FR-03-46

HumRRO

Human Resources Research Organization

CAT-ASVAB Prototype Internet Delivery System: Final Report

Paul J. Sticha
Galt Barber

Prepared for:

Defense Manpower Data Center
400 Gigling Road
Seaside, CA 93955-6771

Contract Number M67004-00-D-0002
Delivery Order 0020 (STP CAT)

June 2003

66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314 • 703.549.3611 • fax. 703.549.9025
www.humrro.org

Table of Contents

CAT-ASVAB PROTOTYPE INTERNET DELIVERY SYSTEM: FINAL REPORT

This report describes the development of a prototype computer-adaptive version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) that can be delivered over the internet. The internet-delivered adaptive test, or iCAT, was designed to include most of the functionality of the version of the CAT-ASVAB that is currently used at Military Enlistment Processing Stations (MEPS) and Mobile Examination Team (MET) sites as well as the version currently being developed to run on the Microsoft® Windows operating system.

This report documents issues for the development of the iCAT prototype and describes how they were addressed in the system. Relevant issues include the security and performance of the system, as well as its compliance with web site design regulations. Some issues for an operational system, such as compliance with the Privacy Act, do not apply to the prototype. The report discusses these issues and describes their implications on the design and implementation of the operational system. In general, the prototype was developed to assess the technical feasibility of implementing adaptive testing methods over the internet and addressed issues related to operational use of such a test in a more cursory manner.

The report begins by describing the goals for the iCAT and specifying which of these goals were addressed in the prototype. It then discusses specific technical, security, and regulatory issues related to the development of the prototype and briefly describes how these issues were addressed in the prototype design. The discussion also summarizes considerations that, although they do not affect the prototype, must be addressed to implement an operational system. The report concludes with a description of the prototype iCAT.

Background and Project Goals

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple-aptitude test battery used by the Armed Forces to select and classify applicants for military enlistment. The ASVAB includes the following nine subtests:

- Arithmetic Reasoning (AR),
- Word Knowledge (WK),
- Paragraph Comprehension (PC),
- Mathematics Knowledge (MK),
- General Science (GS),
- Mechanical Comprehension (MC),
- Electronics Information (EI),
- Auto and Shop Information (AS), and
- Assembling Objects (AO).

This configuration of tests represents a recent change that added the AO test and removed two speeded tests that assessed coding speed (CS) and numerical operations (NO).

The ASVAB is administered as part of the Enlisted Testing Program (ETP) to ensure that applicants for military enlistment are qualified to serve and to assign recruits to entry level jobs

that best match their aptitude. Annually, more than 350,000 applicants for military service take the test. One composite of ASVAB tests, the Armed Forces Qualification Test (AFQT), is used to select applicants who are qualified for military service. Each of the services also calculates unique composites for job assignment that are especially suited to the types of jobs offered to enlisted personnel in that service.

As a part of the Student Testing Program (STP), the ASVAB is given to over 900,000 students in more than 14,000 high schools across the nation. The ASVAB, combined with other tests, allows students to understand their skills, interests, and values; to explore career options that may be appropriate for their interests and skills; and to develop a plan to realize their occupational goals. The career exploration process incorporated in the STP exposes students to the wide variety of military jobs that are available to them, as well as to civilian career opportunities. The STP is administered in the nation's high schools free of charge. The Department of Defense (DoD) gets, in return for the administration of the test, a list of qualified leads for recruiters.

The ASVAB is currently administered in two formats – as a traditional paper-and-pencil test and in a computer-adaptive format called the CAT-ASVAB. The paper-and-pencil version of the test takes about 3.5 hours to administer, and is the only test used for the STP. The CAT-ASVAB version takes about 1.5 hours to administer and is the primary version used to qualify applicants for military service.

Developing an internet version of the CAT-ASVAB has potential benefits for both the STP and the ETP. Allowing applicants to take the test via the web can reduce the number of computers and manpower required to administer the ASVAB at MEPS and MET sites. While it would be necessary to verify the scores obtained over the internet, the verification test can be shortened to less than 15 minutes, compared to the 90 minutes required for the stand-alone version of the CAT-ASVAB (Segall, 2001). Only applicants whose scores on the internet test were not verified would be required to retake the complete test.

Also, the iCAT could be used as a trial version of the test taken at home. The availability of such a version might have several beneficial effects on the recruitment process. Individuals who took the test would receive information regarding the likelihood that they would be qualified for military service. Those who were unlikely to be qualified might be discouraged from applying, thus reducing unnecessary recruiter effort focused on unqualified candidates. On the other hand, qualified individuals might be encouraged to apply, especially if they were qualified for desirable jobs. Consequently, the iCAT could provide a convenient, non-committal way for people to try out the ASVAB and could filter out those with very poor performance.

Internet administration of the CAT-ASVAB may have benefits to the STP, as well. Some schools may choose not to administer the ASVAB because of the length of time required; the number of schools in this category is unknown. An internet version of the CAT-ASVAB would provide a less time-consuming alternative to provide career exploration information to high school students. Such an opportunity might encourage schools to participate in the program. However, the benefits of iCAT would not be guaranteed, especially if individual students were allowed to participate in the STP without school sponsorship. Thus, the potential benefit of iCAT

for the STP would need to be verified through school surveys, focus groups, or other methods before internet administration of the STP is implemented.

It is important to consider the requirements for the prototype in the context of the eventual use of the iCAT. Consequently, we begin by describing the goals for the operational version of the iCAT and discussing some of the challenges that must be met to implement such a system. We then outline the specific goals for the prototype, as well as the potential uses for this more limited capability.

Goals for an Operational iCAT

The following three uses are envisioned for the iCAT:

1. The test could be used by a potential applicant for military service to practice the ASVAB and to become more familiar with the types of items in the test.

2. The test could be used for career exploration, either by individual students taking the test in their homes or by students taking the test in their schools as a part of the STP.

3. The test could be used to qualify an applicant for military service.

The iCAT would provide the ASVAB in a format that could be readily used by schools participating in the STP or by individuals in their homes or other locations. Delivering this test over the internet would allow individuals and participating schools access to the ASVAB using existing hardware and software.

The operational iCAT would have many of the features of the current stand-alone version of the test used in MEPS and MET sites and of the revised version being developed for Microsoft® Windows. Specifically, we envision the following characteristics for the operational version of the test.

- It would be administered in schools or in other testing facilities in a proctored environment to reduce the likelihood of cheating and to enhance test security.
- It would also be available to individuals who would take the test at their home computers.
- The time to administer the test would be about 1½ hours.
- Contact information for those who took the test would be provided as leads to recruiters.
- The test would include all ASVAB components.
- Scores would be available immediately to the test taker or school counselor.
- An applicant could use the scores obtained on the test to qualify for military service.
- To reduce the likelihood that an applicant for military service would cheat, test scores would be verified at the MEPS using a brief, adaptive test. Individuals whose scores were not verified would be given the entire CAT-ASVAB at the MEPS.
- Interactive career exploration would be available, either as a part of the test, or as a separate capability that the student could access at another time.

Implementation of this capability presents formidable challenges to the system developer. Probably the most difficult of these is the need to provide security for both test items and test scores. In addition, the process of implementing the adaptive testing method itself is by no means a trivial problem. A final consideration in the development of an internet CAT-ASVAB is the standards and regulations that control the content and methods included in DoD web sites. These regulations are primarily concerned with privacy and access to the disabled. They can limit the types of procedures that are used to implement an adaptive test.

Goals for the Prototype System

Although the prototype has a much more limited set of goals, it still represents a significant development challenge. The goal for the prototype is to develop an internet delivery system for the CAT-ASVAB that can be used by potential recruits to practice the CAT-ASVAB and get a reasonable indication of their chances of being accepted for military service, as well as other career information. The prototype will have the following characteristics.

- It will use the same adaptive methods as the stand-alone version of the CAT-ASVAB.
- It will not use operational items, and the scores obtained on the test will not qualify an applicant for enlistment.
- Similarly, it will not be used to generate recruiter leads.
- A student or applicant could use the test to get familiar with the types of items that are included in the ASVAB, or to determine whether he or she would likely be qualified for military service.
- Similarly, a recruiter could use the test to "pre-qualify" an applicant.

By removing the requirement for operational use of the test, the prototype avoids most of the security requirements of the operational test. However, the requirements to implement the adaptive test-item selection and to be consistent with relevant regulations and standards remain.

Design and Evaluation Issues

There are several technical hurdles that must be cleared to produce a system to deliver the CAT-ASVAB over the internet. In addition, the design must consider the likely capability of user computers and must be consistent with several relevant regulations. Martin and Hoshaw (1997) described nine criteria for the formal evaluation of CAT-ASVAB. These factors include performance, suitability, reliability, maintainability, ease of use, security, affordability, psychometric acceptability, and expansion/flexibility. McBride, Paddock, Wise, Strickland, and Waters (2001) have discussed issues regarding the use of the internet for the delivery of the ASVAB in terms of these factors.

This report focuses on three of the nine factors considered in earlier studies – performance, suitability, and security – because it is believed that these three factors present the most significant challenges to the development process. Some of the other factors, such as reliability and maintainability, are difficult to evaluate because much of the relevant equipment is not under the control of the test provider. Other factors, such as expansion and flexibility, were excluded because McBride et al. (2001) indicated that they do not present a problem to

development. We have taken a somewhat more general view of performance, and include technical issues in implementing the adaptive model and the psychometric acceptability of the resulting test within this factor.

This section discusses these three design and evaluation issues as they relate to both the prototype and the eventual implementation of the iCAT. For each issue, the discussion identifies the critical challenge, lists the assumptions that were made regarding the difficulty of the challenge, and describes design approaches used address the challenge. The section concludes with a description of regulations that apply to the iCAT.

Performance

The calculations required to select which test item to present can potentially place a load on the server that is controlling the test, especially if many are taking the test simultaneously. The extent of this load depends on the number of servers that are being used. In addition, transmission delays and limits of the user's internet connection can reduce some of the efficiencies that can be obtained through the use of adaptive testing. These factors jointly determine the expected response time of the system.

Response Timing

The design requirements for the stand-alone version of the CAT-ASVAB included a requirement for a maximum 2-second delay between the response to one item and the presentation of the next item. Although this goal may have been a challenge when the CAT-ASVAB was first developed, it does not present any problem to current capabilities. However, the iCAT must contend with transmission time and other delays that are out of control of the system developer. Meeting this goal for an internet version of the test may not be possible because of these additional sources of delay. However, internet users have become accustomed to these delays, and we do not believe that failure to meet a 2-second requirement would indicate a failure of the system, as the following discussion indicates.

We anticipate that transmission delays will have a small effect on overall test time. For example, a 2-second transmission delay per item would add approximately 5 minutes, or roughly 6%, to the total test time (assuming 150 items take approximately 90 minutes in the stand-alone version). In general, transmission time is out of the control of the system developer, but reducing the amount of material to be transmitted can minimize it. Use of large graphics or multimedia files can greatly increase the required transmission time. In this light, the requirements of the CAT-ASVAB are modest because they consist primarily of text and line drawings, which produce relatively small files. Transmission delays can be caused by limitations in the speed of the user's internet connection or bottlenecks somewhere in the path between the server and the user. They can also be caused at the server by the computational load required to process many simultaneous users.

Tradeoffs between Response Time, Computational Load, and Bandwidth

To some extent, it is possible to improve response time, at the expense of computational load and required bandwidth. A procedure that loads items before they are needed could reduce the effects of transmission time. For example, while an individual is reading and answering one question, the system could load the two possible items that would appear next, depending on whether or not the answer was correct. Either of these items would be available immediately, when the individual answered the question. Although such a method could improve response time, its effects on other computational and bandwidth requirements should also be considered. These effects of this procedure are considerable, and include the following items.

- It roughly doubles the calculations required to select items, because two alternative items are chosen at each step.
- It doubles the required bandwidth at the server, because twice as much information is being transmitted to the test taker.
- It requires substantial client-side programming.
- It may have negative impacts on security of correct responses, item parameters, or adaptive item-selection procedures.

The substantial cost of these procedures to reduce response time suggests that they be considered only if delays present a substantial risk to the validity of the resulting scores. We anticipate that DMDC will conduct tests to estimate average response times with a realistic range of network transmission times. We anticipate that the obtained response times will be reasonable, although they may not meet the original CAT-ASVAB standard under all conditions. With a sufficient number of servers supporting the iCAT, a reasonable (e.g., less than 5 sec.) response time should be feasible, even with several hundred simultaneous users per server. Additional tests will determine the effects of interface differences (including differences in response time) on the validity of obtained test scores. While we do not anticipate large effects of these factors, they might require design changes or additional hardware, if they occur.

Suitability

An important concern regarding the suitability of internet administration of the CAT-ASVAB is whether the potential test takers have access to appropriate equipment and internet connections at their homes and schools. In general, recent data indicate that both schools and homes typically have appropriate equipment and access to make internet administration of the CAT-ASVAB feasible.

The National Center for Educational Statistics (Kleiner & Farris, 2002) published data that indicate the extent to which internet access in schools has grown since 1994. At that time, 49% of public secondary schools had internet access, but access was only provided to 4% of instructional rooms. By 2001, the picture had changed substantially. Over 99.5% of public secondary schools had internet access, and access was provided in 88% of instructional rooms. With an average ratio of 4.3 students for each instructional computer with internet access, it seems that most public secondary schools have sufficient technology available to support the use of the iCAT in the STP.

However, the data reported by Kleiner and Farris (2002) also indicated that internet capabilities are not evenly distributed among schools with different characteristics. Most notably, schools in areas with high concentrations of poverty had less access to the internet. Only 79% of instructional rooms had internet access in schools in which 75% or more of the students were eligible for free or reduced-price lunch. In addition, the percentage of instructional rooms with internet access was lower in cities (82%) than it is in urban fringe areas (87%), towns (91%), or rural areas (89%). Although the data from 2001 indicate differences in internet access related to school characteristics, they were much reduced compared to previous years.

An analysis by Parsad, Skinner, & Farris (2001) indicated that development of internet access in private schools lags somewhat behind their public counterparts, although, with some exceptions, rapid progress is being made to incorporate this technology in private educational programs. In 1998-1999 school year, 90% of private secondary schools had internet access, compared to 98% of public secondary schools (in 1999, from Kleiner & Farris, 2002). The difference in the percentage of instructional rooms with internet access is more striking, with 32% of instructional rooms having access in private secondary schools, compared to 67% in public schools. Many of the schools without internet access were planning to obtain connections by the end of 2000. However, 19% of private schools, combined over instructional levels, reported having no plans to be connected to the internet.

These studies also indicate that schools are increasingly using high-speed internet connections. In 2001, 85% of public schools used some kind of broadband internet connection, such as a T1 line, cable modem, or digital subscriber line (DSL; Kleiner & Farris, 2002). These data also indicate that high-speed connections are more common in larger schools, schools with larger proportion of minority students, and schools with high concentration of poverty. Comparable data for private schools were not available. Overall, the speed with which schools are developing internet connections indicates that any problems with access should be minimal by the time an operational version of the iCAT is implemented.

Computers with internet access are less likely in homes than in schools. According to the U.S. Census Bureau (2002), in 2001 about 51% of households had internet access. Among married couples with children under 18, this figure increased to 72%. Access varied substantially by household characteristics, including age, race/ethnicity, education, and income. Because home internet access has nearly doubled between 1998 and 2001 (U.S. Census Bureau, 2002), any problem in access should be substantially reduced in the next few years. Furthermore, it should be noted that students often have internet access at their school or at public libraries even if they lack such access at home. Consequently, internet access within the student population should not be a major limiting factor on the suitability of iCAT for home administration of the ASVAB.

The speed of internet connections might present a barrier to home use to some extent and might limit the extent to which more advanced displays can be incorporated into the iCAT in the future. Two recent studies by Nielsen/NetRatings (2003) and by the UCLA Center for Communications Policy (2003) present similar pictures of the type of internet connections that are available in American homes. At the end of 2002, between two-thirds and three-quarters of

homes that had internet connections used narrowband, dialup connections, while the remaining homes had broadband connections (primarily cable modem and DSL. However, broadband connections have shown tremendous growth, at a rate estimated to be 59% in 2002 (Nielsen/NetRatings, 2003). If rapid growth in the use of broadband connections continues, it seems likely that there will be minimal barriers to home use of the iCAT within the next few years.

An additional suitability consideration regards browser requirements. However, the iCAT prototype places minimal requirements on browser capability, and is compatible with all major browsers including the current versions of browsers from Microsoft and Netscape.

Security

Implementation of a test on a computer can provide the test taker with ways to cheat that are not possible with paper-and-pencil tests. Furthermore, electronic transmission and storage of test items may reduce their security. Although there are technologies available to reduce security risks, trained proctors and test-taking procedures probably provide the best assurance of security. If the test is given in an unproctored environment, then it is safe to assume that both the test items and the test scores may be compromised. Consequently, scores obtained from the internet CAT-ASVAB will need to be verified, if they are used to qualify an applicant for military service.

If the internet test is taken in an unproctored environment, then the security of items cannot be assured. Both the prototype and operational system requirements envision use of the test by individuals at their home computers. Consequently, it must be assumed that the items are compromised. The prototype system will use demonstration test items for which parameters of the item response theory (IRT) model have been estimated. Current plans for the operational system are that the system will use items that have been retired from a current form of the stand-alone version of the CAT-ASVAB.

It is vital that the short verification test, currently being developed by DMDC to be given at MEPS/METS, catches most of the cheaters. Algorithms for estimating likely amount of cheating and setting acceptance thresholds are being developed, tested, and measured. The fact that the scores from the iCAT will be verified at MEPS/METS must be advertised to web users to discourage cheating. Otherwise too many people will fail the verification test and then be forced to retake the full ASVAB at the MEPS/METS and that will reduce the time savings achieved by pre-testing via web. However, because the verification test is significantly shorter than the operational test, some time savings can be obtained even if the iCAT scores are not verified for a substantial proportion of the applicant population.

Relevant Regulations

Several requirements apply to an operational implementation of the iCAT. Only some of them apply to the prototype, because of its more limited uses. We describe three relevant requirements: the Privacy Act, DoD design requirements, and the requirements of Section 508 of the Rehabilitation Act.

Privacy Act

Operational use of the iCAT for either career exploration or application for military service will require the collection of personal information. Currently, one of the outcomes of the STP is a set of recruiting leads that provide contact information for participants in the program. Recruiters can use these leads and the associated scores to identify students who would qualify for military service. The collection of recruiting information has two implications for the requirements of the operational internet CAT-ASVAB. First, since recruiting leads require the collection of personally identifying information the requirements of the Privacy Act will apply. A statement or advisory that describes the uses of the identifying information must be given to the test taker. Furthermore, the data must be secured to restrict access to authorized individuals. Second, the ability of an individual to take the CAT-ASVAB at home also implies that recruiting leads might be obtained from students without agreements with their high schools.

The prototype system will not generate recruiting leads, and consequently will not need to collect any identifying information covered by the Privacy Act. It may be useful for the prototype to include a login procedure, so that an individual can resume an interrupted session or view test results after taking the test. However, the login id and password used could be completely arbitrary, and need not contain identifying information. Currently the prototype collects the user SSN to be consistent with the DOS CAT ASVAB. However, for testing of the prototype, we recommend that this feature be disabled, so that the Privacy Act no longer applies. This change would be feasible without great difficulty.

DoD Design Requirements

DoD has developed policies and procedures that govern the administration of military web sites (Office of the Assistant Secretary of Defense [Command, Control, Communications, and Intelligence], 2001). The requirements that are relevant to this system are enumerated and briefly discussed below.

- A privacy and security notice must be displayed or announced on the home page and the first page of all major sections of the site. The policy and procedures manual provides the required text to be incorporated into the notice.

- IDs and passwords should not be transmitted without encryption using a secure protocol, such as secure sockets layer (SSL).

- The site should not require or encourage users to choose any specific browser software. To the extent possible, it should be compatible with all major browsers.

- Links to software download sites should incorporate text only and should not incorporate any graphics or company logos. For example, a text link to download a reader for Acrobat files should not include a graphic of the logo for the Adobe Corporation.

- In general, the requirements of the paperwork reduction act apply to DoD web sites. This requirement covers user feedback, if the same information is sought from more than nine individuals. However, the ASVAB is exempted from the requirements of the paperwork reduction act. Consequently, it will not be necessary to get OMB clearance to obtain either test data or user feedback. In addition, OMB clearance is not required for attempts to obtain general user comments, since the information obtained from each individual is unique.

- Whenever personally-identifying information is obtained from users, the requirements of the privacy act apply. When the information is maintained in a Privacy Act system of records, a Privacy Act Statement must be posted on the web page. If the information is not maintained in a Privacy Act system of records, a Privacy Advisory is required.

- Use of temporary session cookies is allowed, but must be identified in the site's privacy notice. Temporary cookies are created when the new browser instance is created and cease to exist when the browser is closed.

- Use of persistent cookies is not allowed unless several restrictive conditions are met, including the personal approval by the Secretary of Defense. Other automated means of collecting personally-identifying information are also governed by this requirement.

- Using log files to get information such as the browser type is permitted as long as it is disclosed in the privacy and security notice.

The prototype currently conforms to DoD design requirements.

Section 508 Compliance

Section 508 of the Rehabilitation Act requires any electronic and information based technology products (e.g., web sites, software) created for or by a Federal agency to be made accessible to people with visual, hearing, and/or motor disabilities. The enforcement of Section 508 began June 21, 2001. Because the test is used in the STP as a basis for career counseling, some people will want to ensure that all students are able to benefit from this service, regardless of any disability.

Relevant requirements. Several requirements of Section 508 are particularly relevant to the test. In fact, it may not be possible to satisfy the first of these requirements without jeopardizing the validity of the test results. We present a brief discussion of four particularly relevant requirements.

First, Section 508 mandates that a text equivalent for every non-text element providing information essential to the web page's content or navigation be provided. This requirement affects test items that include graphical information. It specifies that a text tag be associated with the graphic that explains what it is. This requirement cannot be met for some items without substantially changing them and making them useless for testing purposes. For example, an item may include a picture of a saw with the question, "What is this tool?" Section 508 would require

a text tag that describes the content on the graphic. Obviously, the tag, “picture of a saw,” would provide the answer to all test takers, whether or not they were disabled. Rather than providing the kinds of tags that are required by Section 508, we recommend that the test accommodate disabled users by providing alternate assessment methods for those tests that use graphical information.

Second, Section 508 requires that the test (essentially an HTML form) be designed so that test-takers using assistive technology can access the information, input field elements, and perform all functions needed to complete and submit the test. The following techniques can be used to meet this requirement

- Use of explicit field labels ("label" tag, "for" and "id" attributes).
- Location of field labels in proximity and to the left of each field, linking the two when read by a screen reader.
- Use of the "tabindex" attribute to create a logical tab order to the input fields.
- Avoidance of drop-down lists that require the use of the mouse.

Third, Section 508 requires that a disabled user be alerted about timed responses and allowed additional time to give a response if needed. Notification of time limits and of the time remaining on a particular test is included in the current version of the CAT-ASVAB, and hence will be incorporated into both the prototype and the operational internet versions. Provision of additional time will need to be implemented in a fashion such that only the disabled individual is able to obtain additional time. Otherwise, the test results will not accurately represent the abilities of those who take the test.

Finally, navigation and access to various input fields should be possible by using either a mouse or the keyboard ("keyboard enabled" web pages) to accommodate users with mobility impairments.

Prohibited techniques. Several web-programming techniques are not compliant with Section 508, including the following:

- Frames,
- Animated graphics files
- Java applets,
- Third-party plug-ins or applications
- Audio, unless accompanied by a text transcript.
- Video

We currently use none of the techniques prohibited by Section 508.

Restricted techniques. Other techniques are to be avoided or are restricted in some way.

- Color should not be used as the sole distinction between elements on the screen or as the sole method for conveying information.

- If cascading style sheets (CSS) are used, they should be external style sheets that will not interfere with user-defined style sheets used by some users to define viewing preferences that accommodate their particular disability. Also, the web page should be designed to display and function properly without style sheets.

- Server-side image maps should not be used. Client-side image maps should incorporate alt tags to identify the destination of links.

- If data tables are needed, row and column headers should be used so that screen readers can associate a table cell with the corresponding row and column headers. This can be accomplished using the "scope="col"" and "scope="row"" attributes in the "th" or "td" tags.

- Section 508 standards require website authors to avoid the use of tables for layout unless the content is clear when linearized (i.e., read line by line).

- JavaScript event handlers that respond to mouse actions, such as "onMouseOver" and "onMouseDown" (e.g., for rollovers), should be duplicated by keyboard equivalents, such as “onFocus” and “onKeyDown.”

- Test-takers using assistive technology should be able to skip repetitive navigation links.

- We do not anticipate using pop-up windows. However, if they are used, instructions on how to return to the main page should be made available to those using assistive technology.

The procedures used in the prototype currently are in compliance with the requirements in this area, as well. The user can navigate using the keyboard simply by using the Tab key to advance to the desired location, and then pressing Enter. We may at a later date add JavaScript that allows a selection of the multiple-choice answer by choosing the appropriate letter, but that feature will be redundant with existing entry methods. Consequently, with the exception of alternate representation of the graphics used in test items, we believe that the prototype is in compliance with the requirements of Section 508.

System Description

Programming Environment

In establishing the programming environment, we didn't want a solution that only ran on windows because DMDC does not use Windows machines for hosting websites, they use UNIX machines. Also, most computers used by MEPCOM operate under UNIX. However, we didn't want to restrict the environment to UNIX, because we wanted the freedom to develop on either Windows or UNIX. Finally, we wanted the flexibility to be compatible with new operating environments, should DMDC make changes in the future.

The current operational environment includes an Oracle database with Java 2 – Enterprise Edition (J2EE) programming environment and JDBC for Oracle. The development environment mimics the operational environment on a single workstation. It is based on the TOMCAT implementation of the J2EE standard with a Java Development Kit (JDK) from Sun. Java was chosen because it is platform independent, has good performance, can share data among sessions, is not vendor-specific, and is easily integrated with Oracle. The TOMCAT implementation of J2EE was chosen because it is free and has served as the reference standard for J2EE. The database uses My SQL with JDBC. My SQL is a freeware database that provides many of the capabilities of Oracle and can be integrated with J2EE. Given the similarity between the development and operational environments, it should be a straightforward task to move the prototype to the target system. Furthermore, if the requirements or platform change in the future, we will be able to deploy the iCAT using J2EE with any relational SQL database.

We chose to implement graphics in PNG, rather than GIF, to avoid any patent problems with the GIF standard. Graphics files using the PNG standard are readable by all major browsers.

Implementation of Adaptive Item Selection

Algorithms were taken from the DOS CATASVAB C-source code and converted to Java with as little change as possible, to minimize the impact of the conversion and help decrease time spent testing. In some cases, modification was required because of differences in the data structures between the two systems. These modifications were limited wherever possible to the input parameters and the output stages, leaving the heart of the algorithm essentially unchanged.

There were several differences in the data types available between the 16-bit DOS C-code used in the stand-alone demonstration and the Java code used for the iCAT, but most of the time there was enough overlap to avoid significant changes. Some subtle differences in numerical precision may have been introduced, but largely we were able to keep the same data type. Where we could not use the same data type for a particular variable, we maintained or increased the precision of the variable.

Another consideration regarding precision is that the item data were converted from the DOS demonstration to the iCAT prototype in several steps. First, data were converted from the proprietary internal Borland C data structure for floating-point variables to text for import into the SQL database. These text variables get loaded into the Java programs at runtime. The conversion can theoretically produce small differences in values, but no large discrepancies are anticipated from this process.

Final detailed testing of the procedures used in the prototype will require substantial testing by quality control statistical specialists at DMDC. Because the item selection methods are nonlinear, it is possible for small discrepancies in input data to produce substantial effects on the items that are selected and the resulting test score. An assessment of the extent of these differences will need to be conducted before the operational version of the iCAT is implemented.

There are many tables in the CAT-ASVAB that implement different functions required in the system. These were all converted to SQL table data and Java structures/objects were created

to load and hold these in memory on the server at run-time, according to whether they are globally shared data, per-user session data, per-subtest data, or per-item-data. The item-selection algorithm for choosing the next question adaptively was implemented according to CAT-ASVAB methods and the Bayesian algorithm for estimating ability was developed, along with the Owens-Bayes algorithm used during the subtest. This algorithm was implemented with slightly lower accuracy than the DOS implementation, but in a way placing much less demand on computational resources.

The item-exposure control method was implemented to ensure that items are not over-exposed, especially those near the beginning of the test that have "average" difficulty. Item-exposure control makes it harder to cheat, since some questions are passed over randomly, controlled by an exposure parameter. Seeded items were not implemented in the prototype. It is currently believed that this feature will not be used in the iCAT. If it is required at a later date, it will be possible to extend the software to support it.

Certain subtests place special requirements on the prototype. For example, scores for the Automotive Information (AI) and Shop Information (SI) subtests are given separately in the CAT versions of the ASVAB, but get a combined score (AS) to match the paper-and-pencil test. This difference between adaptive and paper-and-pencil versions is needed because there is an extra demand for the content of adaptive tests to be "unidimensional," so that the score being developed relates to a single subject area. Other tests, namely GS and AO used a load-balancing procedure to ensure that the content covered relevant sub-areas. Load balancing is a somewhat less extreme approach than the outright separation that was used for AS. Both of these procedures were implemented in the iCAT prototype.

Scoring is conducted in several steps. First, the raw score is determined. Then penalties for timed-out subtests are added if the test taker had answered only a small number of questions. Next, the scores are normalized and converted to be consistent with their paper-and-pencil equivalents. Composite scores of general ability (AFQT) and other composites for various branches of military are given and converted using various equivalence algorithms or tables. Finally, the full report is shown to the test taker. A potential future modification to the scoring procedure is to run the iCAT in a restricted mode that doesn't show the full report at the end.

Prototype Operation

Both the prototype and operational systems will require the presentation of introductory material, user registration, presentation of test items under control of an adaptive algorithm, calculation and reporting of test results, and collection of user feedback. The operation of the iCAT prototype is very similar to that of the DOS demonstration version on which is it based. It begins with a simple registration procedure in which the test taker provides his or her SSN. To avoid Privacy Act requirements, identifying information should not obtained when the prototype is evaluated.

After registration, the user is presented with instructions on test-taking procedures. Unlike the DOS version, the iCAT prototype does not require a special keyboard. Consequently, that aspect of the instructions is not required. Tests are then presented in a fixed order. The tests

correspond to the current operation version of the ASVAB. However, we have eliminated the two speeded tests that were included in the demonstration (NO and CS), and have added AO.

The prototype provides feedback regarding test scores, including general ability (AFQT) and specific composites for each military service. The report format is based on the comparable format for the demonstration. As was stated previously, a potential future modification to the scoring procedure would allow for restricted score reporting.

Feedback for the prototype will probably most closely mimic the information provided to the applicant. This feedback would be patterned on the information currently provided by the stand-alone version of the CAT-ASVAB. However, since the test uses demonstration items and cannot be used for the purpose of qualification, it will be important to give the user a warning that the scores are not official and only approximate the scores on the operational test.

Collection of User Feedback

One of goals of developing a prototype assessment is to utilize its administration to elicit feedback from test subjects on characteristics that might influence its psychometric properties and/or attitudes towards the test. Participant feedback can also be used to help determine if there are any unforeseen systematic problems with the test. In addition, in the context of internet-based assessment, feedback can be gathered to determine the system specifications of the test taker's system in order to track differential performance for the prototype on systems with varying capabilities (i.e., the performance of the test on a system with a 550 MHz microprocessor as compared to performance on a system with a 2.0 GHz microprocessor). Finally, questions can also be targeted to the performance of the test delivery software and hardware during the user's test session.

The following questions would characterize the computer capabilities of the user population and provide useful information for evaluation of the prototype system. While some of this information could be obtained automatically using client JavaScript programs, most would need to be assessed directly from the user.

System Performance Questions:

- What are your system specifications?
- What type of internet browser are you using?
- The software for the testing program functioned properly? (i.e., buttons included in the graphic environment worked properly)
- Did you receive any error messages? (If so, what were they?)
- On average for all the items, how long did you have to wait between answering one item, and the presentation of the next?
- Did the delay in the presentation of items take a similar amount of time, or did some items take significantly longer to display than others?
- In your judgement, what was the <u>longest</u> amount of time you had to wait for any one item to be displayed?

- In your judgement, what was the shortest amount of time you had to wait for any one item to be displayed?
- Did you have any trouble logging in? (i.e., inputting your Username and/or Password)
- If you used the help function, did the help function perform properly? (When you requested help, you received the relevant information.)

Testing Experience:

- Were the test items visually easy to read off the screen?
- Was any part of the testing session confusing? If so, what part of it?
- Overall, was the testing experience a positive one?

Open-Ended Question:

- What other comments do your have about this test?

Operation of the Operational iCAT

The operational version of the internet CAT-ASVAB will serve three purposes: (1) provide practice for an operational test, (2) serve as a component of a career exploration system, and (3) determine qualification for military service. The operational version will need to be organized so that these three functions will be clear to the user. Consequently, it will require functionality that is not present in the prototype.

The operational version of the iCAT should be as similar as possible to the version that will be in use at the MEPS for two reasons: (1) to maximize the validity of test results and (2) to maximize the utility of practice on the system in preparing an individual for the operational ASVAB. Because a parallel effort is being conducted to develop a stand-along version of the CAT-ASVAB that operates under Microsoft® Windows, it should be possible to make the internet and stand-alone versions essentially identical. Neither of these versions should be arbitrarily constrained to share the features of the DOS version. Of course, the change to a more modern user interface will require additional studies to ensure the validity of test results.

One advantage of computerized testing for result reporting is the ability to develop and display differing levels of data quickly to different groups as defined by their need for the data. In other words, individuals will not receive data that is not relevant to them. For example, the needs of individuals who are applying for military service are different from those who are using the ASVAB for career exploration. All users may be interested in their standard scores on individual tests. However, applicants for military service will need to know their AFQT percentile, as well as their composite scores for the service to which they are applying. Those who are conducting career exploration will need more general information, such as general estimates of verbal, math, and science and technical skills. The operational version of the internet CAT-ASVAB will need to support both types of score reports.

Internet-based centralization of data should resolve issues relating to the current CAT-ASVAB system where results of testing at the MEPS are sent only once a day to DMDC causing

a one-day lag for access to results by upper level decision makers. Centralization of data brought about through internet administration has the potential to eliminate this time lag and increase system efficiency. Centralization of data also facilitates the use of a "single access point" for the system. This single access point can not only generate reports but can deliver other services as well. For example, any individual with internet access (and the proper authorization) may access the system to check results and also access tutorials to assist in the interpretation of those results.

Other Implementation Issues

We conducted an initial load testing to estimate the number of users that could be supported by a single server. The results of the test indicated that a 1GHz machine will be able to simultaneously serve 300-500 people. This level of performance exceeds our goal for the prototype, which was for one machine to be able to serve at least 200 per machine.

Sessions may be interrupted for a variety of reasons, including as power outage, accidental browser close, emergencies requiring the user to abandon the test (temporarily), and so forth. We have developed procedures to allow the user to resume the test after such an interruption. When the server sees no activity from the user after a timeout period (20 min. is the default), then the Java thread monitoring sessions will see the inactivity and proceed to save all user-session state information to the database. When the user logs in again, the entire state will be restored so that the user can resume just where he or she left off. A user who accidentally closes the browser will have to wait 20 minutes for the system to timeout the session before he or she can log in again. A message will tell the user to try again later (in about 20 min.). We expect that this event will happen only rarely. This approach is a simple and reliable way to deal with timed-out sessions. Alternative responses to an attempted re-logon, such as ending and restarting the existing session or taking it over seemed both inadvisable and unnecessary. With the current procedure, sessions can be interrupted and restored safely as outlined above.

Once users have completed a subtest or question, they cannot go back, even if their sessions are interrupted and resumed. The prototype includes a log of session interruptions.

Discussion

The iCAT prototype has established the feasibility of conducting adaptive testing over the internet. It has shown that current adaptive methods can be implemented to support hundreds of test takers simultaneously with a single server. Larger numbers of users can be supported with multiple servers. However, there are several activities that remain to be accomplished before an operational iCAT can be implemented. In addition, the prototype development effort did not resolve several issues, most notably those related to security.

Hosting of the iCAT prototype is a logical next step in the development process. Although this activity does not require any fundamental technical challenge, it does require procurement of server hardware, administrative review of the prototype, and considerable system support. A substantial part of the evaluation requires that the system is hosted and available to internet users.

System testing will be required to ensure that scoring and item selection procedures operate accurately, and that resulting scores correspond to scores from other validated versions of the ASVAB. This phase of testing will require a combination of simulation and empirical studies. Most of the data needed for these studies is already in the database, but there may be other variables required. These items might be added to tables or incorporated into logging/debugging files that could be enabled for test verifications. Although some preliminary evaluation can be conducted using the demonstration items in the prototype, a complete evaluation will require operational items.

The prototype currently does not use any client-side programming. Some JavaScript may be unavoidable in the production version for implementing feature such as hot-key answer-picking or a client-side timeout counter. In addition, JavaScript can be used to record such features as screen size or browser type. This information could be used to track the capabilities of user computers and to evaluate the effects of these factors on the psychometric properties of the test scores. It should be noted that DoD Design Requirements restrict the information that can be obtained and specify that the user be notified that the information is being collected.

References

Kleiner, A., & Farris, E. (2002). *Internet access in U.S. public schools and classrooms: 1994-2001* (NCES 2002-018). Washington, DC: National Center for Educational Statistics.

Martin, C.J. & Hoshaw, C.R. (1997). Policy and program management perspective. In W.A. Sands, B.K. Waters, & J.R. McBride (Eds.), *Computer adaptive testing: From inquiry to operation.* Washington, DC: American Psychological Association.

McBride, J.R., Paddock, A.F., Wise, L.L., Strickland, W.J., & Waters, B.K. (2001). *Testing via the internet: A literature review and analysis of issues for Department of Defense internet testing of the Armed Services Vocational Aptitude Battery (ASVAB)* (FR-01-12). Alexandria, VA: Human Resources Research Organization.

Nielsen/NetRatings (January 15, 2003). *Broadband access grows 59 percent, while narrowband use declines, according to Nielsen/NetRatings.* Press release.

Office of the Assistant Secretary of Defense (Command, Control, Communications, and Intelligence) (2001). *Web Site Administration Policies & Procedures*. Washington DC: Author.

Parsad, B., Skinner, R., & Farris, E. (2001). *Advanced telecommunications in U.S. private schools: 1998-99* (NCES 2001-037). Washington, DC: National Center for Educational Statistics.

Segall, D.O. (April, 2001). *Measuring test compromise in high-stakes computerized adaptive testing: A Bayesian strategy for surrogate test-taker detection.* Paper presented at a meeting of the National Council on Measurement in Education, Seattle, WA.

UCLA Center for Communication Policy (2003). *The UCLA internet report – surveying the digital future: Year three*. Los Angeles, CA: Author.

U.S. Census Bureau (2002). *Statistical abstract of the United States.* Washington, DC: Author.

Appendix

The following list presents a list of the directories included in the system as delivered on CDROM, and a description of the files included in each directory.

Directories on the CDROM:

/jakarta-tomcat-3.3/doc/appdev
Main Java servlet source for iCAT.

/mm.mysql-2.0.7
Driver to connect to MySQL.

/mysql
Has MySQL itself.

/rebol
Contains data conversion functions written in REBOL
for taking demo diskette data and making
the actual SQL import files to be run by MySQL
thus populating the database.

/pcx-png
Has a REBOL utility for converting pcx to png.

/screenshots
Has jpg images of the DOS CATASVAB screens
used for designing the iCAT.

/IrwinHom
Contains dos CATASVAB source obtained from Irwin Hom

/DemoDiskette
Contains snapshots of the DOS CATASVAB
demo diskette and related material.

/dan segall
Contains some files Dan made up and
some missing statistics tables for the AO group
that we put together (AO not included on orig. demo diskette).