Study
Note
99-02

CATBOOK
Computerized Adaptive Testing: From Inquiry to Operation

Edited by

W. A . Sands
Chesapeake Research Applications

Brian K. Waters and James R. McBride
Human Resources Research Organization

United States Army Research Institute for the Behavioral and Social Sciences

January 1999

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical Review by

Ronald B. Tiggle

NOTICES

DISTRIBUTION: Primary distribution of this Study Note has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Study Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research for the Behavioral and Social Sciences.

NOTE: The findings in this Study Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1999	Final April 1994 - October 1996

4. TITLE AND SUBTITLE

CATBOOK - Computerized Adaptive Testing: From inquiry to operation

5. FUNDING NUMBERS

MDA903-93-D-0032, DO 0017
1331C28
65803A
D730

6. AUTHOR(S)

W.A. Sands, Brian K. Waters, and James R. McBride (Eds.)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 400
Alexandria, VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER

FR-EADD-96-26

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: TAPC-ARI-RP
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

Study Note 99-02

11. SUPPLEMENTARY NOTES

Contracting Officer's Representative: Ronald B. Tiggle

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

HumRRO contracted with ARI, sponsored by OASD/P&R (AP), to produce a book for commercial publication by the American Psychological Association (APA) which documents the research and development of computerized adaptive testing (CAT) as a means of administering the Armed Services Vocational Aptitude Battery (ASVAB), the personnel selection test battery used by the Department of Defense (DoD). The CAT-ASVAB program began in 1979, and bore operational fruit in 1992, when CAT-ASVAB went into limited use in an operational test and evaluation. CAT-ASVAB has since been approved to replace conventional, printed versions of ASVAB, beginning in 1996 in all Military Entrance Processing Stations (MEPs).

The principal objective of this book is to document the psychometric research and development of the CAt-ASVAB program and the important practical lesssons learned in developing its delivery system. The approach does this in a historical context. A secondary objective of the book is to provide a case study of the entire CAT-ASVAB program. The book primarily addresses three aspects of CAT-ASVAB history in DoD (adaptive testing measures and strategies; CAT-ASVAB system design issues; and CAT-ASVAB evaluation). It provides reference information useful to practitioners developing a computerized testing system.

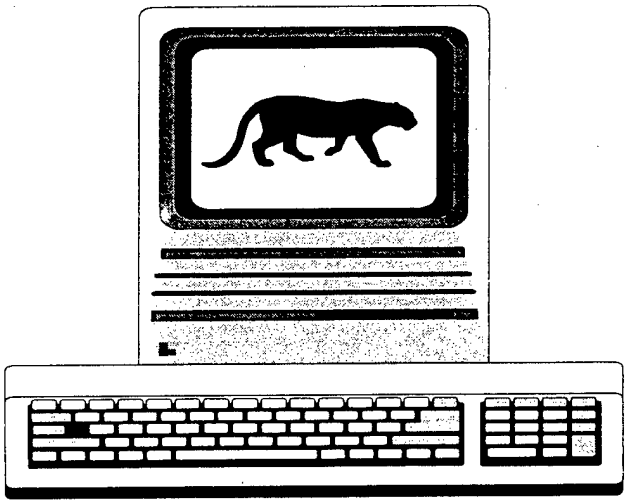
Publication by APA will occur in early 1997.
Reference:
Sands, W.A., Waters, B.K. & McBride, J.R. (1996) CATBOOK - Computerized adaptive testing: From inquiry to operation (HumRRO FR-EADD-96-26). Alexandria, VA: Army Reserach Institute for the Behavioral and Social Sciences.

14. SUBJECT TERMS	15. NUMBER OF PAGES
ASVAB CAT Computerized testing	
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

CATBOOK

Computerized Adaptive Testing: From Inquiry To Operation



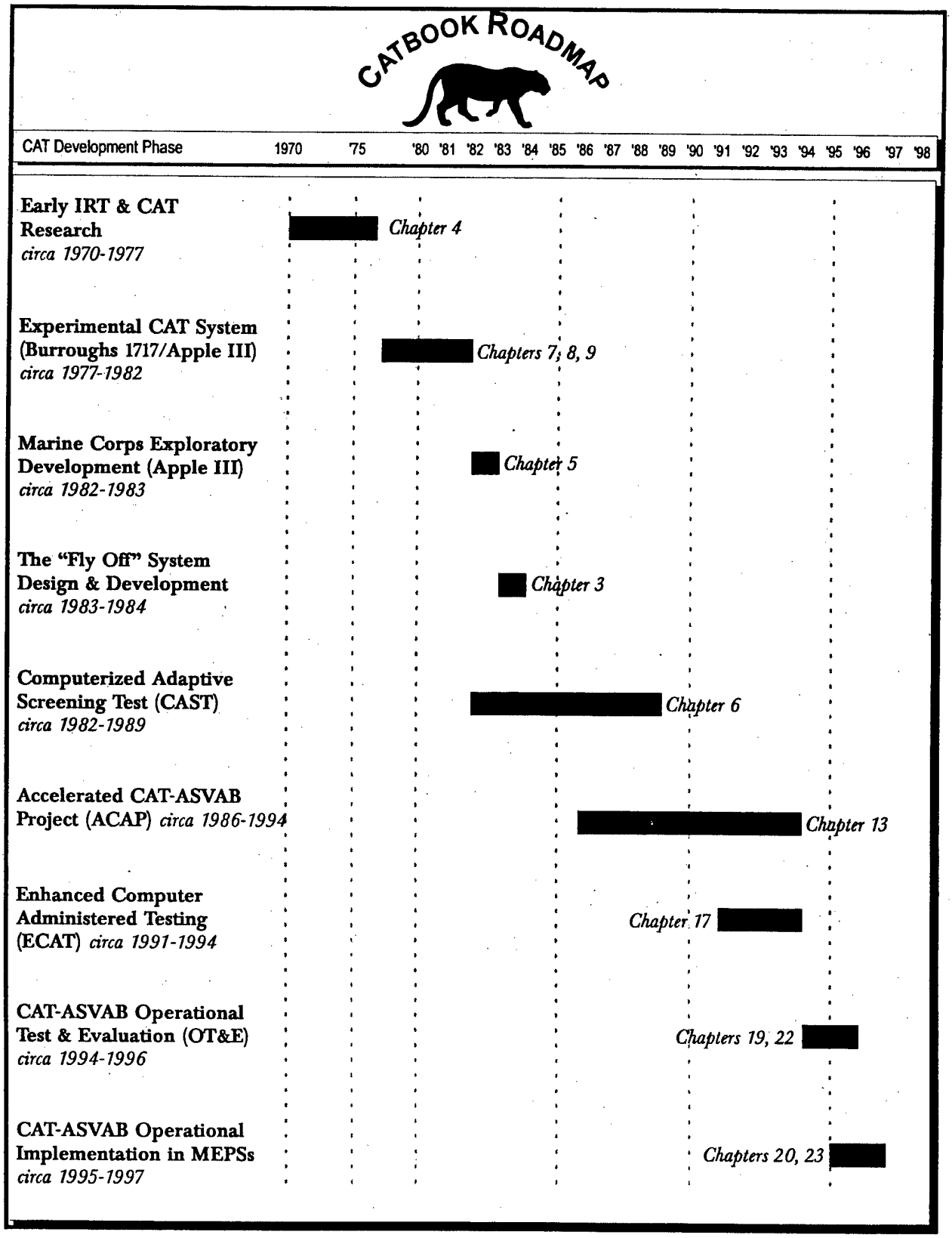
Edited by

W.A. Sands,[a] *B. K. Waters, and J. R. McBride*

Human Resources Research Organization

October 1996

[a] Chesapeake Research Applications (Consultant to HumRRO)


CATBOOK ROADMAP
CAT Development Phase
1970
'75
'80
'81
'82
'83
'84
'85
'86
'87
'88
'89
'90
'91
'92
'93
'94
'95
'96
'97
'98
Early IRT & CAT Research
circa 1970-1977
Chapter 4
Experimental CAT System (Burroughs 1717/Apple III)
circa 1977-1982
Chapters 7, 8, 9
Marine Corps Exploratory Development (Apple III)
circa 1982-1983
Chapter 5
The "Fly Off" System Design & Development
circa 1983-1984
Chapter 3
Computerized Adaptive Screening Test (CAST)
circa 1982-1989
Chapter 6
Accelerated CAT-ASVAB Project (ACAP) circa 1986-1994
Chapter 13
Enhanced Computer Administered Testing (ECAT) circa 1991-1994
Chapter 17
CAT-ASVAB Operational Test & Evaluation (OT&E)
circa 1994-1996
Chapters 19, 22
CAT-ASVAB Operational Implementation in MEPSs
circa 1995-1997
Chapters 20, 23

PREFACE

This book incorporates the ideas and work of many dedicated people, from a variety of professional disciplines, who have made significant contributions to the Computerized Adaptive Testing - Armed Services Vocational Aptitude Battery (CAT-ASVAB) Program from inception in 1979 to the present. A review of the Table of Contents illustrates the large number of authors involved in writing chapters for this book. Numerous other individuals, both inside and outside of the Navy Personnel Research and Development Center (NPRDC), made important contributions over the years. However, four individuals should be singled out for special recognition, based upon the critical roles they played in the success of the CAT-ASVAB Program.

Dr. W. S. Sellman, Director for Accession Policy, Office of the Assistant Secretary of Defense (Force Management Policy) provided vision, on-going guidance, and support for the program from the beginnning until the present. The CAT-ASVAB Program developed as a Joint-Service program, with each Service playing a role, and having its own perspective. Dr. Sellman's central, Department of Defense (DoD) perspective has kept the CAT-ASVAB Program focused on the eventual goal of full-scale, nationwide, DoD implementation of a scientifically sound and practical testing innovation.

Dr. M. F. Wiskoff created the computerized adaptive testing research capability at NPRDC, where the vast majority of the research and development for CAT-ASVAB has been accomplished. He convinced NPRDC management of the merits of the CAT concept, created the organizational structure for the program within his Manpower and Personnel Laboratory, hired new professionals from outside the Center and reassigned key personnel assets from other areas within his laboratory. As the first Officer-in-Charge of the Joint-Service CAT-ASVAB Program, he chaired the CAT-ASVAB Working Group, and headed the CAT-ASVAB Program Office, which included a uniformed officer from each of the Services. His contributions to CAT-ASVAB were crucial to the Program's birth and growth.

Mr. C. R. Hoshaw and, subsequently, Dr. C. J. Martin were key players in the Department of Navy. In the role as policy representative for the lead Service (Navy), they provided a strong headquarters advocacy. As career civilians, they provided a Bureau of Naval Personnel "corporate memory" for the CAT-ASVAB Program. This was essential in working with a succession of rotating senior Naval officers, who were responsible for the program over the years. In addition, they coordinated funding support essential for sustaining the program over many budget years and cutbacks.

This book would never have come to life without the efforts of Mrs. Margie Sands, Ms. Lola Zook, and Ms. Emma James. Mrs. Sands was the Administrative Assistant to Marty Wiskoff at NPRDC during most of the CAT-ASVAB Program. She edited the book chapters from the perspective of someone who had first - hand knowledge of the program over the years. Mrs. Zook (HumRRO) served as a technical/copy editor. Mrs. James (HumRRO) typed many iterations of the entire book. The editors appreciate the important contributions of these individuals.

The book was produced, in part, via an Army Research Institute for the Behavioral and Social Sciences (ARI) delivery order contract: Contract for Manpower and Personnel Research and Studies (COMPRS). Dr. Ron Tiggle (ARI) served as the delivery order Contracting Officer's Representative. Dr. Jane Arabian, Assistant Director for Enlistment Standards, Office of the Assistant Secretary of Defense (Force Management Policy), under Dr. Sellman, was the delivery order monitor.

The views, opinions, and findings contained in this book are those of the authors and editors. They should not be construed as representing an official Department of Defense position, policy, or decision, unless so designated by other official documentation.

About the Editors

W. A. "Drew" Sands has spent most of his career in military personnel research. He earned a Bachelor of Science in Social Sciences and a Master of Arts in Counseling and Testing Psychology from The American University in Washington, DC. In 1967, he joined the Naval Personnel Research and Development Laboratory in Washington as a Personnel Research Psychologist.

In 1973, Mr. Sands transferred to the Navy Personnel Research and Development Center (NPRDC) in San Diego, CA. His projects at NPRDC included the development of biographical/demographic screening and selection instruments for enlisted Navy personnel, and relating measured interests of Naval Academy midshipmen to choice of major academic area.

In 1980, he became the Head of the Computerized Personnel Accessioning Systems Branch of the Personnel Systems Department. He managed the R&D team that developed the Navy Personnel Accessioning System (NPAS) and the Computerized Adaptive Screening Test (CAST). In 1983, he became Head of the Computerized Testing and Accessioning Division in the Personnel Systems Department, which was focused on the Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB). In March 1986, he became the Director of the Personnel Systems Department at NPRDC, where he planned, directed, and evaluated the overall scientific research program in personnel screening, selection, classification, and performance assessment. As the Officer-in-Charge, he had the lead laboratory (NPRDC) responsibility for the Joint-Service CAT-ASVAB Program.

Mr. Sands retired from civil service in March 1994 and returned to Washington, DC. He has authored over 110 journal articles, technical reports, and professional presentations in various areas including: Psychological testing (paper-and-pencil and computerized adaptive tests); personnel screening, selection, and classification; survey design and analyses; computer-based vocational guidance; artificial neural networks; and, expert and decision support systems.

Brian Waters is Program Manager of the Manpower Analysis Program of the Human Resources Research Organization (HumRRO). He joined HumRRO in 1980, after retiring from the Air Force, where he taught and was Director of Evaluation at the Air War College, was an R & D manager and researcher with the Air Force Human Resources Laboratory, and was a navigator.

He holds a Ph.D. and M.S. in Educational Measurement and Testing from Florida State University, and an MBA from Southern Illinois University. His doctoral dissertation in 1974 was one of the earliest empirical studies of computerized adaptive testing (CAT), and he has over 20 years' experience with CAT R&D.

He is a fellow in the American Psychological Association (APA), and is a former President of the Division of Military Psychology of APA. He has authored over 100 journal articles, books, and professional papers, primarily dealing with the selection, classification, and testing of military and civilian personnel.

Jim McBride is a Principal Scientist on the staff of the Human Resources Research Organization HumRRO). A research psychologist, he has been involved in research and development related to computerized adaptive testing since 1972. During his doctoral studies in psychometric methods at the University of Minnesota, he was a research assistant to David J. Weiss, and participated in Weiss' pioneering CAT work for the Office of Naval Research. Since completing doctoral training in 1976, he has done test devel-

opment and personnel research for the Army Research Institute, NPRDC, The Psychological Corporation, and HumRRO.

At NPRDC, he was Principal Investigator on a variety of CAT-related projects ranging from the exploratory development work that provided the first empirical demonstration of CAT's efficiency for military personnel testing, to the design and development of prototype systems intended for nationwide administration of computerized adaptive versions of the Armed Services Vocational Aptitude Battery (ASVAB). At NPRDC, he designed and directed the development of the first complete computerized systems for adaptive ASVAB administration. At the time of his departure from NPRDC, he was Director of the Personnel Systems Department, with responsibility for the entire spectrum of scientific research related to Navy personnel selection, classification, and testing.

He joined The Psychological Corporation in 1984, as Director of its Computer-Based Testing Group; later, his responsibilities there extended to all development and research related to tests designed for personnel assessment in business, government, and career development. Between 1984 and 1990, he designed and directed development of a number of computer-based testing systems, including the first commercial application of CAT: the Computerized Adaptive Edition of the Differential Aptitude Tests.

Since joining HumRRO late in 1990, he has continued his involvement in R&D on computer-based testing in general, and CAT in particular. He directed the development of one of the first CAT systems used for personnel selection in industry, for a Fortune 100 HumRRO client. He has provided consulting services in computer-based testing to several other private-sector firms, and has been a member of an expert panel advising the U.S. Department of Labor on the development and evaluation of a computerized adaptive version of the General Aptitude Test Battery. He is currently directing the HumRRO project team responsible for modifying the Army's Computerized Adaptive Screening Test for use by all of the Armed Services.

FOREWORD

In October 1996, the Department of Defense (DoD) implemented a computerized adaptive testing (CAT) version of its enlistment test battery (the Armed Services Vocational Aptitude Battery or ASVAB) in 65 Military Entrance Processing Stations (MEPSs) across the country. DoD became the first organization to use CAT-derived scores for personnel selection when the system was placed in five MEPSs for operational testing in 1992; now DoD has become the first employer to adopt CAT for its employment system. This is a particularly impressive accomplishment when one considers the size of the program. The Department is the largest single employer of American youth, testing over 350,000 applicants for entrance into the Military Services between October 1, 1994 and September 30, 1995. Efficient enlistment processing and accurate measurement of individuals' aptitudes have been, and continue to be, critical concerns for the Department. Since 1970, DoD has sponsored the Joint-Service research and development of CAT-ASVAB and beginning in June 1992, recruits have joined the military on the basis of their CAT-ASVAB scores.

In the 1960s, the Office of Naval Research (ONR) sponsored work on computerized adaptive testing. The early research focused on the statistical techniques that allowed examinees to respond to different test questions tailored to their particular ability levels. Such statistical underpinning was imperative if CAT scores were to be interpreted against a normative reference group, as well as across time and test versions. Some of the nation's most eminent psychometricians such as Drs. Frederick Lord, Darrell Bock, Fumiko Samejima, Mark Reckase, and David Weiss were involved in this effort. At ONR, Drs. Marshall Farr and Charles Davis provided DoD vision and stewardship.

The Service personnel research laboratories began research directed at selection and classification and training applications in the 1970s. By the early 1980s, DoD had developed concepts for CAT acquisition. At that time, computer costs and portability were significant issues along with technical and psychometric questions. In 1984, the program received an unexpected, and probably unintentional shove forward by Lieutenant General E. A. Chavarrie, then Deputy Assistant Secretary of Defense (Military Manpower and Personnel Policy).

In November 1984, General Chavarrie was the keynote speaker at the Military Testing Association (MTA) conference in Munich, Germany. Part of his speech covered the status of CAT research in the American military. However, the day before the conference opened, General Chavarrie had visited several German recruiting offices where he saw applicants taking an enlistment test via computer. The test was not adaptive, but the General didn't know that; all he knew was that German youth were taking a computerized enlistment test, while the next day he was going to tell over 250 MTA conferees from ten countries that the United States would not be

implementing its computerized testing program for another five years. Consequently, General Chavarrie changed his speech (without informing his staff at the conference) and announced that he was accelerating CAT development by three years. As a result of General Chavarrie's speech, work on CAT assumed a new urgency. However, many technical issues remained that required several more years of intensive research.

In November 1991, W. S. Sellman, Director for Accession Policy in the Office of the Secretary of Defense, presented the opening speech to a NATO Workshop on Computer-Based Assessment of Military Personnel. His address focused on three areas (psychometrics, economics, and politics) pertinent to CAT. A copy of that speech follows this Foreword. In that speech he emphasized the need to resolve issues in all three areas before CAT could become a reality. Now, five years later, we finally have implemented CAT: Technical issues have been resolved, costs of computers have come down (along with their size and weight), and in the current political environment marked by substantial personnel and resource reductions, cost-benefit analyses supported the decision to buy over 1,400 computers for enlistment testing. DoD now is looking ahead for ways to make the most efficient use of CAT (for example, by developing items on-line rather than through separate, labor-intensive data collections) and, in a concepts of operation study, is evaluating alternative approaches for bringing CAT-ASVAB, or some other electronic testing medium, to remote, temporary test locations in a cost-effective manner.

For over 30 years, the CAT-ASVAB program has benefited greatly from the support of military visionaries and users; we expect continued excitement and support in the future. Up to now, the military has especially appreciated CAT because of its potential to reduce testing time, thereby saving valuable resources. But CAT-ASVAB will provide even more benefits once fully implemented. It will not only be easier to incorporate new tests (such as psychomotor tests that require computer administration) and develop new items via on-line item development programs, but it also may be possible to tailor the enlistment testing session to include Service-specific tests for applicants.

Technical issues aside, CAT-ASVAB provides a superior testing situation for all applicants to military service, regardless of their aptitude. Individuals who would struggle through typical paper-and-pencil tests find CAT to be challenging, but not overwhelming. They do not encounter large numbers of items that are far beyond their capabilities. Higher aptitude individuals, on the other hand, are challenged by CAT-ASVAB and, we hope, positively influenced by the military's high technology image. It provides a winning situation for everyone.

The well-justified pride of DoD and Service policy makers and researchers, including civilian scientists working under contract is conveyed in the following pages. This book captures the

long, involved history of CAT-ASVAB implementation. It documents technical information that will be helpful to other scientists and the test development community in general as computerized testing becomes the standard test delivery method for large-scale testing programs.

W. S. (Steve) Sellman, Ph. D.
Director for Accession Policy
Office of the Assistant Secretary of Defense
(Force Management Policy)
U.S. Department of Defense

Jane M. Arabian, Ph. D.
Assistant Director for Enlistment Standards
Office of the Assistant Secretary of Defense
(Force Management Policy)
U.S. Department of Defense

Foreword

COMPUTER ADAPTIVE TESTING
Psychometrics, Economics, and Politics

by

Dr. W. S. Sellman

Director for Accession Policy
Office of the Assistant Secretary of Defense
(Force Management and Personnel)
U.S. Department of Defense

Presentation at the

Workshop on Computer-Based Assessment of Military Personnel
NATO Defense Research Group
Brussels, Belgium

November 26, 1991

Foreword

Computer Adaptive Testing: Psychometrics, Economics, and Politics

by

Dr. W. S. Sellman
Director for Accession Policy,
Office of the Assistant Secretary of Defense (Force Management and Personnel)

Introduction

Good afternoon ladies and gentlemen. It is a pleasure to be here in this beautiful country, on the occasion of the NATO Defense Research Group Workshop on Computer-Based Assessment of Military Personnel to provide opening remarks to such a distinguished group of professionals. The presentations and discussions which will occur here during the next few days will be important to all of our efforts to develop and deliver effective military personnel testing programs. My background is in personnel psychology, and in my current position, I am responsible for setting policy to attract, qualify, and process young people into the military. This includes ensuring the quality of testing for military personnel selection and job classification in the United States.

The U.S. Department of Defense operates the world's largest testing program. Each year, we administer the Armed Services Vocational Aptitude Battery (ASVAB) to over two million young men and women. Last year, the enlistment version of the ASVAB was given to about 900,000 applicants for military service at approximately 1,000 testing sites across the country. ASVAB also was administered to 1.1 million students in over 14,000 secondary and post-secondary schools as part of the DoD Student Testing Program. In addition to operating the world's largest testing program, we also want to operate the world's best testing program. Today, I would like to share with you my views on one of our new testing initiatives--computer adaptive testing--and its promise for improving the way we assess the aptitudes of new recruits.

Computer Adaptive Testing (CAT)

Computer adaptive testing represents the most significant breakthrough in personnel testing in the last 30 years. Although the most noticeable change in the new method of testing is the fact that the test is administered by computer, the essential difference between this method and paper-and-pencil tests is that each examinee answers a special set of test questions "tailored" to his or her ability. Adaptive testing is a way of allowing those tested to answer only those questions that are suited to their individual abilities. This contrasts with conventional group

testing procedures which require many people to spend time on questions that may be either too easy or difficult for them.

Computer adaptive testing (CAT) has major benefits, both in efficiency and test quality. The examination time will be shorter, and the test, as a whole, will be more precise. Because examinees cannot be sure which questions will be asked, CAT also retards, if not eliminates, the problems of test compromise.

With these potential advantages over paper-and-pencil tests, computer adaptive testing should be the testing technology of the future. Yet, it is unclear if the U.S. military will be able to implement an operational CAT system as part of our enlistment process. This is because of the nexus of conflicting pressures that must be resolved before CAT can become a reality. For the next few minutes, I would like to tell you about those pressures, i.e., the factors that ultimately will influence the CAT decision--psychometrics, economics, and politics.

Psychometrics of Computer Adaptive Tests

Let me begin by presenting some psychometric considerations. The enlistment test in use from 1976 through 1980 was miscalibrated. This inflated the scores of low aptitude examinees and resulted in the enlistment of over 300,000 young people who would not have qualified with accurate scores. The revelation of this calibration error led to several major research efforts. The enlistment test was administered to a nationally representative sample of youth ages 16-24 to develop new norms. A large-scale criterion project also was begun, to link enlistment standards to actual job performance.

How and why are these studies relevant to CAT? We report aptitude levels of new recruits to Congress and the American public using a percentile scale that enables comparisons across Services and time. Thus, each version of our test must be calibrated correctly against the normative population. Otherwise, scores would lose their meaning and could not be interpreted. New recruits also qualify for enlistment incentives (e.g., bonuses and educational benefits) and are placed into military occupations on the basis of their scores.

In addition, the Services defend their requests for recruiting resources using aptitude as an index of recruit quality. If the aptitude levels of a Service were low, then that Service would justify additional funds to recruit higher quality young people. These brighter recruits ultimately return the investment on recruiting resources because, when compared with their lower scoring peers, they are more trainable, perform better on the job, have lower rates of indiscipline, and are more likely to complete their obligated tours of duty. Consequently, it is imperative that enlistment test scores are accurate reflections of the ability levels of new recruits.

We know how to calibrate paper-and-pencil tests to one another. However, when we began the CAT research we did not know how to equate a paper-and-pencil test to one administered by computer. For the past five years, we have been collecting data administering the enlistment test and a CAT version to large samples of military applicants. Today, we are

convinced that a person taking a CAT test would receive the same score as if he or she took the paper-and-pencil version.

This ability to calibrate paper-and-pencil and computer tests means we can transition to a computer enlistment test knowing that we can still track aptitude across Services and time. Had we not been able to equate the two types of tests, we could never use CAT because we could not interpret its scores against our normative base or against previous distributions of recruit aptitude. Fortunately with the help of some of the best psychometricians in the United States, we were able to solve that problem.

Economics of Computer Adaptive Tests

In addition to calibrating CAT to our normative population, we also must demonstrate its relative cost utility for selection and classification. In the mid 1980s, we began research to examine the relationship between CAT scores and performance in technical training. The validity coefficients for CAT turned out to be of the same approximate size as those of the paper-and-pencil ASVAB. This was not surprising, since CAT used the same types of questions (verbal, mathematics, reading, technical information) as are found on the operational enlistment test. The only differences between the two types of tests were the "tailored" nature of the questions administered by computers.

While the validity research was underway, we also conducted a cost-benefit analysis for CAT. It would be prohibitively expensive to buy computers for all 1,000 locations where we administer ASVAB. Consequently, we explored a variety of siting strategies that essentially either took the test to the applicant or the applicant to the test. In particular, we considered (1) transporting all applicants to a small number of centralized sites, (2) additional testing at high volume sites, and (3) testing of applicants at portable locations. Costs for each of these strategies were computed, along with costs of paper-and-pencil testing under existing procedures. When the results were in, computer adaptive testing would have increased costs over the paper-and-pencil ASVAB by $17 million for centralized testing and by $132 million for portable testing.

At the same time, the benefits of CAT also were being considered. Using a valid test during selection and classification reduces personnel costs through enhanced performance in training and on the job, and also yields lower attrition. (It costs approximately $20,000 to recruit, train and equip replacements for people who do not complete an obligated tour of duty.) Unfortunately, the validity of CAT was not appreciably higher than for the paper-and-pencil ASVAB. As a result, we could not demonstrate improved enlistment processing though the use of CAT, nor could we justify the costs of purchasing computers for enlistment testing.

New Predictors

One advantage of computerized testing is that new types of tests can be administered that are not possible with paper-and-pencil tests. These include psychomotor tracking, cognitive processing, and tests of short- and long-term memory. If these tests were more valid than

conventional tests, then we should be able to improve selection and classification. With the results of the cost-benefit analysis in mind, we initiated a new phase in the CAT project--development and validation of tests that can only be administered via computer. To date, experimental tests have been constructed, and we are currently administering them to new recruits in a variety of military specialties to learn if they improve our ability to predict performance. Preliminary results are encouraging, but we need more hard data to prove the utility of the new tests.

While this research on new computerized predictors is ongoing, we have returned to the issue of how and where to administer CAT. We have recently awarded a contract to the Human Resources Research Organization (HumRRO) to develop and evaluate alternative procedures for administering and scoring enlistment tests. In particular, HumRRO will devise strategies that vary in mode of administration. Test administration for the different strategies may either be paper-and-pencil or computer (CAT and the new computerized predictors) or a combination of both.

In addition, HumRRO will examine strategies based on a "stage of processing" model. Currently, the paper-and-pencil ASVAB is administered in one-stage (i.e., all examinees take the test during a single session). A viable alternative to this strategy is a two-stage approach where a short test is administered as an initial screen and clearly unqualified applicants eliminated. Only those people with a chance of qualifying would be tested further in a second administration. Dr. Jim McBride, principal investigator for this effort, is here at the workshop and will share his plans for the research with you in more detail.

Politics of Computer Adaptive Tests

Let me close with a brief mention of the politics of CAT. The United States faces a large budget deficit, and our Congress is struggling to discover ways to reduce it. This means that all Government spending receives considerable scrutiny. At the same time, the U.S. military is being reduced from 2.1 million uniformed members to 1.6 million members by FY 1995. The downsizing is a direct result of the reduced threat from the Soviet Union and the Warsaw Pact countries. As the size of the military drops so does the budget for the U.S. Department of Defense. Over the past three years, our recruiting budget has declined by 16 percent, and it will continue to drop as our force reductions continue.

What does all this have to do with CAT? In times of austere resources, any new system must be carefully documented and justified. In order to receive approval for CAT within the Department of Defense and by Congress, we must be able to demonstrate that savings accrued by improved selection and classification can amortize the cost of buying computer hardware. In other words, the benefit of computerized enlistment testing must outweigh the costs of buying the computers. Otherwise, we will never be able to defend our request to implement computer adaptive testing.

Conclusion

Lest I appear overly pessimistic, we have made great progress in the development of computerized tests over the past 10 years. Today, we know a lot that once was only speculation. For example, CAT can reduce testing times by almost one half (3 hours down to 1 1/2). CAT enhances the image of the military with applicants for enlistment who view the technology as an indicator that the military is technically sophisticated. Applicants prefer to take a computerized test versus a paper-and-pencil test. CAT provides more precise measurement for those at the extremes of ability (i.e. high and low aptitude people), although our paper-and-pencil measure still works best for those of average ability. Equivalent scores can be obtained whether paper-and-pencil or computer adaptive versions of our enlistment test are administered. Finally, new measures which can only be administered by computer have shown improvements in the prediction of training and on-the-job success.

As I said at the beginning of this presentation, we must be able to deal with the psychometric, economic, and political issues before implementing an operational CAT system. I believe we have solved most of the psychometric problems, and we are working on the others with a sense of urgency. I am hopeful that the time for computerized testing is close at hand. The development of tests that can only be administered on computer has potential to add incremental validity above that for the paper-and-pencil ASVAB, and the decrease in administration time for CAT may well lead to savings in the costs of enlistment processing.

But there are still lessons to be learned and hard decisions to be made before our recruits are tested by computer. In the near future, we will implement CAT at four sites to examine operational issues and to determine once and for all whether the benefits of computerized testing are real. Obviously, the science and politics of CAT represent complex problems that defy simple solutions. I thank you for the invitation to participate in this workshop and trust that my comments will provoke informed dialogue. In the United States, our goal is to test applicants for military service in the most cost-effective way possible; I believe the CAT program has been developed with that long-term vision in mind.

Foreword

TABLE OF CONTENTS

LIST OF TABLES

List of Tables, Continued

List of Tables, Continued

List of Tables, Continued

LIST OF FIGURES

SECTION I - BACKGROUND

The introductory section of this book provides readers who have little or no familiarity with the Armed Services Vocational Aptitude Battery (ASVAB) and/or computerized adaptive testing (CAT) with some background to lay a foundation for the information presented in the remainder of the book. The three background chapters cover (1) Introduction to ASVAB and CAT, (2) R&D Laboratory Management Perspective, and (3) Technical Perspective. References are made throughout Section I to later chapters which deal with relevant issues in more detail.

Chapter 1, "Introduction to ASVAB and CAT," by Drew Sands and Brian Waters, introduces both the test battery and the concept of computerized adaptive testing. The authors sketch the background of present day ASVAB testing by the U.S. Armed Forces to establish an historical perspective. The ASVAB is administered under two Department of Defense (DoD) programs: The DoD Student Testing Program (DoD-STP), and the Enlistment Testing Program (ETP). The authors first discuss DoD-STP, including the purpose of the student contacts, and describe its vocational guidance tools. Next, they describe the two military test administration environments of the ETP: Military Entrance Processing Stations and Mobile Examining Team Sites.

The two objectives of the ASVAB program are personnel selection and classification. The chapter describes the tests that make up the ASVAB, exploring the aptitudes and qualifications of those who may apply for military service. The process of developing the normative information for ASVAB is also presented. The next section of the initial chapter then addresses CAT, describing this computerized adaptive approach to aptitude measurement and its advantages over conventionally admini4stered, paper-and-pencil aptitude testing.

Chapter 2, "R&D Laboratory Management Perspective," was written by Marty Wiskoff to view CAT-ASVAB as a manager saw it. This chapter describes the major stages of the Navy Personnel Research and Development Center (NPRDC) program, including the process of initiating a CAT R&D capability, performance of the early research under the Marine Corps as the lead Service in the DoD Joint-Service CAT-ASVAB Program, and the transition of lead Service responsibilities to the Navy.

Wiskoff then addresses support and organizational issues, including obtaining management, policy maker, and funding support. Covered also are the topics of professional staffing and organization at NPRDC. The oversight and coordination in the Joint-Service arena and the need for accommodation to changing requirements are discussed, along with examples of international cooperation and technical exchanges.

The next discussion addresses research management issues, and includes (1) psychometric research, (2) the CAT-ASVAB delivery system, (3) economic (cost/benefit) analyses, (4) the introduction of the Enhanced Computer Administered Tests (ECAT), (5) various concepts of operation, and (6) the process of monitoring and coordinating CAT-ASVAB research. Finally, the author offers some recommendations for CAT R&D.

Jim McBride authored Chapter 3, "Technical Perspective." This chapter provides an overview of the CAT-ASVAB project from a technical point of view, both for equipment and for research considerations. After characterizing the testing situation as it existed in 1979, McBride describes CAT delivery system development during the 1980s, when the rapidly changing hardware technology had an important impact on CAT progress and direction. Military CAT hardware evolved from Apple II-plus computers to Hewlett-Packard standalone machines to IBM-compatible personal computers in a little over a decade; meanwhile CAT research on test presentation went forward on a parallel course.

After reporting on the competitive "flyoff" between three competing firms to design and build a prototypical CAT system, McBride describes the CAT psychometric research and development progress over the 15-year period, and the establishment of the research base upon which all current CAT is built.

Chapter 1

INTRODUCTION TO ASVAB AND CAT

by

W. A. Sands[1] and Brian K. Waters[2]

The Armed Services Vocational Aptitude Batttery (ASVAB) and computerized adaptive testing (CAT) are the topics of central importance throughout this book. The purpose of this introductory chapter is twofold: (1) to provide the reader with a brief introduction to ASVAB and CAT, and (2) to consolidate basic information on these two topics, providing a framework for the more detailed presentations in the following chapters.

MILITARY PERSONNEL SCREENING

Aptitude testing plays a central role in the military personnel screening process. Indeed, the military places far more emphasis on aptitude testing as a selection tool than does the civilian sector. This difference is the result of a number of factors:

- The majority of individuals in the primary age group of applicants targeted by the military (17 - 21 years old) has no significant employment history to aid in selection decisions.
- The military selects people for a wide variety of training and jobs.
- The overall military screening process is quite expensive, in part because of the large numbers of people involved. Group-administered tests offer efficiencies in time, cost, and psychometric precision that are quite appealing.
- The large number of people tested enables the military to conduct large-scale, empirical studies to obtain evidence for the validity, reliability, fairness, and differential impact of tests on various subgroups. This information is useful in meeting current professional standards for the use of employment tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985; American Psychological Association, 1980).

HISTORICAL ANTECEDENTS[3]

The early history of military testing is briefly characterized by Eitelberg, Laurence, and Waters, with Perelman (1984).

> *The American military was a pioneer in the field of aptitude testing during World War I. In 1917 and 1918, the Army Alpha and Army Beta tests were*

[1] Chesapeake Research Applications (Consultant to the Human Resources Research Organization).

[2] Human Resources Research Organization.

[3] Additional information on the history of the U.S military's use of aptitude screening tests may be found in a number of Department of Defense publications, (for example: Eitelberg et al., 1984; ASVAB Working Group, 1980;and Department of the Army, 1965).

> *developed so that (1) military commanders could have some measure of the ability of their men, and (2) personnel managers could have some objective means of assigning the new recruits. The Army Alpha test was a verbal, group-administered test used primarily by the Army for selection and placement. The test consisted of eight subtests -- including verbal ability, numerical ability, ability to follow directions, and information -- and served as a prototype for several subsequent group-administered intelligence tests. The Army Beta test was a nonverbal, group-administered counterpart to the Army Alpha test. It was used to evaluate the aptitude of illiterate, unschooled, or non-English-speaking draftees. ...*
>
> *The Army General Classification Test (AGCT) of World War II largely replaced the tests of World War I. The AGCT was described as a test of "general learning ability" and was intended to be used in basically the same manner as the Army Alpha (i.e., an aid in assigning new recruits to military jobs) (Eitelberg et al., 1984, pp. 14-15).*

Between World War II and 1976, each of the Services employed its own set of tests to determine initial eligibility for enlistment and for subsequent classification decisions. These tests included measures of general trainability and specific aptitudes considered important to the Services.

The Selective Service Act (1948) mandated the development and use of a common basis for determining U.S. military enlistment eligibility. At that time, the Army General Classification Test (AGCT) was the most widely used personnel screening instrument in the military. This test became the model for the Armed Forces Qualification Test (AFQT), the Joint-Service selection test designed to address the congressional mandate. The AFQT became operational in 1950.

The original AFQT contained three types of items: verbal, arithmetic reasoning, and spatial relations. Since that first version, various content changes have been introduced. During the period 1972-75, the Services were not required to use the AFQT. Rather, each Service was permitted to use its own test battery and conversion tables to estimate the AFQT score for each person (ASVAB Working Group, 1980).

ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB)

In 1966, the Department of Defense (DoD) directed the individual Military Services to explore the development of a single, multiple-purpose aptitude test battery that could be used in high schools. This direction was designed to prevent costly duplication by the military and schools, and to encourage equitable selection standards across Services (DoD, 1992).

Since 1976, the ASVAB has been the common selection and classification battery for the four (DoD) Services and the Coast Guard (Department of Transportation). New forms of the battery have been produced approximately every three to four years. At the time of this writing, P&P-ASVAB Forms 20 through 22 and CAT-ASVAB Forms 01 and 02 are currently in operational use and Forms 18 and 19 are used in the high schools.

ASVAB Testing Programs

DoD Student Testing Program (DoD-STP). The ASVAB was introduced into the high school setting during the 1968-69 school year. DoD provides the ASVAB, an interest inventory, and a host of supporting materials to participating schools free of charge. The benefit to the schools is a well-researched, multiple-aptitude test battery to provide career guidance and counseling services to students. This benefit is especially important to schools in an era

of budget reductions, as the ASVAB program sometimes is the only vocational guidance information available to counselors and their students.

According to Wall (1995), the purposes of the ASVAB Career Exploration Program (CEP) are to:

- Provide information to students about their abilities, interests, and personal preferences
- Provide information to students on civilian and military occupations
- Help students identify civilian and military occupations that have characteristics of interest to them
- Identify for the Services aptitude-qualified individuals who may be interested in joining the military

ASVAB as a Counseling Tool. The ASVAB CEP provides a comprehensive set of educational and career counseling tools for the student and school counselor for their use as the student learns career decision skills. The program includes ASVAB scores, a DoD-published interest inventory, and exercises designed to help students identify their personal preferences (DoD, 1992).

Interest-Finder. DoD's license to use the *Self-Directed Search* (*SDS*), a commercially published interest inventory, expired in July 1995. Therefore, DoD developed an interest inventory, the *Interest-Finder*, which was implemented in the DoD-STP during the 1995-96 school year. Like the *SDS*, the *Interest-Finder* uses Holland's classification codes (Holland, 1973) to cluster interests into related occupational areas. The instrument has extensive research and development underlying its use in the schools.

ASVAB Career Exploration Program Counseling Materials. A number of CEP printed materials are currently provided to participating schools and students. These materials can be obtained from local military recruiters or from ASVAB Education Services Specialists at Military Entrance Processing Stations (DoD, 1992). Current ASVAB CEP printed materials include:

- *ASVAB 18/19 Educator and Counselor Guide*
- *ASVAB 18/19 Counselor Manual*
- *ASVAB 18/19 Technical Manual*
- *ASVAB 18/19 Student and Parent Guide*
- *Exploring Careers: The ASVAB Workbook*
- *Military Careers*

ASVAB as a High School Recruiting Tool for the Military. A major benefit of the DoD-STP to the military is the recruiting leads provided by the results. ASVAB score information enables Service recruiters to focus on students who will be likely to qualify for enlistment. Hence, the DoD-STP serves as a mechanism to pre-qualify student recruiting prospects. The ASVAB is administered in about 14,000 schools. The number of students tested in the schools has been decreasing, with 931,000 tested during the 1990-91 school year, 882,000 in 1991-92, and 880,294 in 1992-93 (Branch, personal communications, 1995).

Enlistment Testing Program. The Military Services began using the ASVAB in 1976. In FY 1993, about one half million prospects took the ASVAB for active duty (358,755), Reserve (73,244), and National/Air National Guard (67,383) recruiting programs (Branch, personal communication, 1995). As with the DoD-STP, the Defense drawdown has led to decreasing numbers of military applicants taking the ASVAB since 1988.

Active, Reserve, and much National Guard ASVAB testing is conducted in 65 Military Entrance Processing Stations (MEPSs) and their nearly 700 associated, satellite Mobile Examining Team Sites (METSs). The MEPSs and METSs are part of the U.S. Military Entrance Processing Command (USMEPCOM), a Joint-Service agency headquartered in North Chicago, Illinois, which is responsible for administering the ASVAB, physical examination and medical qualification, and other enlistment processing activities for the Armed Forces. USMEPCOM essentially handles all enlistment processing activities from the time that a prospect begins the testing program until he or she ships to a Service recruit training center.

Military Entrance Processing Stations (MEPSs). The approximately 65 MEPSs (the number is shrinking during the Defense drawdown) are geographically dispersed applicant processing centers which have ASVAB test-

ing rooms, answer sheet scanners and computer equipment, medical and physical examining facilities, and offices for Service career counselors (classifiers) to interact with prospects about options for military jobs, training class seats, and shipping dates. The ASVAB is administered by military personnel in the MEPSs, in a carefully controlled testing environment.

Table 1-1
Armed Services Vocational Aptitude Battery (ASVAB) Tests: Description, Number of Questions, and Testing Time[a]

ASVAB Test Title and Abbreviation	Description	Number of Questions	Testing Time (Minutes)
Arithmetic Reasoning (AR)	Measures ability to solve arithmetic word problems	30	36
Word Knowledge (WK)	Measures ability to select the correct meaning of words presented in context and to identify best synonym for a given word	35	11
Mathematics Knowledge (MK)	Measures knowledge of high school mathematics principles	25	24
General Science (GS)	Measures knowledge of physical and biological sciences	25	11
Mechanical Comprehension (MC)	Measures knowledge of mechanical and physical principles and ability to visualize how illustrated objects work	25	19
Electronics Information (EI)	Measures knowledge of electricity and electronics	20	9
Auto and Shop Information (AS)	Measures knowledge of automobiles, tools and shop terminology and practices	25	11
Coding Speed (CS)	Measures ability to use a key in assigning code numbers to words in a speeded context	84	7
Numerical Operations (NO)	Measures ability to perform arithmetic computations in a speeded context	50	3
	Total for All Tests	334	144[b]

[a]Source: Eitelberg, M.J. (1988). *Manpower for military occupations.* Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

[b]Administrative time is 36 minutes, for a total testing and administrative time of 3 hours.

Mobile Examining Team Sites (METSs). Each MEPS has several relatively small, satellite testing sites which operate under its control. In a given METS, testing frequency may range from less than once per week to several times a week. METSs are located in various types of facilities, ranging from post offices and other public buildings to leased space. The METSs administer the ASVAB and some specialized Service tests; qualifying applicants who wish to continue the screening process proceed to the MEPS for medical and physical examinations and other processing. The ASVAB is administered at the METSs by part-time Office of Personnel Management (OPM) test administrators (TAs). The answer sheets are optically scanned at the MEPS, generally a day or two following METS testing, although recruiters are given an unofficial hand-scored AFQT score for their applicants immediately after ASVAB testing.

ASVAB Tests. At present, the paper-and-pencil (P&P) version of the ASVAB contains 10 tests. The name, description, and testing time for each are presented in Table 1-1 on the preceding page. They include eight power (relatively unspeeded) tests (Arithmetic Reasoning [AR], Word Knowledge [WK], Paragraph Comprehension [PC], Mathematics Knowledge [MK], General Science [GS], Mechanical Comprehension [MC], Electronics Information [EI], and Auto and Shop Information [AS]); and two speeded tests (Coding Speed [CS] and Numerical Operations [NO]). The first four are measures of general trainability, while the following four tap learned abilities predictive of success in specific jobs and clusters of military jobs. The two speeded tests predict performance on certain military tasks that require highly speeded activities or rapid information processing. Factor analytic studies of the ASVAB have consistently yielded four factors -- Verbal (WK, PC, and GS), Quantitative (AR and MK), Technical (EI, MC, and AS), and Speed (CS and NO) factors (cf: Waters, Barnes, Foley, Steinhaus, & Brown, 1988).

ASVAB Operational Use. The ASVAB is used for two main purposes in military enlisted accessioning: selection of new recruits from applicants, and subsequent classification of recruits into one of the many jobs available. Scores from AR, WK, PC, and MK are combined into the Armed Forces Qualification Test (AFQT) composite score for each applicant. The AFQT measures trainability and predicts job performance in the military. AFQT has been shown to be valid for these uses in the four Military Services and the Coast Guard. AFQT scores are calculated on a percentile scale ranging from 1 to 99. They are reported to Congress by "AFQT Categories," shown in Table 1-2.

Table 1-2
Armed Forces Qualification Test (AFQT) Categories by Corresponding Percentile Scores and Level of "Trainability" [a]

AFQT Category	AFQT Percentile Score Range	Level of Trainability
I	93 - 99	Well Above Average
II	65 - 92	Above Average
IIIA	50 - 64	Average
IIIB	31 - 49	Average
IV	10 - 30	Below Average
V	1 - 9	Well Below Average/ Ineligible for Enlistment

[a] Source: Department of Defense, *Defense Manpower Quality: Volume I* (Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics), 1985, p. 9.

ASVAB Norms Development. Prior to 1980, ASVAB scores were statistically referenced to the population of all male military personnel on active duty on December 31, 1944. This 1944 reference population served as the

normative base for U. S. military selection tests until the mid-1970s. Since 1984, ASVAB scales have been based upon ASVAB testing of a nationally representative sample of over 12,000 youth 18 to 23 years old (DoD, 1982). The study was part of the National Longitudinal Survey of Youth Labor Force Behavior (NLSY79), sponsored jointly by DoD and the Department of Labor (DoL). The NLSY79 has provided the current normative base for all ASVAB test and composite scores (Waters, Laurence, & Camara, 1987). DoL and DoD are presently planning for a computer-based renorming of the ASVAB, scheduled for 1997.

ASVAB Summary

The ASVAB and its predecessor military tests are exemplars in large-scale, multiple-aptitude selection and classification testing programs. Extensive research and development programs have produced an efficient, accurate, and useful testing program for selecting and assigning hundreds of thousands of young persons annually. With its extensive use in experimental, and now operational, test and evaluation in computerized adaptive testing (CAT), the ASVAB provides a solid basis for the future of military personnel selection and classification.

COMPUTERIZED ADAPTIVE TESTING (CAT)

Traditionally, large-scale aptitude testing has used conventionally-administered, paper-and-pencil, multiple-choice tests. Psychometric developments in item response theory (IRT) (Lord, 1980a), in conjunction with advances in computer technology, have made an alternative approach, computerized adaptive testing (CAT), feasible (McBride, 1979).

Description

As the name indicates, a CAT instrument is computer administered. Less obvious is the way in which the test, dynamically adapts itself to an examinee's ability during the course of test administration. In a conventionally administered, paper-and-pencil aptitude test, every examinee takes the same items, typically in the same order, regardless of the item's appropriateness for a given examinee's ability level. Administering easy items to a high ability applicant is wasteful, as correct responses provide relatively little information about that examinee's ability. In addition, the person may become bored with test items that offer no challenge and may respond carelessly, introducing additional measurement error. Similarly, adminstration of hard items to a low-ability examinee is wasteful as incorrect answers do not provide much information on that person. Moreover, low-ability examinees are likely to find most items too difficult, and may become frustrated and respond randomly, also introducing additional error into the testing process. In contrast, a CAT instrument "tailors" the test to each examinee, as information is collected and evaluated during test administration.

The adaptation process can be illustrated with a hypothetical, 5-item test, shown in Figure 1-1 (Wiskoff, 1981). At the beginning of the test, we have no information about the ability level of the examinee, so we assume that person is average in ability (theta = 0.00). Therefore, an item of average difficulty is chosen for administration. Let us suppose that the examinee correctly answered the first item. Our initial ability estimate (average ability) is updated (in this case, raised to theta = 1.5), and a second (more difficult) item is chosen for administration. Now, suppose that the examinee selected an incorrect answer to the second item, suggesting that it was "too hard." Again, the computer updates the ability estimate (this time in a downward direction to theta = 0.75). Then, the next item is selected for administration at that difficulty level. This third item would be less difficult than the second item, reflecting the latest estimate of the person's ability. Suppose that the examinee also answered this third item incorrectly. Again, the ability estimate is updated (lowered to theta = 0.38) and the next item is chosen. Item 4 would be easier than the third item. If the examinee correctly answered this item, the ability (theta) estimate would be raised, and a more difficult item (theta = 0.56) would be presented as the last item in this hypothetical, 5-item adaptive test.

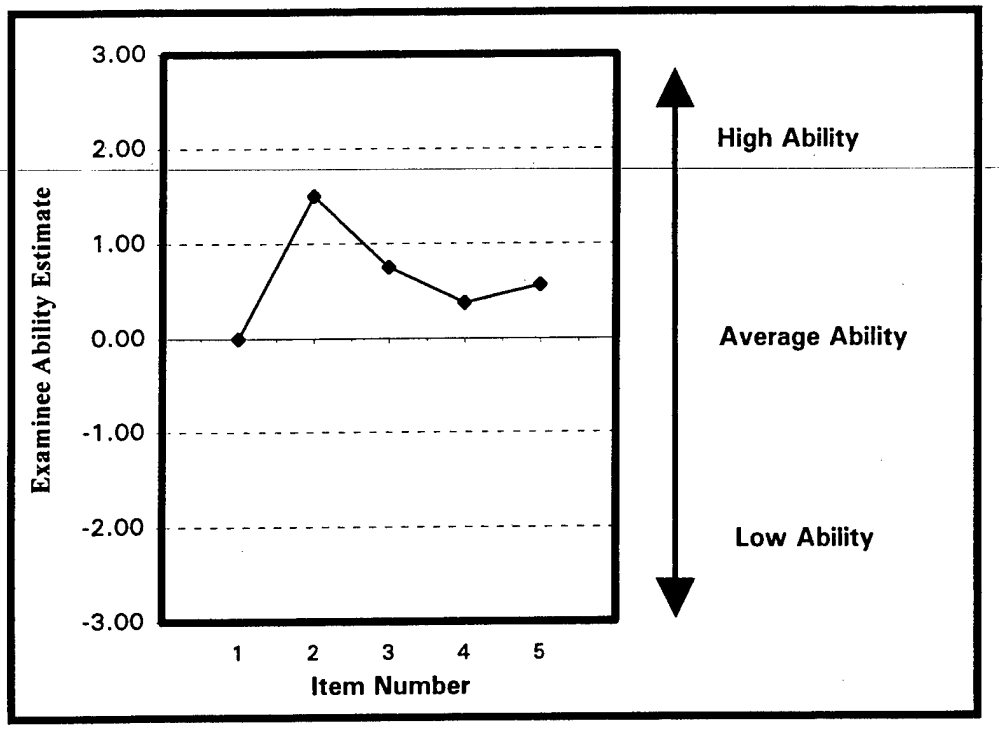


Figure 1-1. Hypothetical 5-Item Computerized Adaptive Test Results.

This process of selecting and administering a test item, scoring an examinee's response, updating his or her ability estimate, and choosing the next item for administration continues until a specified stopping rule is satisfied. The stopping criterion might be administration of a predetermined number of items (fixed-length testing), reduction of the standard error of measurement to a pre-specified level (variable-length testing), or a hybrid combination of the two stopping criteria (see Chapter 4 for discussion).

In comparison to a paper-and-pencil test, the adaptive nature of the CAT instrument produces a very efficient testing session, as illustrated in an example in Figure 1-2. In the example, all paper-and-pencil (P&P) examinees take all 20 test items, regardless of their ability. However, in the CAT test, a low-ability examinee takes a subset of 10 relatively easy items, a person of average ability takes 10 items in the mid-range of difficulty, and a high-ability person takes a subset of 10 relatively more difficult items. In the hypothetical situation portrayed in Figure 1-2, the CAT instrument entails only half the number of items (10) required of the P&P test (20) for comparable test precision, producing a substantial savings in test administration time.

Advantages of CAT

Administrative. A CAT version of a test offers four adminstrative advantages over a P&P version of the same test. Reduced test session length is the first advantage. Since each item presented to a particular examinee is appropriate for the current estimate of that person's ability level, no items are wasted. The number of test items administered in an adaptive test is substantially lower than in a traditional test. This reduction is made possible by obtaining more information about the examinee's actual ability per item administered. This, in turn, reduces the test length required to yield a fixed level of measurement precision.

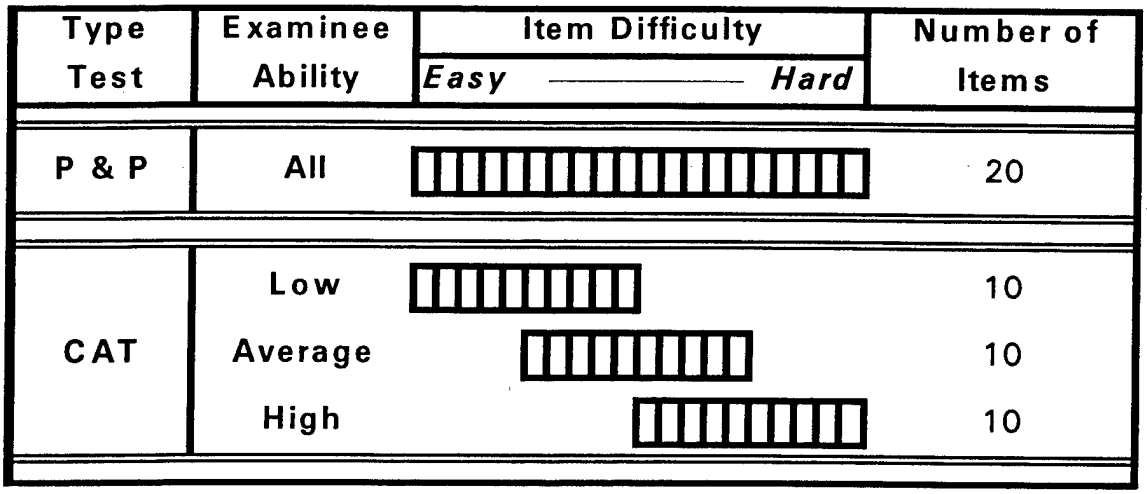


Figure 1-2. Test Item Utilization for Paper-and-Pencil Tests and Computerized Adaptive Tests.

A second administrative advantage of CAT is test session flexibility. The P&P-ASVAB is a group-administered test battery with all examinees starting and ending the test battery together. All examinees are given instructions by the test administrator (TA), and all examinees take each test in the battery simultaneously. Persons finishing a test early must wait for the entire scheduled time for that test to end. Then, all examinees move ahead in lock-step fashion to the next test. In contrast, examinees can begin CAT-ASVAB, individually, at any time. Test battery administrative instructions are provided by the microcomputer. When an examinee finishes a CAT test, that person can proceed directly to the next test. This flexibility increases examinee flow, making the overall testing process more efficient.

A third administrative advantage of CAT is greater standardization. Although P&P-ASVAB is administered with a standard set of instructions and specified time limits for each test, the actual practice may be less standardized than is desirable. While extension is prohibited, the TA might, for example, allow "a little extra time" for a particular test. The testing procedures are more standardized for a CAT instrument, as the computer precisely controls the test administration.

Fourth, CAT administration simplifies test revision. Revision of a P&P-ASVAB is a time-consuming, logistically cumbersome, and expensive process. After a large supply of experimental items is developed, they are organized into sets of overlength forms and administered to groups of recruits in basic training. Since the schedule in recruit basic training is typically quite full, scheduling test administration sessions can often be problematic. The collected data are scored, then analyzed to cull out items that exhibit poor psychometric characteristics. Those items that survive the process are organized into operational-length test forms. The test forms must then be printed and distributed nationwide.

In CAT-ASVAB, a few embedded experimental items can be administered routinely as each person takes the operational battery. Performance on the experimental items has no impact on a person's scores. Administration of experimental items is transparent to both the examinee and the TA. Thus, the computer provides an opportunity to collect a wealth of item data for future item calibration, without the disruption and lengthy development process necessary in P&P-ASVAB form revision.

Scoring. A computer-based delivery system reduces errors that occur due to reliability problems with optical scanning equipment used to score the P&P-ASVAB. In addition, the possibility for clerical error is greater when hand-scoring takes place. Finally, CAT-ASVAB results are available virtually immediately. If policy permits, scores can be given to the applicant and to the recruiter immediately after the test battery is completed.

Measurement Precision. The measurement precision of the typical P&P test is peaked around the average ability level of the target population. This means that most of the items cluster around medium difficulty, while there are relatively few easy or difficult items. Although this strategy of test development usually produces high

measurement precision for "average" people, the measurement precision for examinees at both ends of the ability distribution is typically considerably less. Since each CAT-ASVAB test is designed to be appropriate for each examinee's ability level, measurement precision is improved for both low- and high-ability examinees, while matching the precision of P&P-ASVAB for average-ability examinees.

Test Security. Use of CAT-ASVAB significantly improves test security. There are no test booklets to be stolen or marked. The actual test items are stored in volatile random access memory (RAM) in the microcomputer system. This means that even if an examinee stole the computer, the items would not be compromised, as the information in volatile RAM disappears immediately when the computer is disconnected from its power source.

Motivation/Image. CAT-ASVAB offers advantages in the areas of examinee motivation and military image. Studies have shown that examinees clearly prefer taking a test on a computer to taking a P&P test. Further, the use of microcomputers in the military personnel accessioning process conveys a "high tech" image of the Services to the applicants. This image should assist military recruiters in meeting their goals.

Future Tests. A final area in which CAT-ASVAB offers significant advantages is that it provides a microcomputer-based delivery platform which can be used to administer tests that would be impossible via paper-and-pencil. An example would be a target acquisition and tracking test, which would involve dynamic test items, presented on a computer screen.

Use of the computer to administer tests also makes it possible to measure and record an examinee's response latency for each item. The speed with which an examinee responds to a test question can augment the information provided by the correct/incorrect dichotomous scoring of the item. This may enhance the predictive effectiveness of the ASVAB for some criteria.

CAT Summary

Currently, CAT-ASVAB is being operationally evaluated in five MEPSs and one METS. DoD has decided to implement CAT-ASVAB in MEPSs, and nationwide implementation in METSs is being considered. Conversion of the DoD-STP ASVAB testing to computerized delivery is in the future, if at all, because of logistical, technical, and practical problems in conducting a standardized, computer-based testing program in nearly 15,000 schools. Whatever the outcome of METS and STP implementation decisions, the CAT-ASVAB promises to be one of the largest, if not the largest, operational implementation of CAT in history.

CHAPTER SUMMARY

This chapter was designed to familiarize readers new to the ASVAB program and/or CAT with the concepts, jargon, and applications of the two major focuses of this book, making it unnecessary to redescribe the ASVAB and CAT in each of the following chapters. The 15-year research and development program that has led to CAT-ASVAB operational adoption provides a valuable history of the design, development, implementation, and evaluation of a major CAT effort. The lessons learned are documented in the forthcoming chapters, written by many of the professionals who did the work throughout the years.

Chapter 24

Transfer of CAT-ASVAB Technology

by

James R. McBride [1]

CAT-ASVAB's development cycle has been a lengthy one; from its beginnings in 1979, it has taken over 15 years to approach full-scale operational use. This slow pace of operational introduction, however, belies the pace of its technical development. CAT-ASVAB had successfully demonstrated proof of concept by 1984, when its equivalence to the printed ASVAB was first demonstrated in terms of predictive validity and construct equivalence. Although it took 12 years from that point to the start of operational implementation of CAT-ASVAB, technology developed in the course of the project has been transferred over the years to other projects which have been much quicker to reach practical use. Examples include specific commercial applications of adaptive testing, other military testing programs, and an educational testing program. In addition, key technical developments from CAT-ASVAB are at the core of another major government application of CAT. This chapter will summarize some of the applications of CAT technology that have been the direct beneficiaries of technology developed in the course of the CAT-ASVAB program.

The principal value of technology transfer is perhaps that it makes possible widespread development of practical applications of technology in far less time and expensive than the technology took to develop. Without the transfer of CAT-ASVAB technology, a number of CAT applications that have been in use for up to 10 years might not have been economically feasible. There are at least four aspects of CAT-ASVAB technology that have been either appropriated by other CAT applications, or transferred directly to them. (1) adaptive testing strategy: psychometric technology: (adaptive test design, item selection, and scoring procedures); (2) computer software; (3) equating technology: The extraordinary procedures used to equate IRT-based adaptive test scores to the traditional score metric of conventional tests; and (4) technical standards: The extension of existing professional standards for the development and use of conventional, printed tests to the special situations of computerized test administration in general and adaptive testing in particular. Examples of technology transfer in each of these four areas are presented in this chapter.

ADAPTIVE TESTING STRATEGY

In Chapter 3, I presented a definition of a "strategy" for adaptive testing: An integrated set of methods and criteria for adaptively selecting items one by one, and for placing scores from the resulting tests on the same scale. That chapter reviewed some of the features of a variety of adaptive testing strategies that have been proposed over the years, and described the strategy eventually adopted for use in CAT-ASVAB: A hybrid strategy that administers fixed-length adaptive tests employing Bayesian procedures for ability estimation, a local maximum information criterion for item selection, and a procedure for limiting test item exposure.

---

[1] Human Resources Research Organization.

CAT-ASVAB's adaptive testing strategy was adopted after extensive study of the psychometric characteristics of alternative strategies for adaptive testing, and has been demonstrated to result in efficient adaptive tests that are reliable and valid. Any test user choosing to explore or implement adaptive testing must select a strategy. In doing so, they can either conduct a research program similar to CAT-ASVAB's research into alternative strategies, or they can adopt an already-developed strategy and tailor it to their special requirements. The latter course is less time-consuming, as well as far less expensive. CAT-ASVAB developers have been generous in transferring their accumulated knowledge about various aspects of adaptive testing strategies to other prospective users of the technology; in addition, some CAT-ASVAB researchers have applied CAT-ASVAB procedures to other adaptive test programs after leaving government service.

Examples of the transfer of CAT-ASVAB's adaptive testing strategy to other programs will be given below. First, it may be useful to present a summary of some of the features of that strategy, and to differentiate it from other strategies now in use in major adaptive testing programs (e.g., the computerized adaptive versions of the Graduate Record Examination and the certification testing program of the American Board of Clinical Pathologists). Some key features that differentiate CAT-ASVAB and these programs are (1) their psychometric foundations; (2) their procedures for ability estimation; (3) their criteria for adaptive item selection; and (4) their criteria for test termination.

All of these programs use item response theory (IRT) as a general psychometric foundation. CAT-ASVAB uses the 3-parameter logistic IRT model, as does the GRE programs; the Clinical Pathologists program, in contrast, uses the 1-parameter logistic, also known as the Rasch model. These programs use a wider variety of ability estimation procedures: CAT-ASVAB is unique in this aspect of its overall strategy. It uses Owen's Bayesian sequential procedure for updating the ability estimate after each test item. Then, after the last item in each test, CAT-ASVAB computes a final ability estimate, using Bayesian modal estimation. The GRE uses maximum likelihood estimation to update the ability estimate after each item, and at the end of the test. The Clinical Pathologists program uses Rasch estimation, which in effect is a special case of maximum likelihood estimation.

In their adaptive item selection procedures, CAT-ASVAB and the GRE are similar. Both select items by referring to a pre-computed lookup table in which items are sorted in descending order of their information values at spaced intervals over the ability scale. This is referred to as a "maximum information" item selection criterion. Both programs have modified the maximum procedure somewhat to balance item usage, and thus avoid over-exposure of the most informative test items. Because the Clinical Pathologists testing program uses the Rasch model, it can select items on the basis of the proximity of the item difficulty parameter to the most recent estimate of examinee ability; this is tantamount to the maximum information criterion, but is implemented in a totally different way.

The technology embodied in CAT-ASVAB's hybrid Bayesian sequential adaptive testing strategy has been transferred to a number of other adaptive tests, both within and outside of the federal government. Ironically, although each of the examples presented here is a direct descendant of CAT-ASVAB research and development, each went into practical use years before CAT-ASVAB itself.

The first widespread practical use of adaptive testing was the Army's Computerized Adaptive Screening Test, (CAST), which is available to recruiters to evaluate the likelihood that a prospective recruit will attain a qualifying score on the Armed Forces Qualification Test embedded in the ASVAB. CAST was introduced into operational use in 1985. Its development is described in some detail in Chapter 6. Suffice it to say here that CAST represented the first instance of CAT-ASVAB technology transfer. CAST, which was developed for the Army by the Navy Personnel Research and Development Center (NPRDC), is based entirely on procedures and materials pioneered in the course of CAT-ASVAB research and development. CAST's adaptive testing strategy is identical to the hybrid Bayesian sequential strategy developed for CAT-ASVAB (and reported in Wetzel and McBride, 1986). CAST's item banks were developed in early CAT-ASVAB research reported by Moreno, Wetzel, McBride and Weiss (1983). Decisions about the composition and length of the CAST tests were also based on data reported by Moreno et al. (1983).

One of the first examples of a commercial application of CAT is the Computerized Adaptive Edition of the Differential Aptitude Tests -- the Adaptive DAT -- published by The Psychological Corporation (1986). The printed versions of the DAT have been used to test millions of people since 1947, for educational placement and vocational

guidance of secondary school students, and for personnel selection and career counseling of adults. The Adaptive DAT, like the CAT-ASVAB, is a system for computerized adaptive administration of a traditional multiple abilities battery. Also like the CAT-ASVAB, the Adaptive DAT takes less than half the time it takes to administer the printed version.

The linkage of the Adaptive DAT to military CAT research is very direct: From 1977 to 1983 this writer was principal investigator in the Navy's development of CAT-ASVAB. From 1985 to 1986, I directed the development of the Adaptive DAT, which uses many of the same psychometric procedures pioneered within the Department of Defense, including the hybrid adaptive strategy based on Bayesian sequential ability estimation. (Another CAT system developed by The Psychological Corporation is the Stanford Adaptive Mathematics Screening Test, a brief test of achievement in mathematics that is suitable for use over an extremely wide range of ability -- from fourth through twelfth grade. Like the Adaptive DAT, it uses the hybrid Bayesian strategy. Unlike any other adaptive test I am aware of, it also employs "differential entry levels." Initial ability estimates and difficulty levels vary depending on school grade; thus, the Stanford Adaptive Mathematics Screening Test was the first operational adaptive test to use collateral information -- in this case, school grade -- to guide ability estimation and item selection.

ADAPTIVE TESTING SOFTWARE

Just as the evolution of strategies for adaptive testing was slow and expensive, so was the development of software systems for CAT. The earliest CAT software, developed under the direction of Abraham Bayroff of the Army Research Institute, was very limited in its application. His first system administered adaptive tests via a teletype machine. It was inherently limited to tests that could be presented in printed form, using only numbers, common typographic symbols, and upper-case alphabetic characters. His second system was far more advanced in display capability -- test items were presented by projecting 35mm color transparencies on a small screen, and thus could contain anything that could be photographed. Both of Bayroff's systems were developed for use on mainframe computers, and for research purposes only; they were not easily extended to other applications, and their inherent limitations did not make them attractive candidates for adoption elsewhere.

Starting in the 1970s with the burgeoning availability of minicomputers that could support multiple users simultaneously, and then of microcomputer networks that made it feasible to test each examinee at a dedicated computer, the development of systems for adaptive testing became feasible. Feasibility is one thing; practicality is another. The first general-purpose software systems capable of administering batteries of adaptive tests, and of displaying graphical as well as text-only test items typically took two years or more to design and develop. The cost of software development was commensurate with the time involved. With few exceptions, each early CAT researcher developed a new software system for CAT administration. At first, this was essential, because of the evolution of CAT strategies themselves, and because of the rapid changes that were occurring in computer technology. In time, however, it became feasible to develop flexible, general-purpose systems for administering CAT and other computer-based tests. Once that point was reached, the transfer of DoD-developed CAT software technology to more general use began.

One of the first vehicles of this transfer was MicroCAT (Assessment Systems Corporation, 1984). MicroCAT is an integrated system for both development and administration of tests, including but not limited to, computer administered adaptive tests. It was the first commercially available system for designing, authoring, analyzing and administering adaptive tests. It's the kind of thing that an adaptive test developer would have to invent if it were not itself commercially available. MicroCAT was developed by David Vale of Assessment Systems Corporation under a Navy Small Business Innovations Research contract. Vale had learned his craft under David Weiss at the University of Minnesota. In fact, Weiss was a principal in Assessment Systems Corporation. Hence the link of MicroCAT to military research is a direct one involving both technology and people, and its commercial availability represents the first tangible transfer of adaptive testing software technology from DoD to public use.

Other instances of the transfer of adaptive testing software outside DoD have occurred subsequently. One such transfer took the form of publishing CAT software in the public domain. The entirety of the Navy's experimental adaptive testing software system -- including source code and system documentation -- was published in an NPRDC technical report by Quan, Park, Sandahl and Wolfe (1984). This publication made it possible for the public to obtain, and use without charge, software implementations of all or part of a computer adaptive testing system, including provisions for test item bank storage and retrieval, text and graphic item design and display, examinee response processing, and features specific to adaptive testing, such as dynamic selection of test items, ability estimation, test scoring, and storage of detailed data for each test administered.

More recently, DoD and the Navy have made portions of the CAT-ASVAB software system available to other users, both within and outside the federal government. For example, some CAT-ASVAB software was incorporated into a system developed by the U.S. Department of Labor (DoL) to administer a computerized version of the General Aptitude Test Battery (GATB). Additionally, the Navy has made its software available for administration of other agencies' computerized adaptive tests, and has provided technical support in adapting the software for those agencies' use. An adaptive testing system under development for administering personnel tests for the U.S. Immigration and Naturalization Service is based entirely on the CAT-ASVAB software platform. Additionally, the NPRDC performed a similar adaptation of the CAT-ASVAB software system for use in an experimental educational test administration system developed for the North Carolina State Department of Education.

ADAPTIVE TEST EQUATING METHODS

Among the thorniest technical challenges to the developers of CAT-ASVAB was the problem of test equating. The problem itself is straightforward: For some period of time after implementation of CAT-ASVAB, adaptive and conventional versions of the battery will be in use at the same time. Consequently, scores from both versions of the battery must be interchangeable. Adaptive tests, however, use a different score metric than conventional tests (IRT continuous ability metric rather than number correct scores), and typically have different degrees of measurement precision. Methods used to equate alternate forms of conventional tests were not applicable to the problem of equating adaptive and conventional test scores. Segall discusses this problem -- and CAT-ASVAB's solution to it -- in detail in Chapter 18. Solving the equating problem was essential, not only in the case of CAT-ASVAB but also for any other adaptive test developed to be used interchangeably with a conventional test. DoD and the Navy have published their equating technology, and have made it, and the expertise of its developers, available to other organizations faced with analogous situations.

The first example of this is the DoL's computerized GATB program, mentioned above. The computerized version of GATB contains both adaptive and conventional versions of many of the printed GATB tests. The computerized paper-and-pencil GATB tests are speeded by design, and measurement differences between printed and computerized implementations of speeded tests are well-documented (e.g., Greaud & Green, 1986). The adaptive versions of some of the GATB tests represent a particular challenge, as they are designed as power tests yet are to be used interchangeably with counterpart printed tests which are speeded. Through a cooperative arrangement between the U.S.

DoD and DoL, NPRDC staff involved in equating CAT-ASVAB with its printed counterpart have taken responsibility for equating the new and old versions of GATB as well. That effort is incomplete at this writing, pending collection of printed and computer-administered GATB data by DoL.

ADAPTIVE TESTING STANDARDS

Another difficult issue arose early in the development of CAT-ASVAB as an alternative to, and replacement for, the P&P-ASVAB: The absence of precedents. The CAT-ASVAB program began in 1979; by 1982, enough research

data had been accumulated to indicate that adaptive versions of some of the ASVAB tests were highly correlated with their conventional counterparts, and were much more efficient. While the early data were promising for the new technology, there were no professional standards or guidelines available to evaluate the suitability of computerized tests in general, and adaptive tests in particular, as replacements for conventional tests in ongoing testing programs.

Even though the early results were promising from a research standpoint, it was not clear what kind and amount of evidence would be required to support the use of CAT-ASVAB as a replacement for the traditional version used for DoD enlisted personnel selection. There were many unanswered questions: Would the computerized adaptive tests measure the same ability constructs as the conventional tests?, what about the speeded tests?, Would they be as reliable and valid for personnel selection?, What evidence would be needed to answer the preceding questions in the affirmative?, Would computerized test administration give examinees with computer experience an advantage over others?, and Would computerized tests put some population subgroups -- such as males, females, majority or minority group members -- at an advantage or disadvantage? Existing professional standards, particularly the then current *Standards for Educational and Psychological Tests* (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1974) neither addressed, nor anticipated, the use of computers or adaptive testing in test administration.

The absence of applicable standards was a matter of some concern; among other things, it left open the possibility that computerized adaptive testing might be technically attractive yet unacceptable on legal or other grounds. To address the absence of standards, Dr. Charles Davis of the Office of Naval Research arranged for a panel of experts, independent of the DoD, to study the matter and develop a set of technical recommendations on the kinds of research needed to evaluate the suitability and technical acceptability of a computerized adaptive version of the ASVAB. The evaluation plan proposed by that panel (Green, Bock, Humphreys, Linn & Reckase, 1982) constituted what may be the most rigorous standards ever imposed on a psychometric test development project. Its contents were transferred to the public domain by the subsequent publication of an article in the *Journal of Educational Measurement* (Green et al., 1984) that applied similar evaluation standards to CAT in general. The contents of the evaluation plan also influenced the 1985 revision (American Psychological Association, 1985) of the 1974 test standards, as well as the later *Guidelines for Computer-Based Tests and Interpretations* (American Psychological Association, 1986).

SUMMARY

As the examples given above indicate, long before its recent introduction into large-scale operational use, CAT-ASVAB had a profound impact on other practical applications of CAT programs. The psychometric testing strategy developed in the early 1980s for the CAT-ASVAB system has been incorporated in a number of other adaptive tests, beginning as early as 1985. Software developed for CAT-ASVAB has been incorporated, in whole or in part, into a number of other public sector adaptive testing systems; some CAT-ASVAB software has also been published in the public domain, and is potentially available for all to use. Technology initially developed for equating CAT-ASVAB test scores to scores of counterpart conventional ASVAB tests has been extended for use in other programs, notably the DoL's development of a computerized version of the GATB. Finally, and perhaps most important of all, technical standards developed specifically to guide the evaluation of CAT-ASVAB as a potential replacement for its conventional version have become de facto professional standards for evaluating any CAT system.

CONCLUSION

There is little doubt that computerized testing, in general, and CAT, in particular, is poised for a broad technology transfer to the entire spectrum of testing -- cognitive testing, surveys and polling, personality measurement, clinical diagnosis, and myriad other applications. Government, industry, and academia all are carefully venturing into the

unfamiliar waters. The 20-year research and development sponsored and conducted by the Military Services will provide the foundation upon which that technology is built.

Psychological testing has finally reached the point predicted over a quarter of a century ago by Dr. Bert Green: *"most of these changes lie in the future....in the inevitable computer conquest of testing."* (Green, 1970).

CONSOLIDATED REFERENCE LIST

Ackerman, T.A. (1985). "An investigation of the effect of administering test items via computer." Paper presented at the Midwest Educational Research Association, Chicago.

Alderton, D.L., & Larson, G.E. (1992). *ECAT battery: Descriptions, constructs, and factor structure.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations.* Washington, DC: Author.

American Psychological Association. (1980). *Principles for the validation and use of personnel selection procedures.* Division of Industrial-Organizational Psychology. Berkeley, CA: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests.* Washington, DC: American Psychological Association.

Angoff, W.H. (1971). "Norms, scales, and equivalent scores." In R.L. Thorndike (Ed.), *Educational measurement (2nd ed.).* Washington, DC: American Council on Education.

Armor, D.J., Fernandez, R.L., Bers, K., & Schwarzbach, D. (1982). *Recruit aptitudes and Army job performance: Setting enlistment standards for Infantrymen* (R-2874-MRAL). Santa Monica, CA: The RAND Corporation.

Assessment Systems Corporation. (1984). *User's manual for the MicroCAT testing system* (RR 85-1*)*. St. Paul, MN: Author.

ASVAB Working Group. (1980). *History of the Armed Services Vocational Aptitude Battery (ASVAB) 1974-1980.* A report to the Principal Deputy Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics). Washington, DC: Author.

References

Automated Sciences Group & CACI, Inc.-Federal. (1988). *CAT-ASVAB program: Concept of operation and cost/benefit analysis*. Fairfax, VA: Author.

Baker, H.G. (1985). *NPRDC research and development in support of recruiting.* Paper presented at the Army Recruiting Research Coordination Conference, Northbrook, IL.

Baker, H.G. (1983a). *Navy Personnel Accessioning System (NPAS): Background and overview of the person-job matching (PJM) and recruiting management support (RMS) subsystems* (SR 83-34). San Diego: Navy Personnel Research and Development Center.

Baker, H.G. (1983b). *Navy Personnel Accessioning System (NPAS): II. Summary of research and development efforts and products* (SR 83-35). San Diego: Navy Personnel Research and Development Center.

Baker, H.G., Rafacz, B.A., & Sands, W.A. (1984). *Computerized Adaptive Screening Test (CAST): Development for use in military recruiting stations* (NPRDC TR 84-17). San Diego: Navy Personnel Research and Development Center.

Baker, H.G., Rafacz, B.A., & Sands, W.A. (1983). *Navy Personnel Accessioning System: III. Development of a microcomputer demonstration system* (SR 83-36). San Diego: Navy Personnel Research and Development Center.

Baker, M. S. (1983c). *Predictors of performance in Navy electronics skills: The effects of mathematical skills* (NPRDC TR-83-10). San Diego: Navy Personnel Research and Development Center.

Bejar, I.I., & Weiss, D.J. (1978). *A construct validation of adaptive achievement testing* (RR 78-4). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Betz, N.E., & Weiss, D.J. (1975). *Empirical and simulation studies of flexilevel ability testing* (RR 75-3). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Betz, N.E., & Weiss, D.J. (1974). *Simulation studies of two-stage ability testing* (RR 74-4). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Betz, N.E., & Weiss, D.J. (1973). *An empirical study of computer-administered two-stage ability* testing (RR 73-4). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Birnbaum, A. (1968). "Some latent trait models and their use in inferring an examinee's ability." In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bloxom, B.M., McCully, R., Branch, R., Waters, B.K., Barnes, J.D., & Gribben, M.R. (1993). *Operational calibration of the circular-response optical-mark-reader answer sheets for the ASVAB (*Report 93-009*)*. Monterey, CA: Defense Manpower Data Center.

Bock, R.D. (1972). "Estimating item parameters and latent ability when responses are scored in two or more nominal categories." *Psychometrika, 37,* 29-52.

Bock, R.D., & Aitkin, M. (1981). "Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm." *Psychometrika, 46,* 443-459.

Bock, R.D., & Mislevy, R.J. (1981). "Adaptive EAP estimation of ability in a microcomputer environment." *Applied Psychological Measurement, 6,* 431-444.

Bodzin, L.J. (1986). *An image capturing and editing system for the HP-Integral computer.* Unpublished paper. San Diego: Naval Ocean Systems Command.

Booth-Kewley, S. (1983). *Validation of the Armed Services Vocational Aptitude Battery Forms 5/6/7 and 8/9/10 selection criteria for strategic weapons system electronic "A" school.* San Diego: Navy Personnel Research and Development Center.

Branch, Rick. (March 7, 1995). Personal communication. North Chicago, IL: U.S. Military Entrance Processing Command.

Brogden, H.E. (1977). "The Rasch model, the law of corporate judgment, and additive conjoint measurement." *Psychometrika, 42,* 631-635.

Burke, M.J., Michael, & Normand, J. (1986). *Examinee attitudes toward computer-administered psychological testing.* Paper presented at the American Educational Research Association, San Francisco.

References

Carey, N.B. (1994). "Computer predictors of mechanical job performance: Marine Corps findings." *Military Psychology, 6,* 1-30.

Clark, C.L. (1975). *Proceedings of the First Conference on Computerized Adaptive Testing* (PS 75-6). Washington, DC: Personnel Research and Development Center, U.S. Civil Service Commission.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum, Associates.

Congressional Budget Office. (1980). *Costs of manning the active duty military* (Staff working paper). Washington, DC: Author.

Crichton, L.I. (1981).*Effects of error in item parameter estimates on adaptive testing.* Unpublished doctoral dissertation, University of Minnesota.

Croll, P.R. (1982). *Computerized adaptive testing system design: Preliminary design considerations* (NPRDC TR 82-52). San Diego: Navy Personnel Research and Development Center.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, IL: University of Illinois Press.

Cudeck, R. (1985). "A structural comparison of conventional and adaptive versions of the ASVAB." *Multivariate Behavioral Research, 20,* 305-322.

Datacopy Corporation. (1985a). *Datacopy Model 700 User's Guide (Version 1.4).*

Datacopy Corporation. (1985b). *Datacopy WIPS Editor User's Guide (Version 1.0).*

Department of the Army. (1965). *Marginal man and military service.* Washington, DC: Author.

Department of Defense. (1992). *ASVAB 18/19 counselor manual.* North Chicago, IL: U.S. Military Entrance Processing Command.

Department of Defense. (1985). *Defense manpower quality: Volume I.* Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Installations and Logistics).

Department of Defense. (1982). *Profile of American Youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery.* Washington, DC:Office of the Assistant Secretary of Defense (Manpower, Installations and Logistics).

Divgi, D.R. (1990). *Calculating the performance gain due to improved validity* (CRM 89-254). Alexandria, VA: Center for Naval Analyses.

Divgi, D.R. (1986). *On some issues in the Accelerated CAT-ASVAB Project* (CRM 86-231). Alexandria, VA: Center for Naval Analyses.

Divgi, D.R., & Stoloff, P.H. (1986). *Effect of the medium of administration on ASVAB item response curves* (Report 86-24). Alexandria, VA: Center for Naval Analyses.

Dorans, N.J., & Kingston, N.M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. *Journal of Educational Measurement, 22,* 249-262.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley and Sons.

Drasgow, F., & Parsons, C.K. (1983). "Application of unidimensional item response theory models to multidimensional data." *Applied Psychological Measurement, 7,* 189-199.

Eitelberg, M.J. (1988). *Manpower for military occupations.* Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Eitelberg, M.J., Laurence, J.H., Waters, B.K., with Perelman, L.S. (1984). *Screening for service: Aptitude and education criteria for military entry.* Washington, DC: Office of Assistant Secretary of Defense (Manpower, Installations and Logistics).

Ezekiel, M. (1930). *Methods of correlational analysis.* New York: John Wiley and Sons.

Fairbank, B.A. (1987). "The use of presmoothing and postsmoothing to increase the precision of equipercentile equating." *Applied Psychological Measurement, 11,* 245-262.

Folchi, J.S. (1986). "Communication of computerized adaptive testing results in support of ACAP." *Proceedings of the Annual Conference of the Military Testing Association, 28,* 618-623.

References

Frankel, S. (1986). *Computer adaptive tests: The engines that will drive the revolution in education.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Garrison, W.M., & Baumgarten, B.S. (1986). An application of computer adaptive testing with communication handicapped examinees. *Educational and Psychological Measurement, 46,* 23-35.

Grafton, F., Czarnolewski, M.Y., & Smith, E.P. (1989). *Relationship between Project A psychomotor and spatial tests and TOW2 gunnery performance A preliminary investigation* (ARI-WP-RS-89-1). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Greaud, V.A., & Green, B.F., Jr. (1986). "Equivalence of conventional and computer presentation of speed tests." *Applied Psychological Measurement, 10,* 23-24 .

Green, B.F., Jr. (1983). "The promise of tailored tests." In H. Wainer & S.A. Messick (Eds.), *Principles of modern psychological measurement: A Festshrift in honor of Frederic Lord.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Green, B.F., Jr. (1970). "Comments on tailored testing." In W.H. Holtzman (Ed.), *Computer-assisted instruction, testing, and guidance.* New York: Harper & Row. 194.

Green, B.F., Jr., Bock R.D., Humphreys, L.G., Linn, R.L, & Reckase, M.D. (1984). "Technical guidelines for assessing computerized adaptive tests". *Journal of Educational Measurement,* 347-360.

Green, B.F., Jr., Bock R.D., Humphreys, L.G., Linn, R.L, & Reckase, M.D. (1982). *Evaluation plan for the Computerized Adaptive Armed Services Vocational Aptitude Battery. Baltimore, MD:*: The Johns Hopkins University, Department of Psychology.

Greenberg, I.M. (1980). *Mental standards for enlistment performance of Army personnel related to AFQT/ASVAB scores* (Final report MGA-0180-WRO-02). Monterey, CA: McFann, Gray and Associates.

Hardwicke, S.B., Eastman, L., & Cooper, R. (1984). *Computerized adaptive testing (CAT) A user's manual* (NPRDC TR 84-32). San Diego: Navy Personnel Research and Development Center.

Hardwicke, S.B., Vicino, F.L., McBride, J.R., & Nemeth, C. (1984). *Computer adaptive testing of the Armed Services Vocational Aptitude Battery.* San Diego, CA: Navy Personnel Research and Development Center.

Hardwicke, S.B., & White, K.D. (1983). *Predictive utility evaluation of computerized adaptive testing.* (Contract N00123-8). San Diego:Navy Personnel Research and Development Center.

Hedges, L.V., Becker, B.J., & Wolfe, J.H. (1992). *Detecting and measuring improvements in validity* (TR-93-2). San Diego: Navy Personnel Research and Development Center.

Hedl, J.J., O'Neill, H.L., & Hansen, D.N. (1973). Affective reactions toward computer-based intelligence testing. *Journal of Counseling and Clinical Psychology, 40,* 217-222.

Henley, S.J., Klebe, K.J., McBride, J.R., & Cudeck, R. (1990). "Adaptive and conventional versions of the DAT: The first complete test comparison." *Applied Psychological Measurement, 13,* 463-471.

Hetter, R.D., & Segall, D.O. (1986). "Relative precision of paper-and-pencil and computerized adaptive tests." In *Proceedings of the 1986 Annual Conference of the Military Testing Association.* Mystic, CT.

Hetter, R.D., Segall, D.O., & Bloxom, B.M. (1994). "Evaluating item calibration mode in computerized adaptive testing." *Applied Psychological Measurement, 18 (3),* 197-204.

Hogan, P.F., McBride, J.R., & Curran, L.T. (1995). *An evaluation of alternate concepts for administering the Armed Services Vocational Aptitude Battery to applicants for enlistment* (DMDC TR 95-013). Monterey, CA: Personnel Testing Division, Defense Manpower Data Center.

Holland, J.L. (1973). *Making vocational choices:A theory of careers.* Englewood Cliffs, NJ: Prentice-Hall.

Hom, I. (1994). *PC-CAT Examinee Testing Station Software.* San Diego: RGI.

Hornke, L.F., & Sauter, M.F. (1980). "A validity study of an adaptive test of reading comprehension." In D.J. Weiss, (Ed.), *Proceedings of the 1979 Computerized Adaptive Testing Conference.* Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota, 57-67.

References

Hotelling, H. (1931). "A generalization of Student's ratio." *Annual of Mathematical Statistics, 2,* 360-378.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory.* Homewood, IL: Dow Jones-Irwin.

Jensema, C.J. (1977). "Bayesian tailored testing and the influence of item bank characteristics." *Applied Psychological Measurement,* 111-120.

Jensema, C.J. (1975). *Bayesian tailored testing and the influence of item bank characteristics.* Paper presented at the Conference on Computerized Adaptive Testing. Washington, DC: Personnel Research and Development Center, U. S. Civil Service Commission, 82-89.

Jensema, C.J. (1974a). "An application of latent trait mental test theory." *British Journal of Mathematical Statistical Psychology, 27,* 29-48.

Jensema, C.J. (1974b). "The validity of Bayesian tailored testing." *Educational and Psychological Measurement, 34,* 757-766.

Jensema, C.J. (1972). *An application of latent trait mental test theory to the Washington Precollege Testing Battery.* Doctoral thesis, University of Washington. (University Microfilms 72-20,871, Ann Arbor, MI).

Jensen, H.E., & Valentine, L.D. (1976). *Development of the Enlistment Screening Test (EST) Forms 5 and 6* (TR 76-42). Brooks AFB, TX: Air Force Human Resources Laboratory.

Johnson, M.F., & Weiss, D.J. (1979). "Parallel forms reliability and measurement accuracy of adaptive and conventional testing strategies." In D.J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference.* Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Joint Chiefs of Staff. (1982). *United States military posture for FY83.* Washington, DC: Author.

Jöreskog, K.G., & Sorbom, D. (1984). *LISREL VI, Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods.* Mooresville, IN: Scientific Software.

Kass, R.A., Mitchell, K.J., Grafton, F.C., & Wing, H. (1983). "Factorial validity of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10: 1981 Army applicant sample." *Educational and Psychological Measurement, 43,* 1077 -1087.

Kendall, M.G., & Stuart, A. (1977). *The advanced theory of statistics, Section 10-6.* New York: Hafner.

Kershaw, S.W., & Wainer, H. (1985). "Reviewing an item pool for its sensitivity to the concerns of women and minorities: Process and outcomes." *Proceedings of the Annual Conference of the Military Testing Association,* 288-293. San Diego: Military Testing Association.

Kieckhaefer, W.F., Moreno, K.E., & Segall, D.O. (1985). "Medium of administration effects on attitude toward ASVAB testing." *Proceedings of the Annual Conference of the Military Testing Association.* San Diego: Military Testing Association.

Kieckhaefer, W.F., Ward, D.G., Kusulas, J.W., Cole, D.R., Rupp, L.M., & May, M.H. (1992). *Criterion development for 18 technical training schools in the Navy, Army, and Air Force* (Contract N66001-90-D-9502, DO 7J08). San Diego: Navy Personnel Research and Development Center.

Kingsbury, G.G., & Weiss, D.J. (1981). *A validity comparison of adaptive and conventional strategies for mastery testing* (Report 81-3). Minneapolis: Department of Psychology, University of Minnesota.

Knapp, D.J. (1987a). *Final report on a national cross-validation of the Computerized Adaptive Screening Test (CAST)* (ARI Selection and Classification Technical Area Working Paper 87-05). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J. (1987b). *Display of results: Alternatives for the computerized adaptive screening test (CAST)* (ARI TR 768). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Pliske, R.M. (1986a). *An update on the Computerized Adaptive Screening Test (CAST).* Paper presented at the annual meeting of the Military Testing Association. Mystic, CT: Military Testing Association.

References

Knapp, D.J., & Pliske, R.M. (1986b). *Preliminary report on a national cross-validation of the Computerized Adaptive Screening Test (CAST)* (ARI RR 1430). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kolen, M.J. (1991). "Smoothing methods for estimating test score distributions." *Journal of Educational Measurement, 28,* 257-282.

Kronmal, R., & Tarter, M. (1968). "The estimation of probability density and cumulatives by Fourier series methods." *Journal of the American Statistical Association, 69,* 925-952.

Larkin, K.C., & Weiss, D.J. (1975). *An empirical comparison of two-stage and pyramidal adaptive ability testing* (RR 75-1). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Larkin, K.C., & Weiss, D.J. (1974). *An empirical investigation of computer-administered pyramidal ability testing* (RR 74-3). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota. (AD 783 553)

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. *Royal Society of Edinburgh Proceedings, Section A, 62,* 28-30.

Lord, F.M. (1980a). *Application of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F.M. (1980b). "Small N justifies Rasch methods." In D.J. Weiss (Ed.) *Proceedings of 1979 Computerized Adaptive Testing Conference.* Minneapolis: Computer Adaptive Testing Laboratory, Psychometric Methods Program, Department of Psychology, University of Minnesota.

Lord, F.M. (1979). "Discussant remarks." In D.J. Weiss (Ed.) *Proceedings of the 1979 Computerized Adaptive Testing Conference,* 439-441. Minneapolis: Computer Adaptive Testing Laboratory, Psychometric Methods Program, Department of Psychology, University of Minnesota.

Lord, F.M. (1977). "A broad range tailored test of verbal ability." *Applied Psychological Measurement,* 95-100.

Lord, F.M. (1974). "Individualized testing and item characteristic curve theory." In Krantz, Atkinson, Luce, & Suppes, *Contemporary developments in mathematical psychology.* San Francisco: W.H. Freeman.

Lord, F.M. (1971a). "The self-scoring flexilevel test." *Journal of Educational Measurement, 8,* 147-151.

Lord, F.M. (1971b). "A theoretical study of two-stage testing." *Psychometrika, 36,* 227-241.

Lord, F.M. (1971c). "Robbins-Munro procedures for tailored testing." *Educational and Psychological Measurement, 31,* 3-31.

Lord, F.M. (1970). "Some test theory for tailored testing." In W. Holtzman (Ed.), *Computer-assisted instruction, testing, and guidance.* New York: Harper & Row.

Lord, F.M. (1952). *A theory of test scores.* Psychometric Monograph No. 7. Princeton: Educational Testing Service.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lukin, M.E., Dowd, T., Plake, B., & Kraft, R.G. (1985). "Comparing computerized versus traditional psychological assessment." *Computer in Human Behaviors, 1,* 49-58.

Maier, M.H. (1983). *The predictive validity of the AFQT for forms 8, 9, and 10 of the ASVAB* (83-3163/27). Alexandria, VA: Center for Naval Analyses.

Maier, M.H., & Truss, A.R. (1983). *Validity of the ASVAB Forms 8, 9, and 10 for Marine Corps training courses: Subtests and current composites* (83-3107/1). Alexandria, VA: Center for Naval Analyses.

Martin, J.T., McBride, J.R., & Weiss, D.J. (1983). *Reliability and validity of adaptive tests in a military recruit population* (ONR TR 83-1). Arlington, VA: Personnel and Training Programs, Office of Naval Research.

Mathews, J.J., & Ree, M.J. (1982). *Enlistment Screening Test Forms 81a and 81b: Development and calibration* (AFHRL TR 81-54). Brooks AFB, TX: Air Force Human Resources Laboratory.

References

McBride, J.R. (in press). "Innovations in computer-based ability testing: Promise, problems, and peril." In M.D. Hakel, (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection.* Hillsdale, NJ: Lawrence Erlbaum Associates.

McBride, J.R. (1982). *Computerized adaptive testing project: Objectives and requirements* (NPRDC TN 82-22). San Diego: Navy Personnel Research and Development Center.

McBride, J.R. (1980). "Adaptive verbal ability testing in a military setting." In D.J. Weiss (Ed.), *Proceedings of the 1979 Computerized Adaptive Testing Conference* Minneapolis: Department of Psychology, University of Minnesota.

McBride, J.R. (1979). *Adaptive mental testing: The state of the art* (ARI Report 423). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (NTIS 088000)

McBride, J.R. (1976a). *Research on adaptive testing 1973-1976: A review of the literature.* Unpublished manuscript. Department of Psychology, University of Minnesota.

McBride, J.R. (1976b). *Simulation studies of adaptive testing: A comparative evaluation.* Unpublished doctoral dissertation, University of Minnesota.

McBride, J.R. (1975). "Scoring adaptive tests." In D.J. Weiss (Ed.), *Computerized adaptive trait measurement -- problems and prospects*. Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

McBride, J.R., Corpe, V.W., & Wing, H. (1987). *Equating the computerized adaptive edition of the Differential Aptitude Tests*.. Paper presented at the annual convention of the American Psychological Association, New York.

McBride, J.R., & Martin, J.T. (1983). "Reliability and validity of adaptive verbal ability tests in a military setting." In D.J. Weiss, (Ed.), *New horizons in testing,* 223-235. New York: Academic Press.

McBride, J.R., & Sympson, J.B. (1982). "The computerized adaptive testing system development project." *Proceedings of the 1982 Item Response Theory and Computerized Adaptive Testing Conference.* Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

McBride, J.R., & Weiss, D.J. (1976). *Some properties of a Bayesian adaptive ability testing strategy* (RR 76-4). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

McHenry, J.J., Hough, L.M., Toquam, J.L., Hanson, M.A., & Ashworth, S. (1990). "Project A validity results: The relationship between predictor and criterion domains." *Personnel Psychology, 43,* 335-354.

Mead, A.D., & Drasgow, F. (1993). "Effects of administration medium: A meta-analysis." *Psychological Bulletin, 114 (3),* 449-458.

Mislevy, R.J., & Bock, R.D. (1981). *BILOG--Maximum likelihood item analysis and test scoring: LOGISTIC model* Chicago: International Educational Services.

Mitchell, P., Hardwicke, S.B., Segall, D.O., & Vicino, F.L. (1983). *Computerized adaptive testing:A preliminary study of user acceptability.* Paper presented at the Annual Conference of the Military Testing Association. Gulf Shores, AL: Military Testing Association.

Moe, K.C., & Johnson, M.F. (1986). *Participants reaction to computerized testing.* Paper presented at the American Psychological Association, Washington, DC.

Monzon, R.I., Shamieh, E.W., & Segall, D.O. (1990). *Subgroup differences in equipercentile equating of the Armed Services Vocational Aptitude Battery: Development of an adaptive item pool.* NPRDC Technical Report. San Diego: Navy Personel Research and Development Center.

Moreno, K.E., Segall, D.O., & Kieckhaefer, W.F. (1985). "A validity study of the Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery." *Proceedings of the Annual Conference of the Military Testing Association,* 29-33. San Diego: Military Testing Association.

Moreno, K.E., Wetzel, C.D., McBride, J.R., & Weiss, D.J. (1984). "Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and Computerized Adaptive Testing (CAT) subtests." *Applied Psychological Measurement, 8 (2),* 155-163.

Moreno, K.E., Wetzel, C.D., McBride, J.R., & Weiss, D.J. (1983). *Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and computerized*

adaptive testing (CAT) subtests (NPRDC TR 83-27). San Diego: Navy Personnel Research and Development Center. (NTIS ADA131683)

Mullins, C.J., Earles, J.A., & Ree, M. (1981). *Weighting of aptitude components based on differences in technical school difficulty* (AFHRL TR-81-19). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Muraki, E. (1984). *Implementing full-information factor analysis:TESTFACT program.* A paper presented at the annual meeting of Psychometric Society, University of California, Santa Barbara.

Nellis, J., Carlson, F.R., Gray, P., Hayes, J., Holman, M., & White, M.J. (1980). *Technology assessment of personnel computers.* Los Angeles: Center for Future Research, University of Southern California.

Olivier, P.A. (1974). *An evaluation of the self-scoring flexilevel tailored testing model.* Unpublished doctoral dissertation, Florida State University.

Owen, R.J. (1975). "A Bayesian sequential procedure for quantal response in the context of adaptive mental testing." *Journal of the American Statistical Association, 70,* 351 - 356.

Owen, R.J. (1969). *A Bayesian approach to tailored testing* (RB-69-92). Princeton, NJ: Educational Testing Service.

Palmer, D.R., & Busciglio, H.H. (1996). "Coaching on the ASVAB: Analysis of post-test questionnaire responses." *Military Psychology, 8 (4)* __-__.

Park, R.K., & Dunn, M.L. (1991). *Compatibility evaluation and research on the Computerized Adaptive Screening Test (CAST). Final Report: User and programming guide.* Washington, DC: American Institutes for Research.

Park, R.K., & Rosse, R.L. (1991). *Functional requirements of an automated data collection system for the Computerized Adaptive Screening Test (CAST).* Washington, DC: American Institutes for Research.

Parker, S.B., & McBride, J.R. (1990). *A comparison of Rasch and three-parameter logistic models in computerized adaptive testing.* Unpublished manuscript.

Peterson, N.G., Hough, L.M., Dunnette, M.D., Rosse, R.L., Houston, J.S., & Toquam, J.L. (1990). "Project A: Specification of the predictor domain and development of new selection/classification tests." *Personnel Psychology, 43,* 247-276.

Pliske, R.M., Gade, P.A., & Johnson, R.M. (1984). "Cross-validation of the Computerized Adaptive Screening Test (CAST)." *Proceedings of the Annual Conference of the Military Testing Association, I,* 333-338. Munich, Federal Republic of Germany: Psychological Service of the German Federal Armed Forces.

Prestwood, J.S., & Vale, C.D. (1984). *Development of an adaptive item pool for the ASVAB.* Brooks AFB, TX: Air Force Human Resources Laboratory.

Prestwood, J.S., Vale, C.D., Massey, R.H., & Welsh, J.R. (1985). *Armed Services Vocational Aptitude Battery. Development of an adaptive item pool* (TR 85-19). Brooks AFB, TX: Air Force Human Resources Laboratory.

The Psychological Corporation. (1986*). The Stanford Adaptive Mathematics Screening Test.* San Antonio, TX: Author.

The Psychological Corporation. (1986*). The Computerized Adaptive Edition of the Differential Aptitude Tests*. San Antonio, TX: Author.

Quan, B., Park, T.A., Sandahl, G., & Wolfe, J.H. (1984). *Microcomputer network for computerized adaptive testing* (CAT) (NPRDC TR 84-33). San Diego: Navy Personnel Research and Development Center.

Rafacz, B. A. (1995). *Computerized Adaptive Testing version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB): Computer system development* (NPRDC-TN-95-8). San Diego: Navy Personnel Research and Development Center.

Rafacz, B.A. (1994). *The design and development of a computer network system to support the CAT-ASVAB program*. San Diego: Navy Personnel Research and Development Center.

Rafacz, B.A., & Moreno, K.E. (1987). *Interactive screen dialogues for the examinee testing (ET) station.* San Diego: Navy Personnel Research and Development Center.

Rasch, G. (1960). *Probabilistic model for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Reckase, M.D. (1979). "Unifactor latent trait models applied to multifactor tests: Results and implications." *Journal of Educational Statistics, 4*, 207-230.

Reckase, M.D. (1974). *Ability estimation and item calibration using the one and three parameter logistic models: A comparative study* (RR 77-1). Columbia, MO: Tailored Testing Research Laboratory, Educational Psychology Department, University of Missouri.

Ree, M.J. (1977). "Implementation of a model adaptive testing system at an Armed Forces Entrance and Examining Station." In D.J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference.* Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Ree, M.J., Mullins, C.J., Mathews, J.J., & Massey, R.H. (1982). *Armed Service Vocational Aptitude Battery: Item and factor analysis of Forms 8, 9, and 10* (AFHRL TR-81-55). Brooks AFB, TX: Air Force Human Resources Laboratory.

Ree, M.J., & Wegner, T.G. (1990). "Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population." *Military Psychology, 2,* 157-169.

Robinson, C.A., Tomblin, E.A., & Houston, A. (1981). *Computer-managed instruction in Navy technical training: An attitudinal survey* (NPRDC TR 82-19). San Diego: Navy Personnel Research and Development Center.

Samijima, F. (1977). "A method of estimating item characteristic functions using the maximum likelihood estimate of ability." *Psychometrika, 42,* 163-191.

Samejima, F. (1976). "Graded response model of the latent trait theory and tailored testing." In C.L. Clark (Ed.), *Proceedings of the First Conference on Computerized Adaptive Testing* (PS 75-6). Washington, DC: Personnel Research and Development Center, U.S. Civil Service Commission.

Samiuddin, M. (1970). "On a test for an assigned value of correlation in a bivariate normal distribution." *Biometrika, 57*, 461-464.

Sands, W.A. (1983). *Computerized adaptive testing for the U.S. Army JOIN system.* Paper presented at the annual convention of the American Psychological Association, Anaheim, CA.

Sands, W.A. (1981). "The Navy personnel accessioning system." *Proceedings of the Annual Conference of the Military Testing Association.* Arlington, VA: Military Testing Association.

Sands, W.A. (1980). "The automated guidance for enlisted Navy applicants (AGENA) system." *Proceedings of the Annual Conference of the Military Testing Association*, Toronto: Military Testing Association.

Sands, W.A. & Gade, P.A. (1983). "An application of computerized adaptive testing in U.S. Army recruiting." *Journal of Computer-Based Instruction, 10 (3 & 4),* 87-89.

Sands, W.A., Gade, P.A., & Bryan, J.D. (1982). "Research and development for the JOIN system". *Proceedings of the Annual Conference of the Military Testing Association.* Gulf Shores, AL: Military Testing Association.

Sands, W.A., & Rafacz, B.A. (1983). "Field test of the Computerized Adaptive Screening System (CAST)." *Proceedings of the Annual Conference of the Military Testing Association.* San Antonio, TX: Military Testing Association.

SAS Institute. (1990). *SAS/STAT User's Guide* (4th ed.). Raleigh-Durham, NC: Author.

Schmid, J., & Leiman, J.M. (1957). "The development of hierarchical factor solutions." *Psychometrika, 22,* 53-61.

Schmidt, F.L., & Gugel, J.F. (1975). *The Urry item parameter estimation technique: How effective?* Paper presented at the annual convention of the American Psychological Association, Chicago.

Schmidt, F.L, Hunter, J., & Dunn, W. (1987). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery* (TN-95-5). San Diego: Navy Personnel Research and Development Center. (NTIS AD-A279580)

Schmidt, F.L, Urry, V.W., & Gugel, J.F. (1978). *Item parameterization procedures for the future.* Washington, DC: Personnel Research and Development Center, U.S. Civil Service Commission, 107-112.

Schmitz, E. (January 1995). Personal communications.

References

Segall, D.O. (1989). *Score equating development analyses of the CAT-ASVAB.* Unpublished manuscript, Navy Personnel Research and Development Center.

Segall, D.O. (1987). *ACAP item pools:Analysis and recomendations.* Unpublished manuscript, Navy Personnel Research and Development Center.

Segall, D.O. (1986). *An asymptotic hypothesis test for dependent correlations* (Draft Technical Report). San Diego: Navy Personnel Research and Development Center.

Sellman, W.S. (1988, December 14). *Memorandum for Manpower Accession Policy Steering Committee: Appointment of technical representative to a Joint-Service computerized test selection committee.* Washington, DC: Office of the Assistant Secretary of Defense.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Sims, W.H., & Hiatt, C.M. (1981). *Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6, 7, 6E, and 7E* (CNA Study 1152). Alexandria, VA: Center for Naval Analyses.

Skinner, H.A., & Allen, B.A. (1983). "Does the computer make a difference? Computerized versus face-to-face versus self report assessment of alcohol, drug and tobacco used." *Journal of Consulting and Clinical Psychology.*

Slack, W.W., & Slack, C.W. (1977). "Talking to a computer about emotional problems: A comparative study." *Psychotherapy:Theory, Research and Practice, 14,* 156-164.

Smith, E.P., & Walker, M.R. (1988). "Testing psychomotor and spatial abilities to improve selection of TOW gunners." *Proceedings of the Annual Conference of the Military Testing tion, 30,* 647-652. Arlington, VA: Military Testing Association.

Spray, J.A., Ackerman, T.A., Reckase, M.D., & Carlson, J.E. (1989). "Effect of medium of item presentation on examinee performance and item characteristic." *Journal of Educational Measurement, 26,* 261-271.

Stocking, M.L., & Lord, F.M. (1983). "Developing a common metric in item response theory." *Applied Psychological Measurement, 1,* 201-210.

Swanson, L. (1979). *Armed Services Vocational Aptitude Battery, Forms 6 and 7: Validation against school performance in Navy enlisted schools (July 1976 - February 1978)* (NPRDC TR 80-1). San Diego: Navy Personnel Research and Development Center.

Swanson, L. (1978). *Armed Services Vocational Aptitude Battery Forms, 6 and 7: Validation against school performance - interim report* (NPRDC TR 78-24). San Diego: Navy Personnel Research and Development Center.

Swanson, L., Fischl, M.A., Ross, R.M., McBride, J.R., Wiskoff, M.F., Valentine, L.D. Jr., Mathews, J.J., & Wilfong, H.D. (1978). *Validity of the Armed Services Vocational Aptitude Battery (ASVAB) for predicting performance in Service technical training schools* (TR 77-4). Chicago: Directorate of Testing, U.S. Military Enlistment Processing Command.

Sympson, J.B., & Hartmann, L. (1985). Item calibrations for computerized adaptive testing (CAT) item pools. In D.J. Weiss (Ed.). *Proceedings of the 1982 Item Response Theory and Computerized Adaptive Testing Conference.* Minneapolis: Computerized Adaptive Testing Laboratory, Department of Psychology, University of Minnesota.

Sympson, J.B., & Hetter, R.D. (1985). *Controlling item exposure rates in computerized adaptive tests.* Paper presented at the Annual Conference of the Military Testing Association. San Diego: Military Testing Association.

Sympson, J.B., Weiss, D.J., & Ree, M.J. (1984). *Predictive validity of computerized adaptive testing in a military training environment.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Thompson, J.G., & Weiss, D.J. (1980). *Criterion-related validity of adaptive testing strategies* (RR 80-3). Minneapolis: Department of Psychology, University of Minnesota.

Tiggle, R.B., & Rafacz, B.A. (1985). "Evaluation of three local CAT-ASVAB network designs." *Proceedings of the Annual Conference of the Military Testing Association,* 23-28. San Diego: Navy Personnel Research and Development Center.

Tsutakawa, R.K. (1984). *Final report on project NR 150-464 improved estimation procedures for item response functions* (RR 84-2). Columbia, MO: Department of Statistics, University of Missouri.

Tucker, L.R. (1986). "Maximum validity of a test with equivalent items." *Psychometrika, 11,* 1-13.

Upchurch, C. (November 1994). Personal communications.

Urry, V.W. (1983). *Tailored testing and practice: A basic model, normal ogive models, and tailored testing algorithms.* Washington, DC: Office of Personnel Management. (Also NPRDC TR 83-32)

Urry, V.W. (1977). "Tailored testing: A successful application of latent trait theory." *Journal of Educational Measurement, 14,* 181-196.

Urry, V.W. (1976). "Ancillary estimators for the item parameters of mental test models." In W.A. Gorham (Chair), *Computers and testing:Steps toward the inevitable conquest* (PS-76-1). Washington, DC: U.S. Civil Service Commission.

Urry, V.W. (1974a). *Computer-assisted testing: The calibration and evaluation of the verbal ability bank* (Technical study 74-3). Washington, DC: Research Section, Personnel Research and Development Center, U.S. Civil Service Commission.

Urry, V.W. (1974b). "Approximations to item parameters of mental test models and their uses". *Educational and Psychological Measurement, 34,* 253-269.

Urry, V.W. (1971). "Individualized testing by Bayesian estimation." *Research Bulletin* 171-177, Seattle: Bureau of Testing, University of Washington.

Urry, V.W. (1970). *A Monte-Carlo investigation of logistic mental test models.* Unpublished doctoral dissertation, Purdue University.

USMEPCOM. (1983). *U.S. MEPCOM Mobile Examining Team Site requirements for computerized adaptive testing.* Director of Testing, USMEPCOM Letter Report dated December 22, 1983. Vol. I, II, and III.

Vale, C.D. (1986). *The ongoing design of a micrococmputer-based adaptive testing system.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Vale, C.D. (1975). "Problem: Strategies of branching through an item pool." In D.J. Weiss (Ed.), *Computerized adaptive trait measurement: Problems and prospects* (RR 75-5). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota. (AD A018675)

Vale, C.D., & Gialluca, K.A. (1985). *ASCAL:A microcomputer program for estimating logistic IRT item parameters* (RR ONR 85-4). St. Paul, MN: Assessment Systems Corp.

Vale, C.D., & Weiss, D.J. (1975). *A simulation study of stradaptive ability testing* (RR 75-6). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota. (AD A020961)

Vineberg, R., & Joyner, J.N. (1982). *Prediction of job performance: Review of military studies* (NPRDC TR 82-37). San Diego: Navy Personnel Research and Development Center.

Wainer, H., Dorens, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wall, J.E. (1995). Briefing presented to the Defense Advisory Committee on Military Personnel Testing. Apache Junction, AZ, February 16, 1995.

Walter, G.H., & O'Neill, H.F., Jr. (1974). "On-line user-computer interface: The effects of interface flexibility, terminal type, and experience on performance." *AFIPS Conference Proceedings, 43,* 379-384.

Waters, B.K. (1994). "It ain't over 'til it's over: Implementing behavioral science R & D." T*he Military Psychologist.* Spring/Summer 1994. Washington, DC: Division of Military Psychology, American Psychological Association.

Waters, B.K. (1975). "An empirical investigation of Weiss' Stradaptive Testing Model." *In Proceedings of the First Conference on Computerized Adaptive Testing.* (PS-75-6). Washington, DC: Personnel Research and Development Center, U.S. Civil Service Commission, 54-63.

Waters, B.K. (1974). *An empirical investigation of the stradaptive testing model for the measurement of human ability.* Unpublished doctoral dissertation, Florida State University.

References

Waters, B.K., Barnes, J.D., Foley, P., Steinhaus, S.D., & Brown, D.C. (1988). *Estimating the reading skills of military applicants: Development of an ASVAB to RGL conversion table* (HumRRO Final Report, FR-PRD-88-22). Alexandria, VA: Human Resources Research Organization.

Waters, B.K., Laurence, J.H., & Camara, W.J. (1987). *Personnel enlistment and classification procedures in the U.S. military.* Washington, DC: National Academy Press.

Waters, B.K., & Lee, G.C. (1981). *Legal and political considerations in large-scale adaptive testing.* Paper presented at the annual conference of the Military Testing Association, Arlington, VA.

Weiss, D.J. (1983). *Computer-based measurement of intellectual capabilities.* Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Weiss, D.J. (1975). "Computerized adaptive ability measurement." *Naval Research Reviews,* 1-18.

Weiss, D.J. (1974a). *Strategies of adaptive ability measurement* (RR 74-5). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota.

Weiss, D.J. (1974b). *The stratified adaptive computerized ability test* (RR 73-3). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota. (AD 768 376)

Weiss, D.J., & Betz, N.E. (1973). *Ability measurement: Conventional or adaptive?* (RR 73-1). Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota

Weltin, M.M., & Popelka, B.A. (1983). *Evaluation of the ASVAB 8/9/10 clerical composite for predicting training school performance* (TR 594). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wetzel, C.D., & McBride, J.R. (1986). *Reducing the predictability of adaptive item sequences.* Paper presented at the annual conference of the Military Testing Association, San Diego: 43-48.

Wetzel, C.D., & McBride, J.R. (1983). *The influence of fallible item parameters on test information during adaptive testing* (TR 83-15). San Diego: Navy Personnel Research and Development Center.

Wilbourn, J.M., Valentine, L.M., & Ree, M.J. (1984). *Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to Air Force technical school final grades* (AFHRLTP-84-8). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Wilson, C. (1994). *PC-CAT Test Administrator Station Software* (Contract Deliverable).

Wilson, D., Wood, R., & Gibbons, R.D. (1984). *TESTFACT:Test scoring, item statistics, and item factor analysis*. Mooresville, IN: Scientific Software.

Wingersky, S., Barton, M.A., & Lord, F.M. (1982). *LOGIST user's guide* Princeton, NJ: Educational Testing Service.

Wingersky, S., & Lord, F.M. (1973). *A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses* (RM 73-2). Princeton, NJ: Educational Testing Service.

Wise, L.L., McHenry, J.J., Chia, W.J., Szenas, P.L., & McBride, J.R. (1990). *Refinement of the computerized adaptive screening test*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wiskoff, M.F. (1985). *Military psychology and national defense: Making a difference.* Division 19 Presidential Address at the annual convention of the American Psychological Association, Los Angeles.

Wiskoff, M.F. (1981). "Computerized adaptive testing." *Proceedings of the National Security Industrial Association First Annual Conference on Personnel and Training Factors in Systems Effectiveness*. San Diego: National Security Industrial Association.

Wolfe, J.H. (Ed.). (in press). Special issue on ECAT. *Military Psychology, 9,1* .

Wolfe, J.H. (1985). "Speeded tests: Can computers improve measurement?" *Proceedings of the Annual Conference of the Military Testing Association, I.* San Diego: Military Testing Association, 49-54.

Wolfe, J.H., & Alderton, D.L. (1992). "Navy incremental validity study of new predictors." *Proceedings of the Annual Conference of the Military Testing Association.* San Diego: Military Testing Association, 39-44.

Wolfe, J.H., Alderton, D.L., & Larson, G.E. (1993). *Incremental validity of new computerized aptitude tests for predicting training performance in nine Navy technical schools.* Navy Personnel Research and Development Center, Unpublished manuscript.

Wolfe, J.H., Alderton, D.L., Larson, G.E., & Held, J.D. (1995). *Incremental validity of Enhanced Computer Administered Testing (ECAT)* (TN 96-6). San Diego: Navy Personnel Research and Development Center.

Wood, R.L., Wingersky, M.S., & Lord, F.M. (1976). *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters* (RM-76-6). Princeton, NJ: Educational Testing Service.

Wright, B.D. (1977). "Solving measurement problems with the Rasch model." *Journal of Educational Measurement, 14,* 97-166.

Wright, B.D., & Douglas, G.A. (1977). "Conditional versus unconditional procedures for sample-free item analysis." *Educational and Psychological Measurement, 37,* 47-60.

Wright, B.D., & Mead, R.J. (1977). *BICAL: Calibrating items and scales with the Rasch model* (RM No. 23). Chicago: Statistical Laboratory, Department of Education, University of Chicago.

Yoes, M.E., & Hardwicke, S.B. (1984). *The effect of mode of presentation on test performance.* San Diego: RGI.

LIST OF ACRONYMS

ACRONYM	DEFINITION
1PL	One-Parameter Logistic IRT Model
3PL	Three-Parameter Logistic IRT Model
ACAP	Accelerated CAT-ASVAB Program
AFHRL	U.S. Air Force Human Resources Laboratory
AFQT	Armed Forces Qualification Test
AFQT CATEGORY	AFQT Score Group
AGCT	Army General Classification Test
AI	Aotomotive Information Test of the CAT-ASVAB
AO	Assembling Objects Test in ECAT Battery
AR	Arithmetic Reasoning Test of the ASVAB
ARI	Army Research Institute for the Behavioral and Social Sciences
ART	ASVAB Review Technical Committee
AS	Auto and Shop Information Test of the ASVAB
ASD/FM&P	Assistant Secretary of Defense for Force Management and Personnel
ASD/M&L	Assistant Secretary of Defense for Manpower and Logistics
ASVAB	Armed Services Vocational Aptitude Battery
ASVAB CEP	Armed Services Vocational Aptitude Battery - Career Exploration Program
ATG	Acceptance Testing Group
BBN	Bolt, Beranek, and Newman
BDM	BDM Federal Inc.
BIT	Built-in Test Software
BME	Bayesian Modal Estimation
BRTT	Broad Range Tailored Test
BSE	Bayesian Sequential Estimation
CAST	Computerized Adaptive Screening Test
CAT	Computerized Adaptive Test
CAT-ASVAB	Computerized Adaptive Testing Version of the ASVAB
CATICC	CAT Inter-Service Coordinating Committee
CATWG	CAT Working Group
CBO	Congressional Budget Office
CDF	Cumulative Distribution Function
CMOA	Calibration Mode of Administration
COPE	Concepts of Operation Planning and Evaluation Panel
CPU	Central Processing Unit
CS	Coding Speed Test of the ASVAB
CSMA/CD	Carrier Sense Multiple Access/Collision Detection
CT	Mental Counters Test in ECAT Battery
CTC	Contract Testing Center
DAC	Defense Advisory Committee on Military Personnel Testing
DAT	Differential Aptitude Tests
DEP	Delayed Entry Program
DHC	Data Handling Computer

LIST OF ACRONYMS

ACRONYM	DEFINITION
DIF	Differential Item Functioning
DMDC	Defense Manpower Data Center
DoD	U. S. Department of Defense
DoD-STP	Department of Defense Student Testing Program
DoL	U. S. Department of Labor
DOS	Disk Operating System
DRP	Digital Response Pad
ECAT	Enhanced Computer-Administered Tests
EI	Electronics Information Test of the ASVAB
EST	Enlistment Screening Test
ET	Examinee Testing Station
FEDSIM	Federal Computer Performance Measurement and Simulation Center
FR	Figural Reasoning Test in ECAT Battery
FSG	Final School Grade
GRE	Graduate Record Examination
GS	General Science Test of the ASVAB
HP-IPC	Hewlett Packard Integral Personal Computer
HumRRO	Human Resources Research Organization
ID	Integrating Details Test in ECAT Battery
IOT&E	Initial Operational Test and Evaluation
IRT	Item Response Theory
ISA	Industry Standard Adapter
JOIN	Joint Optical Information Network
K-S	Kolmogorov-Smirnov Statistical Test
LAN	Local Area Network
LCN	Local CAT-ASVAB Network
MAP	Manpower Accession Policy Steering Committee
MAPWG	Manpower Accession Policy Working Group
MC	Mechanical Comprehension Test of the ASVAB
MCRD	Marine Corps Recruit Depot
MDAC	McDonnell-Douglas Astronautics Corporation
MEPS	Military Entrance Processing Station
METS	Mobile Examining Team Site
MHz	Megahertz
MIRS	USMEPCOM Integrated Resource System
MISE	Mean Integrated Square Error
MK	Mathematics Knowledge Test of the ASVAB
MLE	Maximum Likelihood Estimation
MLMI	Maximum Likelihood/Maximum Information

LIST OF ACRONYMS

ACRONYM	DEFINITION
MOA	Mode of Administration
NIC	Network Interface Controller
NLSY79	1979 National Longitudinal Survey of Youth Labor Force Behavior
NO	Numerical Operations Test of the ASVAB
NOS	Network Operating System
NPAS	Navy Personnel Accessioning System
NPRDC	Navy Personnel Research and Development Center
NSFH	National Survey of Families and Households
NVSNP	Navy Validity Study of New Predictors
O&S	Operations and Support
OASD	Office of the Assistant Secretary of Defense
OPM	U. S. Government Office of Personnel Management
OSD	Office of the Secretary of Defense
OT&E	Operational Test and Evaluation
P&P	Paper-and-Pencil
P&P-ASVAB	Paper-and-Pencil Version of the ASVAB
PACE	Professional and Administrative Career Examination
PAY80	1980 *Profile of American Youth* Study
PC	Paragraph Comprehension Test of the ASVAB
R&D	Research and Development
RAM	Random Access Memory
RFP	Request for Proposal
RGL	Reading Grade Level
RISC	Reduced-Instruction-Set-Computing
ROI	Return on Investment
ROM	Read-Only Memory
SDS	Self-Directed Search
SED	Score Equating Development
SEV	Score Equating Verification
SI	Shop Information Test of the CAT-ASVAB
SM	Sequential Memory Test in ECAT Battery
SO	Spatial Orientation Test in ECAT Battery
SOS	Sophisticated Operating System (Apple Computer)
SSAN	Social Security Account Number
STMI	Stratified Maximum Information
STRADAPTIVE	Stratified Adaptive Testing Strategy
SVGA	Super Video Graphics Array
T1	One-Handed Tracking Test in ECAT Battery
T2	Two-Handed Tracking Test in ECAT Battery
TA	Test Administrator

LIST OF ACRONYMS

ACRONYM	DEFINITION
TASP	Technical Advisory Selection Panel
TCC	Test Characteristic Curve
TI	Target Identification Test in ECAT Battery
TIF	Test Information Function
TSR	Terminate-and-Stay-Resident Driver
USAREC	U.S. Army Recruiting Command
USMEPCOM	U.S. Military Entrance Processing Command
UID	Unique Identification Number
VE	Verbal Composite of the ASVAB
VGA	Video Graphics Adaptor
WK	Word Knowledge Test of the ASVAB