Technical Report 625

AD-A162 709

A Theory-Based Approach to Reading Assessment in the Army

Rebecca L. Oxford-Carpenter
Linda J. Schultz-Shiner

Selection and Classification Technical Area
Manpower and Personnel Research Laboratory

DTIC FILE COPY

DTIC ELECTE DEC 3 0 1985

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

July 1984

Approved for public release; distribution unlimited.

Best Available Copy

85 12 27 0

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Technical Reviews by

Joan Harman
Judith Orasanu
Paul Rossmeissl

NOTICES

DISTRIBUTION: Primary distribution of this report 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: PERI-POT, 5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600.

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

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report 625	2. GOVT ACCESSION NO. AD-A162 709	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Theory-Based Approach to Reading Assessment in the Army		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Rebecca L. Oxford-Carpenter & Linda J. Schultz-Shiner		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Research Institute for the Behavioral & Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263731A791
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 1 July 1984 13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSISFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Assessment; Reading assessment; Reading theory; Army testing; ASVAB; Reading comprehension; Reading tests; Reading test scores; Grade equivalent scores

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report addresses practical Army problems in reading assessment from a theory base that reflects the most recent and most sound research on reading comprehension. Six major conclusions are drawn from both theory and practice. First, reading is important in military and civilian work life. Second, reading assessment is a highly visible and important issue in the Army. Third, reading theories--especially the new interactive-inferential theory--can positively influence reading measurement practices in the Army. Fourth,

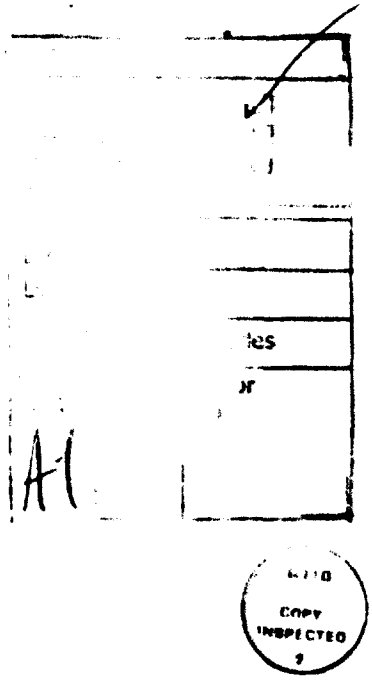
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (Continued)

reading tests are not all alike; they differ widely in terms of psychometric characteristics and overall quality as evaluated using theory-based standards. Fifth, high correlations exist between the Armed Services Vocational Aptitude Battery (ASVAB) and various reading tests, although caution needs to be exercised in using any part of the ASVAB as a reading-test surrogate. Sixth, alternatives to grade equivalent scores are available and should be considered for Army use.



COPY INSPECTED

Technical Report 625

A Theory-Based Approach to Reading Assessment in the Army

Rebecca L. Oxford-Carpenter
Linda J. Schultz-Shiner

Submitted by
Newell K. Eaton, Chief
Selection and Classification Technical Area

Approved as technically adequate
and submitted for publication by
Joyce L. Shields, Director
Manpower and Personnel Research Laboratory

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

July 1984

Army Project Number
2Q263731A791

Manpower and Personnel Research

Approved for public release; distribution unlimited

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Manpower and Personnel Research Laboratory of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducts research in areas related to manpower and personnel issues. One key issue is reading performance of Army soldiers as shown through reading assessment. Reading assessment has become an important problem not just for training but also for personnel selection and classification. This report addresses Army reading assessment from a theory base that is both current and useful. The research was conducted under ARI Project Number 2Q263731A791. The information reported here will be helpful to policymakers and scientists concerned with military selection, classification, training, and education.

EDGAR M. JOHNSON
Technical Director

A THEORY-BASED APPROACH TO READING ASSESSMENT IN THE ARMY

EXECUTIVE SUMMARY

Requirement:

To address practical Army problems in reading assessment from a theory base that reflects the most recent and most sound research on reading comprehension.

Procedure:

Specific aspects of the Army reading assessment problem were identified. The most current and relevant reading theory and its implications for reading assessment in general were then explained. Finally, the relationship of those implications to Army reading assessment was described.

Findings:

Six major conclusions are drawn from both theory and practice. First, reading is important in military and civilian work life. Second, reading assessment is a highly visible and important issue in the Army. Third, reading theories, especially the new interactive-inferential theory, can positively influence reading measurement practices in the Army. Fourth, reading tests differ widely in terms of psychometric characteristics and overall quality as evaluated by theory-based standards. Fifth, the Armed Services Vocational Aptitude Battery (ASVAB) and various reading tests have a high correlation, although caution needs to be exercised in substituting any part of the ASVAB for a reading test. Sixth, alternatives to grade equivalent scores are available and should be considered for use in the Army.

Utilization of Findings:

This report has utility for Army selection, classification, training, and education. The findings may also be helpful to other military services concerned with reading assessment issues.

A THEORY-BASED APPROACH TO READING ASSESSMENT IN THE ARMY

CONTENTS

A THEORY-BASED APPROACH TO READING ASSESSMENT IN THE ARMY

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has been involved in research on reading assessment in the Army from both practical and theoretical perspectives. The purpose of this paper is to address practical Army problems in reading assessment from a theory base that reflects the most recent and most sound research on reading comprehension. This paper has seven sections. The first two sections, which concern the importance of reading in work life and the importance of reading assessment in the Army, provide the basis for addressing the overall issue of reading assessment in the Army. The reading theories presented in the third section are helpful in generating evaluative criteria for reviewing potential measures of reading for Army use. In the fourth and fifth sections, three categories of measures are considered: civilian, military, and the special case of the Armed Services Vocational Aptitude Battery (ASVAB)--a military test of general cognitive ability with potential value as a surrogate reading measure. The sixth section deals with scoring issues that need to be addressed no matter what measure is used. The final section offers conclusions based on both theory and practice.

READING IS IMPORTANT IN MILITARY AND CIVILIAN WORK LIFE

Military research on reading has shown that reading is an important part of military life. Sacher and Duffy (1978) found that military workers' overall job performance suffered if they read more than two grade levels below job demands. Sticht (1980) observed a low generalizability of general reading skills to specific literacy tasks in the military and emphasized the need for job-related reading tests. Sticht (1982) also investigated empirical relations between reading proficiency and job proficiency in the military setting and found that the armed services pose demands for basic skills equal to or greater than those of civilian jobs. Even with higher enlistment standards, many military recruits are below average in basic skills, including not only reading but also other skills. However, Sticht (1982) found that basic skills competence as measured by a variety of tests did not appear to be the overriding determinant of success in the military. The precise relationship between reading skills and job skills needs further investigation.

In addition to military research, civilian research may have some implications for Army literacy. Diehl and Mikulecky (1980) observed 100 workers representing a cross-section of occupations and found that 90% of the workers participated in some form of reading each day. In a later study, Mikulecky (1982) compared high school reading to work reading and found that students read less for school than most workers did for work. Workers read more often for application or to make judgments, while students read to gather facts. Mikulecky and Strange (in preparation) noted that the workplace calls for a wider range of literacy strategies than does school literacy, 95% of which is based on textbooks. These civilian research results may, by implication, underscore the importance of reading in many, if not all, military jobs.

READING ASSESSMENT IS IMPORTANT IN THE ARMY

In the last few years reading assessment has become a key issue in the Army primarily because of changes in the reading ability levels of the Army population. Population shifts are, in turn, related to a multitude of factors, such as changes in the national economic picture and variations in norming of the ASVAB used for Army selection and classification. In recent years the Training and Doctrine Command (TRADOC), the Forces Command (FORSCOM), the Education Directorate of The Adjutant General's Office (TAGO), and the Office of the Deputy Chief of Staff for Personnel (ODCSPER) have all turned to ARI for practical advice on reading assessment. For example, both TRADOC and FORSCOM have asked for ARI's help in measuring the reading level of noncommissioned officers (NCOs). TRADOC Systems Analysis Activity (TRASANA) has requested assistance in interpreting its data on reading tests of officers and enlisted personnel. TRADOC is currently establishing a program to improve officers' communication skills and has asked for ARI's advice on the use of reading tests as part of the program. TAGO is funding an ARI project, known as the Job Skills Education Program (JSEP), a computer-based effort that involves the assessment of job-oriented basic skills. ARI has also advised ODCSPER on the difficulties of using reading as a reenlistment criterion for midterm noncommissioned officers.

Army agencies often ask the following questions about reading assessment: "What reading tests can we use?" "What do our reading test scores mean?" and "Since ASVAB scores relate to reading test scores, can the ASVAB be used as a substitute for a reading measure?" Recent requests for assistance have not focused on the more central question, "How does reading proficiency relate to job proficiency?" The last question--perhaps the most crucial question for Army reading assessment--is an empirical one that research has not yet answered adequately. However, both theory and research provide answers to the first three questions. Before addressing these questions in detail, we will summarize several key types of reading theories and indicate their implications for reading assessment.

THREE TYPES OF READING THEORIES ARE EXAMINED

Many theories of reading exist. One theory unites perceptual and cognitive elements. A second set of theories relates to information processing. A third theory, the most recent, concerns inferencing in an interactive mode. We will describe the essence of these theories and their implications for reading assessment. The greatest emphasis is on the last theory because of its major implications for reading assessment.

Perceptual-Cognitive Theory Involves Strategies

Gibson and Levin (1975) presented a theory that is both perceptual and cognitive. Although they stressed most heavily the perceptual underpinnings of the reading process, these researchers also indicated that psychological processes of mature reading go far beyond perception to remembering, problem solving, and organization of conceptual knowledge for better extraction of

meaning. Two implications of perceptual-cognitive theory for reading assessment are (1) the cognitive strategies that help people improve their reading can also be used to help people perform better on reading tests, and (2) these strategies are not innate but can be taught.

Information Processing Theories Analyze Reading Components

As noted by Carpenter and Just (in preparation), information processing is characterized by efforts to understand what information is represented in memory, what information is acquired, how processes are acquired and invoked, how long these processes take, and what sources of error exist. A major contribution of information processing theory is that it acknowledges the role of environment (not just previously acquired knowledge and thought patterns) in explaining reading behavior. The information processing model of Sticht, Beck, Hauke, Kleiman, and James (1974) emphasizes the interaction between the environment and cognitive components such as sensory information storage, short-term memory, and long-term memory to generate literacy skills. Massaro's (1975) information processing model of reading delineates four component processes or stages between the language stimulus and the meaning response: feature detection (sensation), primary recognition (perception), secondary recognition (conception), and recoding and rehearsal. Information processing analyzes reading into component processes; however, the parts are not always easily reconstituted into the reading act (Carpenter & Just, in preparation). Information processing theory is now being used in computer-reading simulations, in which successful computer programs must have a great deal of "knowledge" about vocabulary, language structure, and the topic of the text. Carpenter and Just (in preparation) described some recent computer simulations of reading based on information processing.

Information processing theories of reading have some implications for reading assessment. First, these theories imply that environment and short- and long-term memory are very important in the reading process and that therefore reading is not a simple thing to measure. Second, one application of information processing theory (that of Sticht et al., 1974) has shown that reading tests need to be functional (i.e., related to the person's environment) for results to be most meaningful.

Interactive-Inferential Theory Provides a New Perspective

Although the emerging reading theory lacks a consistent, descriptive name, we will call it interactive-inferential theory. This theory builds largely on the information processing model but also goes beyond it by demonstrating the interaction between higher order, interpretive, metacognitive processes and less complex processes. Interactive-inferential theory focuses on interactive rather than linear movement in reading and has generated a spate of research that centers on reading comprehension instead of decoding. We devote more time to interactive-inferential theory than to earlier theories, because it offers a more "evolved" view and a number of implications for reading assessment. According to this theory, reading (1) is an active process in which the reader constructs meaning through inference and interpretation, (2) is purposeful and hence involves motivation, (3) can be

improved through use of cognitive and learning strategies, (4) is processed by the individual through interactions among several levels of information, and (5) can be divided into a set of subskills that fit into a unified process but that are not as yet fully defined or understood (Farr, Carey, & Tone, in preparation; J. Orasanu, personal communication, January 12, 1984). Each of these characteristics of reading is discussed below, along with its implications for reading assessment.

Action, interpretation, and inference. The reader actively constructs meaning from text cues by calling on knowledge of language, text structure, writing conventions, and the topic itself. The reader applies inference and interpretation to go beyond what is given (Farr, Carey, & Tone, in preparation). Researchers have studied text cues through methods of text analysis and text linguistics. The active, inferential, interpretive nature of reading creates at least seven implications for reading assessment.

First, reading comprehension tests may be most valid when they are functional, that is, when they involve materials similar to those needed to perform everyday and vocational tasks (Miller, 1973). "Reading assessment ought to reflect the schema domains, syntax, vocabulary, style, and structure of materials that will need to be read by the individuals taking the tests" (Farr, Carey, & Tone, in preparation, p. 24). As a consequence, job-relevant reading tests of a criterion-referenced nature (i.e., keyed to a specified criterion or standard of performance) may be more useful for Army populations than are other kinds of reading tests, at least for the purpose of assessing comprehension. However, if a test is given with the intent of ranking group members and not with the intent of carefully assessing comprehension, then it might be acceptable to use a norm-referenced, general reading test that is not job related.

Second, in order to employ a somewhat uniform cognitive schema, reading tests need to be developed, tested, and normed, using populations similar to those who will take the test. For example, an adult reading test should not be developed using just children, and items developed for children may not be relevant for adults. These facts, while seemingly straightforward, are overlooked with surprising frequency.

Third, the active, inferential, interpretive characteristic of the new theory implies that tests should be appropriate to the examinees insofar as possible. Therefore, adaptive (tailored) testing may be useful. Adaptive testing allows the individual to take only items that are at a relevant difficulty level and may shorten test administration time (McBride, 1979). The joint services are now designing and testing for the ASVAB computerized adaptive testing, which could be used for adult reading tests as well.

Fourth, because readers of various cultural backgrounds can interpret reading passages differently and come to different conclusions, test bias (known as "differential validity") can occur unless background knowledge is controlled or explained. However, because reading comprehension depends greatly on background knowledge, such knowledge cannot be artificially factored out or eliminated by using esoteric content (Farr, Carey, & Tone, in preparation).

Fifth, reading tests of literal recall are not as revealing or lifelike as reading tests that demand some sort of inferencing. Old-fashioned tests of literal recall are therefore to be avoided.

Sixth, the theory's encouragement of relevance of the test to the examinee indicates that child-based scores such as reading grade levels (grade equivalents) are not useful for adults, although those scores are ubiquitous in the armed forces (Farr, Carey, & Tone, in preparation; Miller, 1973).

Seventh, text linguistics as used in the new theory can provide guidance on how to create more reliable reading test items, that is, items that accurately and consistently measure the skill in question. Sample guidelines include avoiding items that require stylistic and other ambiguous judgment; not testing for incidental, insignificant information; not using harder vocabulary in questions than in text; and avoiding list-like density of ideas in test items (J. Orasanu, personal communication, March 9, 1984).

In addition to the active, inferential, interpretive aspect of the new theory, other aspects--purposefulness, cognitive strategies, nonlinearity, and subskills--also have implications for reading assessment.

Purposefulness. The fact that reading is purposeful and involves motivation implies that reading test developers should select and present test passages that engage the reader in a valid purpose, not just the purpose of passing the test (Farr, Carey, & Tone, in preparation). The interest level and degree of relevance must be kept high to maintain the reader's sense of purpose. A negative sense of purpose (e.g., "If I don't pass this reading test I may not be allowed to reenlist") may have a detrimental effect on the performance of some readers but not of others. Functional reading tests might instill a greater sense of purpose than more general reading tests do.

Cognitive strategies. An implication of the existence of cognitive strategies is that such strategies can improve not only reading in general but also performance on reading tests. Different strategies useful for a variety of purposes and genres can be taught. Such strategies can develop readers' sensitivity and can thus improve comprehension of a passage in a reading test or any other text (Brown & Armbruster, in preparation). Cognitive strategies are also important in earlier theories, such as that of Gibson and Levin (1975).

Nonlinearity. Readers use interactions among several information levels, moving "top-down" and "bottom-up" and mixing higher order inference with simpler perceptual processes. Because of the nonlinear, complex interaction that occurs, reading should not be assessed as the automatic decoding of a sequential string of letters.

Subskills. The last few decades have seen a proliferation of subskills in reading tests (Farr, Carey, & Tone, in preparation), despite lack of adequate subskill definition, lack of a coherent reading theory from which subskills derive, and lack of a consensus on how many items are needed to measure a particular subskill. The new theory may yet lead to a better understanding of these subskills and how they can best be measured.

Summary of Theories

We have discussed three types of reading theories: perceptual-cognitive, information processing, and interactive-inferential. These theories all have important implications for reading assessment. Some measure of agreement exists across theories on the importance of examinee context, background knowledge, and cognitive strategies for reading. The last theory seems to be the most helpful in offering concrete implications for reading measurement. We will now discuss psychometric qualities of available reading tests in light of these implications.

PSYCHOMETRIC CHARACTERISTICS OF AVAILABLE TESTS DIFFER

In this section, the standards implied by reading theory will be applied to the review of reading tests currently available for Army use. This review will cover both commercially and militarily developed reading tests. Theory and practice tell us that good reading tests should have certain psychometric characteristics, including appropriate validity, reliability, norms, and standards. All reading theories imply the need for valid measures, that is, tests that measure what they purport to measure. The emerging interactive-inferential reading theory particularly emphasizes the need for a type of content validity (or at least content relevance) that reflects the schema the reader ordinarily employs. The need for reliability of measurement is implied in the text-linguistic basis of this new reading theory. Appropriate norms (for norm-referenced tests) and appropriate standards (for criterion-referenced tests) are also implicit in the emerging theory.

Armed Services Have Used Commercial Reading Tests

Standardized, norm-referenced, commercial reading tests have dominated the armed forces market in the past. Widely used tests have included the Adult Basic Literacy Examination (ABLE), the Test of Adult Basic Education (TABE), the Nelson-Denny Reading Test, and the Metropolitan Achievement Test (MAT), Reading, Form D. All of these tests have been reviewed elsewhere (Buros, 1972, 1975; Oxford-Carpenter & Schultz, 1983). We will provide a brief assessment of these tests in light of the standards implied by reading theory.

The ABLE is one of a few commercially available, psychometrically sound reading tests originally designed for and normed on adults. From a purely technical standpoint, the ABLE appears to be a good commercially developed reading test for Army use. The ABLE has well-documented reliability (high .80s and .90s). Concurrent validity is based on administration of the ABLE and the Stanford Achievement Test to elementary and junior high school students (.60-.76) and to an adult job corps group (.36-.72). The test was designed for adults with varying achievement levels and for adults who have not completed formal eighth-grade education. Norms are based on 6,000 elementary and junior high school students, 800 job corps members, and 450 adult basic education students. Vocabulary, reading, spelling, and arithmetic subtests are available in two parallel forms. The reading subtest focuses on comprehension. The ABLE was originally selected for use in the Army's Basic Skills

Education Program (BSEP). However, the test proved unpopular with some BSEP personnel for reasons related to administration not technical quality (Raines, 1983). The ABLE was replaced by the TABE for BSEP use about five years ago.

Though possessing some technical merit, the TABE was chosen more on practical than technical grounds. The TABE covers reading, language, and arithmetic; its reading subtest contains both vocabulary and comprehension. The test is actually a revision of the California Achievement Test (CAT), which was designed for children. TABE developers removed patently childish references from the CAT or changed them to adult references. TABE subtests have adequate reliability (in the high .70s through .90s). Validity documentation for the TABE is not compelling, because its concurrent validity is based on a correlation of .56 with the General Educational Development examination using a small sample. The idea of "inherited" validity from the CAT is spurious. Furthermore, the TABE lacks adult norms, and its use in testing adults in the Army has drawn strong criticism (Bachem, 1982). Clearly, key career decisions for soldiers should not be made on the basis of children's norms, such as those of the TABE. However, the TABE has been useful in BSEP for diagnosing soldiers' strengths and weaknesses.

The Army has frequently used the Nelson-Denny Reading Test and the MAT. Designed to assess reading for college placement and adult reading classes, the Nelson-Denny has adult norms. The MAT was given a new, Army cover and used as the U.S. Armed Forces Institute (USAFI) Reading Test with few changes in the civilian-to-military transformation. Its norm group includes no adults. For both tests reliability is strong, but validity is not. The Army has also occasionally used other commercial reading tests, such as the Gates-McGinitie Reading Test and the Iowa Test of Basic Skills. A review of these commercially developed reading tests is presented in the Appendix.

These tests need to be evaluated on the basis of whether they provide appropriate validity, reliability, and norms demanded by reading theory and by practicality. The Army particularly needs to examine the validity of commercial reading tests in light of Army needs. Although a given reading test may be highly valid for a group of school children, it may not be valid for adults in an intense, job-oriented Army setting. Furthermore, the issue of standards is pertinent to commercial reading tests as used in the Army. Most commercial reading tests are purely norm referenced and do not advertise acceptable standards of performance. However, the Army has occasionally used these tests in a quasi-criterion-referenced way by designating a given grade equivalent test score as an acceptable minimum standard for soldiers. For example, ninth-grade reading level has become a magic number to some Army leaders. Before any standards are chosen, those standards need to be shown to correlate highly with Army job performance (not just with Army job materials). Insufficient research has been done in this fruitful area.

Armed Forces Have Developed Their Own Reading Tests

While the armed forces have commonly used commercial, norm-referenced reading tests, the military has also created reading tests. The Army has taken the lead in much of this test development. Throughout the 1970s and the early 1980s, the Army worked on a series of job-related reading tests for Army enlisted personnel (Clauoy & Caylor, 1982; Sticht, 1975, 1982; Sticht, Hooke, &

Caylor, 1982; Sticht, Caylor, & James, 1978; and Sticht, Caylor, Kern, & Fox, 1971). These efforts produced the Job Reading Task Tests (JRTT) and the Job Reading Test (JRT), which are MOS-specific and emphasize functional literacy. Unlike the JRTT, the JRT is normed and machine-scorable. As mentioned earlier, the Army is currently involved in a massive test development undertaking under the Job Skills Education Program (JSEP), which will provide computerized, job-related testing in basic skills. The other armed services have also developed their own reading measures, but the efforts appear to have been independent rather than coordinated. A leading developer of Army reading tests stated that the Army's own reading tests have never been accepted because of lack of internal advocacy. Support existed for test development but not for test use (T. Sticht, personal communication, August 16, 1983). A review of selected militarily developed reading tests is found in the Appendix.

An important merit of some reading tests developed by the military is also one deficiency: job-specificity. The emerging reading theory encourages the use of functional, job-related reading tests. However, because approximately 350 jobs exist in the Army alone, the task of developing just one test form per job is a monumental endeavor. Of course, to enhance reliability there should be several test forms per job. Furthermore, rapid technological change in job content means that job-related tests need to be continually updated. Such an endeavor is obviously very costly and difficult. Other options might be to develop reading tests for career management fields instead of jobs or to develop reading tests for only the highest density jobs. Even these options are not simple.

Given the issues associated with use of commercial and military reading tests, the Army has occasionally sought a surrogate measure of reading, a test that can be said to measure reading while not actually measuring it. Many Army administrators have looked longingly toward the ASVAB as such a surrogate reading test.

ASVAB SCORES ARE CORRELATED WITH READING TEST SCORES

Because most reading tests at present cannot adequately account for the role of such factors as background knowledge and reasoning ability, such tests may be seen as measures of intelligence and experience as well as measures of reading ability (Farr, Carey, & Tone, in preparation). Consequently, it is not surprising that reading tests are often highly correlated with general measures of aptitude or "trainability" (Jensen, 1981), such as the ASVAB. Like similar tests, the ASVAB is "reading-dependent," that is, reading is necessary but not sufficient to perform well on the test. Moderate to large correlations have been demonstrated between the ASVAB and several reading tests. For example, Sticht (1975) showed a .65 correlation between an unspecified reading test and the Armed Forces Qualifying Test (AFQT), which covers four key ASVAB subtests. The Job Reading Task Tests developed by Sticht for the Army also have moderate (in the .50s and .60s) correlations with the AFQT (Sticht, 1975). Fischl (1981) found that the USAFI Reading Test correlated .80-.95 with various composites of the ASVAB and with the total ASVAB for a sample of 600 soldiers. In a recent investigation involving 2,385 Army and Marine recruits, a .85 correlation was found between the ABLE and the

General Technical (GT) composite of the ASVAB (P. Grafton, personal communication, August 15, 1983). Among several groups of soldiers whose English proficiency was limited and who were headed for Army English-as-a-second-language (ESL) training, Oxford-Carpenter (1982) found very low correlations (in the teens and .20s) between the ABLE and a number of ASVAB composites. After ESL instruction, the correlation between the same soldiers' ABLE and ASVAB scores rose to the .40s and .50s. In summary, reading tests have been shown to correlate moderately to highly with the ASVAB as long as language problems (for individuals with limited proficiency in English) do not intervene.

A large verbal element clearly exists in the ASVAB (Valentine, n.d.), and some researchers have conjectured that the ASVAB indirectly measures reading ability. In fact, the Army has devised a so-called literacy index for the ASVAB by equating the ASVAB-GT with a composite of three ABLE subtests, reading, vocabulary, and arithmetic reasoning (P. Grafton, personal communication, August 15, 1983). The index involves ABLE grade equivalent scores, which, while apparently simple, have some complexities that are discussed next.

GRADE EQUIVALENT SCORES CAN BE MISLEADING

Grade equivalent scores, popularly known in the military as reading grade levels when applied to reading tests, are the most used and abused of all score types. The emerging interactive-inferential reading theory implies that reading tests should be as appropriate as possible to the examinees. Therefore, use of grade equivalent scores with adults is potentially misleading (Miller, 1973). Bachem denounced the use of grade equivalent scores in the military setting: "The use of elementary school grade levels to categorize adult combat soldiers seems little short of an insult, no matter how desperate their need for remedial work may be" (Bachem, 1982, p.4). Grade equivalent scores are not applicable to high school or adult levels because these scores have ambiguous meaning beyond the earlier years of constant growth (Ysseldyke & Marston, 1982), and because they compare adults not with their peers but with children. Even using grade equivalent scores with children can be misleading (Oxford-Carpenter & Schultz, 1983). Scores other than grade equivalents are abundant, and many are potentially useful for reading assessment in the Army. Some useful norm-referenced scores include percentile ranks, standard scores, stanines, and some types of adjusted gain scores. Criterion-referenced scores, which are frequently encountered in Army training programs, include number of objectives mastered or passed, number of trials to mastery, and time to mastery. These scores may also be useful for Army reading measurement. The emerging theory of reading suggests that criterion-referenced scores may be very helpful, because criterion-referenced tests are geared toward specific objectives and have relevant standards and may therefore be more germane than other types of tests. However, the theory does not rule out use of norm-referenced tests and their associated scores.

SIX MAJOR CONCLUSIONS ARE REACHED

Six major conclusions may be drawn from both theory and practice. First, reading is important in military and civilian work life. Second, reading assessment is a highly visible and important issue in the Army. Third, reading theories--especially the interactive-inferential theory--can positively influence reading measurement practices in the Army. Fourth, reading tests differ widely in terms of psychometric characteristics and overall quality as evaluated using theory-based standards. Fifth, high correlations exist between the ASVAB and various reading tests, although caution needs to be exercised in using any part of the ASVAB as a reading surrogate. Sixth, alternatives to grade equivalent scores are available and should be considered for use by the Army.

REFERENCES

Bachem, M. B. (1982). Test of Adult Basic Education (TABE or not TABE). In Proceedings, 24th Annual Conference of the Military Testing Association (pp. 48-51). Alexandria, VA: Defense Technical Information Center.

Brown, A. L., & Armbruster, B. B. (in preparation). The role of metacognition in reading and studying. In J. Orasanu (Ed.), Reading comprehension: From research to practice. Hillsdale, NJ: Erlbaum.

Buros, O. K. (1972). Seventh mental measurements yearbook. Highland Park, NJ: Gryphon Press.

Buros, O. K. (1975). Eighth mental measurements yearbook. Highland Park, NJ: Gryphon Press.

Carpenter, P. A., & Just, M. A. (in preparation). Recent progress in basic research on the cognitive processes in reading. In J. Orasanu (Ed.), Reading comprehension: From research to practice. Hillsdale, NJ: Erlbaum.

Claudy, J. G., & Caylor, J. S. (1982). Development of job-based reading tests (ARI Technical Report 479). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A123 324)

Diehl, W., & Mikulecky, L. (1980). The nature of literacy at work. Journal of Reading, 24, 221-227.

Farr, R., Carey, R., & Tone, B. (in preparation). Recent theory and research into the reading process: Implications for reading assessment. In J. Orasanu (Ed.), Reading comprehension: From research to practice. Hillsdale, NJ: Erlbaum.

Fischl, M. A. (1981). Soldier reading ability: The advocacy point of view. In Proceedings, 23rd Annual Conference of the Military Testing Association (pp. 1447-1450). Alexandria, VA: Defense Technical Information Center.

Gibson, E. J., & Levin, H. (1975). The psychology of reading. Cambridge, MA: MIT Press.

Jensen, A. R. (1981). Straight talk about mental tests. New York: Free Press.

Massaro, D. W. (1975). Understanding language: An information-processing analysis of speech perception, reading and psycholinguistics. New York: Academic Press.

McBride, J. R. (1979). Adaptive mental testing (ARI Technical Report 423). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A088 000)

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workspace actuality. Reading Research Quarterly, 17, 400-419.

Mikulecky, L., & Strange, R. L. (in preparation). Effective literacy training programs for adults in business and municipal employment. In J. Orasanu (Ed.), Reading comprehension: From reasearch to practice. Hillsdale, NJ: Erlbaum.

Miller, G. A., Ed. (1973). Linguistic communication: Perspectives for research. Newark, DE: International Reading Association.

Oxford-Carpenter, R. (1982). Data tables for English-as-a-second language. Unpublished manuscript, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Oxford-Carpenter, R. L., & Schultz, L. J. (1983, November). Reading assessment in the Army. (ARI Working Paper RS-WP-83-5). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Raines, J. (1983). Information paper: Literacy testing for mid-term reenlistment. Alexandria, VA: Education Directorate, The Adjutant General's Office, Department of the Army.

Sacher, J., & Duffy, T. (1978, March). Reading skill and military effectiveness. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario.

Sticht, T. G. (Ed.). (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Sticht, T. G. (1980, April). Literacy and vocational competence. (Report 2, No. 12). Washington, DC: U.S. Department of Labor.

Sticht, T. G. (1982). Basic skills in defense. (HumRRO-PP-3-82). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Beck, L. J., Hauke, R. N., Kleiman, G. M., & James, J. H. (1974). Auding and reading: A developmental model (AFHRL-TR-74-36). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Caylor, J. S., & James, J. H. (1978). Job-related reading tasks: Teaching marginally literate adults to read (HumRRO-PP-10-78). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. (1971). Literacy skill requirements in four military occupational specialties (HumRRO-TR-71-23). Alexandria, VA: Human Resources Research Organization.

Sticht, T. G., Hooke, L. R., & Caylor, J. S. (1982). Literacy, oracy, and vocational aptitude as predictors of attrition and promotion in the Armed Services (HumRRO-PP-2-82). Alexandria, VA: Human Resources Research Organization.

Valentine, L. D., Jr. (n.d.). Development of the Air Force Reading Ability Test. Unpublished manuscript, Air Force Human Resources Laboratory, Brooks Air Force Base, TX.

Ysseldyke, J. E., & Marston, D. (1982). A critical analysis of standardized reading tests. School Psychology Review, 11, 257-266.

APPENDIX

Table 1

Review of Commercially Developed Reading Tests

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY
Adult Basic Learning Exam (ABLE)[1] AND	Psychological Corporation	1974	To determine general education level of adults; to evaluate adult education programs	Adults with achievement levels grades 1–4, 5–9, 10–12; adults who have not completed formal 8th grade education.	Grade equivalents, based on equating ABLE to Stanford Achievement Test (SAT) using scores of students in grades 2–7; authors encourage use of local norms (percentiles and stanines).	Concurrent validity low to moderate based on: (1) administration of ABLE and SAT to elementary and junior high students (.60–.76), (2) administration of ABLE and SAT paragraph meaning to job corps group (.36–.72).
SelectABLE	Psychological Corporation	1974	To screen for determination of which ABLE level to administer	Same as for ABLE	Raw score, percent correct	INA[1]
Carver-Darby Chunked Reading Test[1]	Revrac Publication (developed by American Institutes for Research)	1972	To measure information stored during reading.	Grades 9–16 and adults	Efficiency, accuracy, and rate scores; no standardized scores.	Validity best viewed as pilot study
Gates-McGinitie Reading Test	Houghton Mifflin	1978	To assess three areas of reading	Grades 1–12.	Raw score, extended standard score, percentiles	Content validity suitable, may have high verbal-IQ component (has high correlation with Large Thorndike verbal IQ).
Gray Oral Reading Test	Bobbs-Merrill	1967	To assess oral reading skill	Grades 1–16 and adult.	Grade equivalents (total score only)	INA[1]
Metropolitan Achievement Test[2]	Psychological Corporation	1978	To assess achievement in a number of skill areas	Grades 2.5–9.5 for reading subtest	Scaled score, percentiles, stanines, grade equivalents.	Content validity good, no predictive or other validity mentioned in reviews, test authors suggest local curriculum validity be checked.
Nelson-Denny Reading Test	Houghton-Mifflin	1973	To assess reading for college placement and adult reading classes	Grades 9–16 and adults.	Percentiles, grade equivalents	Little data on concurrent and predictive validity, nothing on content or construct validity.
RBH Basic Reading and Word Test	Richardson, Bellows, Henry & Company	1969	To assess reading comprehension and vocabulary	Disadvantaged adults.	Percentiles, standard scores	Lacks convincing validity data, no statistics
RBH Test of Reading Comprehension	Richardson, Bellows, Henry & Company	1963	To test reading comprehension using six articles related to business and industry	Business and industry	Percentiles, standard scores.	Lacks convincing validity data, low validity coefficients (.17–.45)
Tests of Adult Basic Education (TABE)	CTB/McGraw-Hill	1976	To measure reading and arithmetic levels of adults using an adapted version of a children's test.	Adults reading at levels of children in grades 2–4 (Level E), 4–6 (Level M), 7–9 (Level D).	Grade equivalents based on California Achievement Tests (CAT), 1970.	Content validity based only on item selection procedures from CAT. Questionable validity. Concurrent validity .56 with GED test.
Wide Range Achievement Test Level II[2]	Jastak Assessment Systems	1978	To assess quickly three discrete areas of achievement	[illegible] years to adult.	Standard scores, graded equivalents, percentiles based on age	Content validity questionable

1 – Information not available
2 – Review concerns reading-related subtests only (e.g., reading, vocabulary, spelling)
3 – Timing for whole test

Table 1 (Continued)

Review of Commercially Developed Reading Tests

RELIABILITY	NORMS	GROUP VS. INDIVIDUAL	TIMING	SCORING	FORMS	SUBTESTS OR SCORE AREAS	ITEM QUALITY
Split-half reliability good to excellent for school group (.87–.95), job corps group (.85–.96), and adult basic education students (.91–.94). No information on test-retest or other reliability	Based on 6,000 elementary and junior high students, 800 job corps members, and 450 adult basic education students.	Group	Level 1 145 min Level 2 145 min Level 3 207 min [2]	Hand or Machine	A & B	Vocabulary, reading, spelling, arithmetic	Items appear technically well constructed. Most items relevant to adults. Reading items are cloze types and do not measure all facets of reading—just reading comprehension
KR-20 .84–.87, split-half .86–.91.	452 military recruits	Group	15 min	Hand	One Form	Total	Adequate
Reliability best viewed as pilot study—not high (.43–.81).	Totally unsatisfactory, means and standard deviations available on 41 "college student volunteers"	Group	25 min. optional	Machine	A & B	Efficiency, accuracy, rate	Divides sentences into chunks but without a consistent theoretical basis related to the reading act
Alternate form reliability (.72–.90)	Norming procedures good	Group	Level A, B 50 min Level C 50 min Level CS 7 min Level D 40 min Level E 44 min	Hand or Machine	1, 2, 3	Speed/accuracy, vocabulary comprehension	Adequate
Standard errors of measurement adequate	Norms based on very small samples (N = 40 per grade)	INA[1]	INA[1]	INA[1]	A, B, C, D	Comprehension, total	Items satisfactory to good, oral reading only
Internal consistency reliability .90+, standard errors of measurement excellent	Norm group carefully selected	Group	Level P-2 160–170 min.; Level E, 225–235 min.; Level I 270 min.; Level A 265 min [2]	Hand or Machine	J & K	Word knowledge, reading, total word analysis (or language, depending on test level), spelling, mathematics (computation, concepts problem solving, total)	Items written for children not adults. Content problems because content remote from adult life
High alternate forms reliability for most subtests, based on small samples, however	Standardization OK, adult norms based on cut time administration only	Group	30–35 min.	Hand or Machine	A & B	Vocabulary comprehension, total, rate	Mostly well constructed items
Lacks convincing reliability data (.86 probably an overestimate)	INA[1]	Group	25–30 min	Hand	One Form	Total	Items cover mostly vocabulary. Choice of vocabulary tested is questionable
Reliability data not adequate, no means and standard deviations given	Test probably was too easy for norm group, norms not adequately standardized	Group	25–25 min.	Hand	One Form	Total	INA[1]
KR-20 for reading test section .86–.94, test-retest for reading .79–.85	No adult norms, norms based on students in grades 2–9	Group	Level E 127 min Level M 209 min Level D 191 min [2]	Hand	3 & 4	Reading (vocabulary comprehension, total) Arithmetic (reasoning, fundamentals, total) Language (mechanics, spelling, total)—for top two levels Total (Locator test also available)	Items were written for children but have been revised to omit purely childish references
High reported reliabilities (.98) are suspect	No national norming sample. Norms developed from continuous age data. Identity and nature of norm group not clear	Part individual	20–30 min	Hand	One Form	Spelling, arithmetic, reading (based on "clinical factor analysis").	Questionable item development and quality

Table 2

Review of Selected Military Developed Reading Tests

TEST SERIES	DEVELOPER OR PUBLISHER	LATEST COPYRIGHT	MAIN INTENDED PURPOSE(S)	INTENDED TEST POPULATION	SCORES	VALIDITY
Air Force Reading Ability Test (AFRAT)[1]	U.S. Air Force	INA[2]	To assess reading ability among fair to excellent readers in the Air Force	5th grade through college level; not intended for poor readers	GE, percentile	Concurrent validity is .72 with California Achvt. Test, .75 with Nelson-Denny. Predictive validity moderate (.40 in predicting Officer Training School grades, -.13-.61 in predicting technical training grades). Comprehension is better predictor than vocabulary
Job Reading Task Tests[4]	U.S. Army	circa 1971-1973	To assess performance on Army job reading tasks with items based on 3 military occupational specialties (MOS), for research purposes only	Army enlisted personnel	Percent correct, reading grade level	Moderate (.64-.80) concurrent validity with standardized achvt. test. Lower concurrent validity with AFQT. Moderate correlations with course grades and job knowledge tests
Job Reading Task Tests[4]	U.S. Army	circa 1975	To assess performance on Army job reading tasks with items based on 6 MOS	Army enlisted personnel	GE, percentiles	Content validity good, based on items from empirically determined job rating tasks. concurrent validity ranges .69-.79 with USAFI RGL criterion.
Job Reading Tests[5]	U.S. Army	circa 1982	To assess performance on Army job reading tasks with items based on 6 MOS	Army enlisted personnel	Percentile	Validity data not convincing. Item validity appears adequate. Only information on test validity is that the content is drawn carefully from 6 MOS
U.S. Armed Forces Institute (USAFI) Reading Test (same as Metropolitan)	←					

1 - Information from Mathews (personnal communication, September 7, 1983) and Valentine (n.d.)
2 - Information not available from published source
3 - One form assumed unless more than one is specified
4 - Information from Sticht (1975)
5 - Information from Claudy & Taylor (1982)

Table 2 (Continued)

Review of Selected Military Developed Reading Tests

RELIABILITY	NORMS	GROUP VS. INDIVIDUAL	TIMING	SCORING	FORMS[3]	SUBTESTS OR SCORE AREAS	ITEM QUALITY
Internal consistency reliability approx .90 for whole test	Adult male and female Air Force and Army enlistees.	Group	50 min. total test	Hand or Machine	A, B	Comprehension, vocabulary, total.	Vocabulary and comprehension items above average in quality.
Test-retest reliability 75-80. No other reliability information.	INA[2]	Group	INA[2]	INA[2]	Three forms, cook, supply clerk, vehicle repairman.	Tables, standards and specifications, identification and description, procedural directions, check points, functional description.	INA[2]
KR-21 .93-.94, alternate forms .68-.76, SE of msmt. .02-.96	750 young adult male Army recruits	Group	1 hr +	Hand	A, B, C	Using index to locate information, extracting information from tables and narrative prose, following procedural directions	Free-response, fill-in-blank items cause slow scoring
INA[2]	Norming appears adequate	Group	30-40 min.	Machine	A, B, C	Locating job information in tables, index, graphs, and narrative; forms completion	Item statistics appear adequate. Multiple choice items