certain informative items, an exposure rate control process (Simpson & Hetter, 1985) is employed by ACAP. This procedure compares an item's exposure control parameter (e.c.p.) with a random number between 0 and 1. If the random number is less than or equal to the item's control value, the item is a candidate for administration; otherwise, it is rejected for administration and another item is considered. The ACAP procedure for using exposure control parameters in selecting and displaying items is discussed in the next section of this appendix.

7. Test scoring. In the case of power tests, ACAP uses the Owen (1969, 1975) Bayesian procedure to develop a new estimate of examinee ability as a function of the examinee's score on the current item and preceding items. This new estimate of ability is then used in determining the next item for display; the response to the last item in a test can then be used to obtain a final Owen estimate of ability. However, as pointed out by Segall (1990), this process results in the undesirable feature of the final scores depending on the order in which the items were administered. In order to avoid the situation of two examinees being administered the same set of items in a different sequence, giving the same item responses, and yet getting different final scores, a final estimator that is order-independent was required. In the case of the ACAP, the mode of the posterior distribution (i.e., the Bayesian mode) is used. This results in a final (power test) score that is not influenced by the order of item administration, and yet provides slightly greater precision than the Owen estimator.

Speeded test scoring within ACAP is accomplished using a rate-score statistic that is proportional to the proportion of correct responses (corrected for guessing) divided by the geometric mean of the examinee response times for responding to the items. The correction for guessing is necessary to avoid allowing an examinee to get a very high score by simply pressing any valid response key quickly. Finally, a constant scaling factor is used in the formula for the rate-score that permits the rate-score to be interpreted as the number of correct responses per minute. Segall (1990) provides additional information on the scoring procedures used in the ACAP.

8. Scoring of incomplete tests. A practical concern for the ACAP is that not all examinees will complete all test items allowed for a given test. This is true because the tests are timed at a maximum limit (Table A-1) and some examinees (regardless of ability) will respond slowly to test items. The problem is that the Bayesian modal estimate (used to score the power tests) includes a bias that is inversely proportional to test length. In fact, a low ability examinee could achieve a high modal estimate by just answering one or two items. Therefore, to address this concern and discourage the use of a strategy that would compromise the Bayesian modal scoring process, a penalty procedure that includes certain desirable properties was incorporated into the ACAP. Segall (1990) discusses in more detail the development of the penalty procedure when applied to incomplete tests within the ACAP.

9. Producing number-right scores. In order to deploy the ACAP computer system, it was necessary for the output from the ACAP test administration process to produce number-right scores that are interchangeable with those produced by the currently operational P&P-ASVAB program. Use of these ACAP "equated scores" will permit the currently operational selection and

classification procedures to be applied to applicants for all of the military services. This requirement was introduced in Section 2.1. In order to convert the Bayesian modal estimates and rate-scores produced from the ACAP adaptive power and speeded tests (see paragraph 7), with due consideration to incomplete tests, equated scoring tables needed to be developed that permitted the 10 ACAP scores (see Table A-1; AI and SI are combined before conversion) to be converted into number-right scores equivalent to the present form of the P&P-ASVAB. These tables were developed as a result of the Score Equating Development (SED) data collection effort of the ACAP, and implemented during Score Equating Validation (SEV) deployment. The details for applying the equating tables, and producing the desired number-right scores, have been documented by Segall (1989).

Power Test Item Selection and Display

Figure A-2 includes a flowchart of the decision process used within the ET station software to determine which power test item should be displayed to an examinee. Initially, an ability estimate of average is assumed for the examinee (mean = 0, standard deviation = 1). In addition, one of the numbers 2, 3, or 4 is selected at random in order to determine the position (in the sequence [i.e., 1, 2, 3, . . .] of presentation) the seeded item is to be displayed.

The decision logic proceeds from the first to the last item to be presented to the examinee; at each iteration, the information table is consulted to determine the item that meets certain criteria for administration. Once the best item is found, it is presented to the examinee, the response is scored, a new estimate of ability is developed, and the next item is determined. Specifically, considering Figure A-2, at each iteration the first concern is whether the order of presentation matches that of the seeded item. If so, the seeded item is displayed immediately, scored, and the performance of the examinee (answer, response latency, etc.) recorded. If a seeded item is not to be displayed at this point, the information table is searched (in the column corresponding to the current estimate of ability) for the item with the highest information value in that column. If this item has already been displayed, or previously "rejected for display," it is rejected and the next item in the same column of the information table is considered. Finally, an item will be found that was neither previously displayed nor previously rejected for display. A random number is generated (between 0 and 1) and compared with that item's e.c.p.; see paragraph 6. If the random number is greater than the item's e.c.p., that item is marked as "rejected for display" and rejected; otherwise, the item is considered acceptable and displayed. Once displayed, the item is marked as "having been displayed." Eventually, the examinee will have confirmed a response to the item, at which point the item is scored, a new ability estimate is calculated, and certain examinee performance information in responding to the item is recorded. Finally, if the correct number of items for the subject power test has been administered (second column of Table A-1), that test is terminated and the examinee continues with the next CAT-ASVAB test. Otherwise, another item is selected for display.

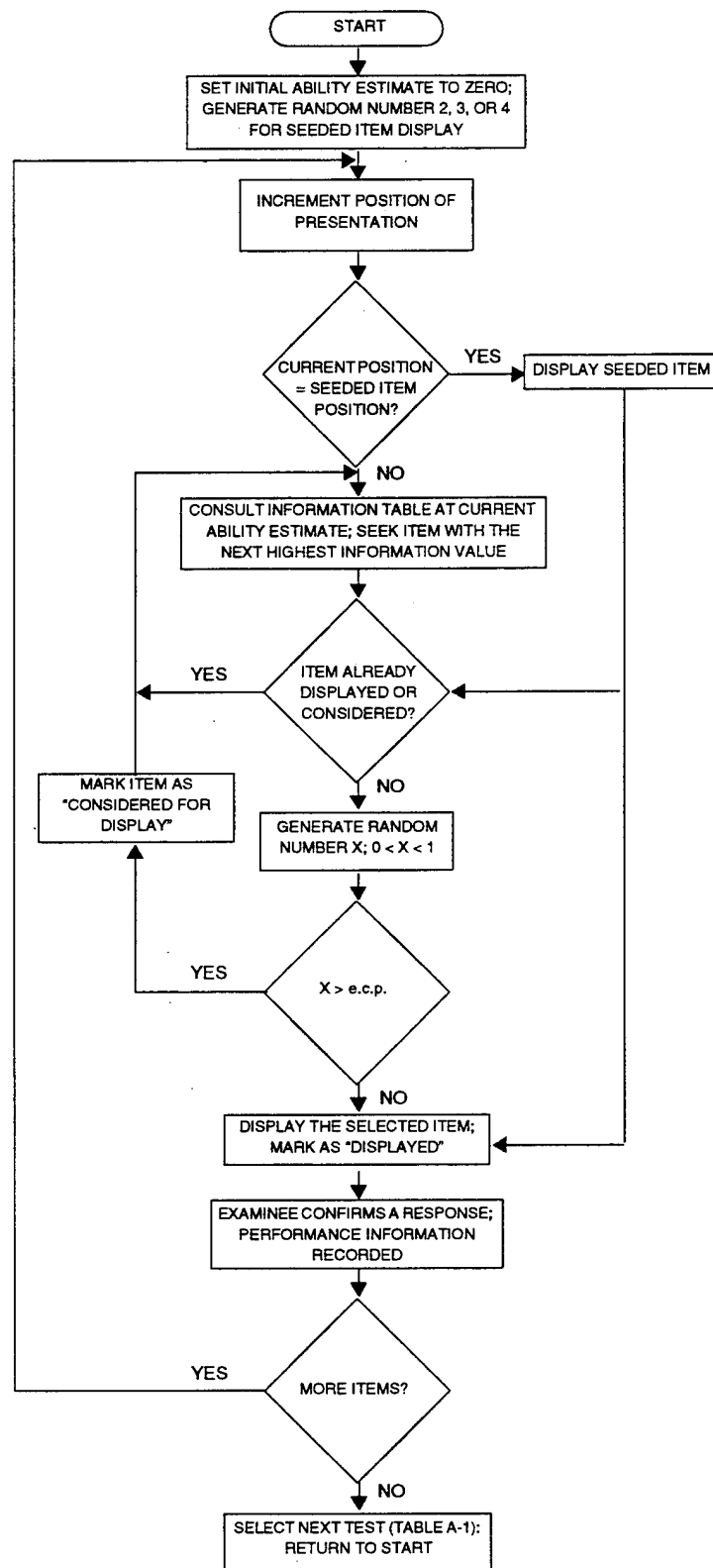


Figure A-2. Flowchart of the power test item selection process.

Summary

To summarize the CAT strategy for ability estimation used in ACAP:

1. Because of the complexity and real-time nature of the test administration process, adaptive tests must be administered by computer.
2. The CAT testing strategy results in different examinees receiving different items, enhancing the security of test items by minimizing the exposure of items.
3. The average time to complete the CAT-ASVAB tests is 1.5 hours; however, virtually all examinees will complete CAT-ASVAB in approximately 2.5 hours.
4. CAT-ASVAB tests are self-paced (i.e., examinees may complete any test at their own pace, up to the maximum test time limits). Therefore, not all examinees will complete a test at the same time.
5. Scoring of the tests is accomplished by computer and, therefore, some final score estimates will be available at the testing site at the conclusion of the CAT-ASVAB.
6. It is possible to collect data on new test items on the computer during the actual test administration of operational items. This process is called "seeding" and during ACAP, one seeded item will be administered with a set of power test items.
7. The accuracy of the scores obtained by using CAT-ASVAB is better than the scores obtained by using P&P testing, particularly for examinees at low and high ability levels.

Distribution List

Chief of Naval Personnel (PERS-23), (PERS-2FF/234)

HQ USMEPCOM/MEPCIM

HQ USMEPCOM/Director of Testing

Defense Technical Information Center (DTIC) (4)