

DOCUMENT RESUME

ED 109 265

95

TM 004 864

AUTHOR Nafziger, Dean H.; And Others
 TITLE Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments.
 INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
 SPONS AGENCY Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
 PUB DATE Jun 75
 CONTRACT OEC-300-75-0098
 NOTE 108p.

EDRS PRICE MF-\$0.76 HC-\$5.70 PLUS POSTAGE
 DESCRIPTORS Adult Literacy; *Adults; Criterion Referenced Tests; Evaluation; Evaluation Criteria; *Functional Illiteracy; *Reading Tests; Standardized Tests; Testing; *Test Reviews; *Tests; Test Validity

ABSTRACT

Currently available measures of functional literacy for adults are reviewed and evaluated. This report concentrates on tests that are referenced to literary skills important to an adequately functioning adult, such as life skills, coping skills, etc. Because functional literacy has frequently been defined in terms of a grade level equivalent or some other norm, adult reading tests referenced to a norm group are also included. A common set of 40 criteria categorized under four main headings are used: measurement validity, examinee appropriateness, technical excellence, and administrator usability. The report provides teachers and administrators in Right-to-Read and other adult education programs a reference for use in identifying and judging the value of tests available for assessing adult functional literacy. To increase its utility as a reference, summaries of a number of tests designed for adults are included. The report consists of six major parts: (1) Problems in Defining and Measuring Literacy; (2) Test Identification; (3) Evaluative Criteria; (4) Test Reviews; (5) Test Evaluations; and (6) Summary. Because many tests of functional literacy are newly developed or still being developed, there may be tests which should have been--but could not be--included in this report. No one set of criteria is appropriate for judging all tests. Thus, these test evaluations must be interpreted with respect to the intended use of each test. (Author/RC)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

PL 9265

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to T.M. C.S.

In our judgment, this document is also of interest to the clearing houses noted to the right. Indexing should reflect their special points of view.

Tests of Functional Adult Literacy

an evaluation of currently available instruments

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT THE NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

TM004 864

ERIC
Full Text Provided by ERIC

Tests of Functional Adult Literacy:
An Evaluation of Currently Available Instruments

Dean H. Nafziger
R. Brent Thompson
Michael D. Hiscox
Thomas R. Owen

Assessment Projects
Northwest Regional Educational Laboratory
710 S. W. Second Avenue
Portland, Oregon 97204

June 1975

Prepared for the
U. S. Office of Education
Office of Planning, Budget and Evaluation
Contract No. 300-75-0098

The work presented or reported herein was performed pursuant to a Contract from the U. S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U. S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

ACKNOWLEDGMENTS

Special thanks are due many people who assisted in the project and in the preparation of this report. The authors are grateful to Robert C. Hall of the U. S. Office of Education for his guidance and assistance during all phases of the evaluation. We wish to thank W. Aubrey Gardner and James R. Sanders for their assistance in planning and conceptualizing the study, and for their guidance and consultation throughout the project. Also, we appreciate the work of Diane S. Thompson and Ann Helmick in preparing parts of the report; and of Robert J. Silverman and William J. Wright, who provided many valuable suggestions on an earlier draft of the manuscript.

The contributions of Vicki Spandel in editing the report and of Julie Stange, Peggy Hootstein, and Trena Jabbour in their careful preparation of the manuscript are especially appreciated.

Finally, we express our gratitude to many professionals in Right-to-Read and adult education programs at the federal, state, and local levels who answered our written and oral requests for information. Naturally, the authors accept all responsibility for the contents of the report; and no responsibility for any shortcomings inherent in this report should be assigned to those who gave generously of their time in its preparation.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iii
TESTS OF FUNCTIONAL ADULT LITERACY: AN EVALUATION OF AVAILABLE INSTRUMENTS	1
Organization of the Report	3
PROBLEMS IN DEFINING AND MEASURING FUNCTIONAL LITERACY	5
Estimates of the Extent of Literacy	7
Literacy Definitions	11
Functional Literacy	13
Choosing Tasks to be Measured	14
Summary	15
TEST IDENTIFICATION	17
EVALUATIVE CRITERIA	21
Measurement Validity	24
Examinee Appropriateness	27
Technical Excellence	31
Administrative Usability	32
TEST REVIEWS	39
CRITERION-REFERENCED FUNCTIONAL LITERACY TESTS	45
Adult Performance Level Functional Literacy Test (APL)	45
Basic Reading Skills Mastery Test	48
Reading/Everyday Activities in Life (R/EAL)	50
Wisconsin Test of Adult Basic Education (WITABE)	53
STANDARDIZED TESTS	57
Adult Basic Learning Examination (ABLE), Level I	57
Basic Occupational Literacy Test (BOLT), Fundamental Level	60
General Educational Performance Index (GEPI)	63
SRA Reading Index	66
Tests of Adult Basic Education (TABE), Level E	69

TABLE OF CONTENTS (Continued)

	Page
INFORMAL TESTS	72
Adult Basic Reading Inventory	72
Cyzyk Pre-Reading Inventory	76
Harris Graded Word List and the Informal Textbook Test	78
Idaho State Penitentiary Informal Reading Inventory	81
An Informal Reading Inventory for Use by Teachers of Adult Basic Education	85
Individual Reading Placement Inventory	88
Initial Testing Locator Tests	91
Reading Evaluation--Adult Diagnosis (READ)	93
 TEST EVALUATIONS	 97
 SUMMARY	 109
 FOOTNOTES	 115
 BIBLIOGRAPHY	 121

TESTS OF FUNCTIONAL ADULT LITERACY:
AN EVALUATION OF CURRENTLY AVAILABLE INSTRUMENTS

Adult illiteracy was recently designated a major target area of the Right-to-Read program in the United States. The extent of their commitment to reduce adult illiteracy is reflected in a national goal of the Right-to-Read program: To eliminate functional illiteracy by 1980 among 90% of the population over 16 years of age.¹ In particular, Right-to-Read seeks to teach necessary reading skills to adults who have not been successful participants in society. Increasing emphasis on functional literacy has led to a proliferation of reading programs designed to teach reading tasks important to social survival. The desire to determine the efficacy of these programs has led, in turn, to a need for instruments that measure functional literacy.

The purpose of this report is to review and evaluate currently available measures of functional literacy. The report concentrates on tests that are referenced to literacy skills important to an adequately functioning adult. These skills have been referred to as life skills, survival skills, coping skills, and so on.² Because functional literacy has frequently been defined in terms of a grade level equivalent or some other norm, adult reading tests referenced to a norm group are also included. A common set of criteria, which address characteristics important for any test, were used to evaluate all tests included in this report.

The report summarizes the current availability of tests of adult functional literacy. It is also intended to provide administrators and teachers in Right-to-Read and other adult education programs a reference for use in identifying and judging the value of tests available for assessing adult

functional literacy. To increase its utility as a reference, the report includes summaries of a number of tests designed for adults.

It is also important to note what this report does not attempt to provide. First, the contractual mandate of this study was to review and evaluate only those tests developed strictly for adults. Therefore, this report does not provide a comprehensive listing of all tests used in measuring adult reading ability, since many such tests were developed for children; excellent resources that list these tests are already available.³ Second, this report does not identify and evaluate tests which are inextricably bound to specific instructional materials, curricula, or programs; only tests appropriate for general use are listed.

In addition, the report has certain limitations. Because many tests of functional literacy are newly developed or still being developed, their existence is not widely known. Despite the national mail survey that preceded this report, some such tests may not have been identified and included. Also, some authors requested that their tests be excluded from consideration until further work on them was completed. As a result, there may be tests which should have been--but could not be--included in this report.

Another limitation concerns the fact that no one set of criteria is appropriate for judging all tests. Most tests have some unique characteristics that bring into question the applicability of some criteria. This problem is intensified when different kinds of tests--e.g., norm-referenced and criterion-referenced--are being judged by the same criteria. Moreover, standard criteria may not reflect the interests or priorities of a particular audience for evaluation results. Thus, the reader must interpret the test evaluations in this report with respect to the intended use of each test.

A further limitation lies in the fact that the tests were evaluated according to existing public data provided by the test authors or publishers. On criteria for which no data were available, tests received unfavorable evaluation. While this approach was judged by the report authors to be the most equitable known, it has the disadvantage of appearing overly stringent in relation to tests still in the early stages of development.

Organization of the Report

Following this introductory section, this report consists of six major parts:

1. The Problems in Defining and Measuring Literacy. This section includes estimates of the extent of illiteracy, definitions of literacy, notions of functional literacy, and problems in choosing tasks to measure literacy.
2. Test Identification. This section includes a discussion of three major activities undertaken to identify tests: a literature search, requests to publishers and professionals involved in adult education, and telephone and personnel interviews with persons active in teaching and measuring adult reading.
3. Evaluative Criteria. This section includes a discussion of the 41 criteria used in evaluating the tests, and an explanation of how these criteria are categorized under four main headings: measurement validity, examinee appropriateness, technical excellence, and administrative usability.
4. Test Reviews. This section presents descriptive reviews of the tests, which are grouped into three subsections: criterion-referenced functional literacy tests, standardized tests, and informal tests.
5. Test Evaluations. The section presents evaluations of the tests, which are grouped according to the same overall organization as the test reviews.
6. Summary. This section notes some general strengths and weaknesses of different types of tests. It also describes continuing work related to the measurement of functional literacy being conducted by three groups.

PROBLEMS IN DEFINING AND MEASURING FUNCTIONAL LITERACY

Conducting adult literacy assessment necessarily requires an understanding of what literacy is. Achieving that understanding is difficult because literacy is not a solitary trait; it comprises many sub-skills. Also, one might be considered literate (able to comprehend) in some content areas, though not in others.

The multifaceted nature of literacy has often been glossed over through the use of such composite scores as standard scores and grade level equivalents. For example, one might say, "He is reading at grade level 7.2"; in a very general way, this kind of normative statement relates a particular person's performance on some unknown reading task to the performance of others at a particular--in this case educational--level. It is not usually clear how this level of performance would relate to any other possible literacy tasks. One could argue that, with young children, general reading ability that can be applied to a broad range of tasks is most important to consider; but with adults, especially those who are only marginally literate, one is more concerned with whether they can perform particular sets of life- or work-related literacy tasks.

Estimates of the Extent of Illiteracy

Estimates of the extent of illiteracy in the United States vary considerably, depending somewhat on the method of assessment used. The Census Bureau considers literate anyone 14 years of age or older who has completed sixth grade.⁴ Those who have not completed the sixth grade are asked whether they can read and write a simple message in any language; if they say, "Yes," they are considered literate. Based on this method, it is estimated that approximately one percent of those aged 14 years and older are illiterate.

However, the self report feature of the inquiry, the concern that the description "simple messages" may not be adequate, and the uncertainty about the relationship of literacy to graduating from the 6th grade together cast considerable doubt on the Census Bureau's method of estimating literacy--except perhaps as a way of deriving a lower bound estimate.

Assessing levels of literacy using grade level equivalent scores on some type of reading test is a common practice. The National Center for Health Statistics has conducted a survey using their Brief Test of Literacy, which shows that 4.8 percent of individuals 12-17 years old score below the average 4th grader on the instrument and can therefore be regarded as illiterate.⁵ This method of assessment is not useful because it is uncertain what specific performances are implied by success on the test.

Indices of literacy such as those discussed here may be useful at the "first guess" level. They are inadequate beyond that point, however, either because they do not relate directly to literacy, or because they do not permit inferences about what sorts of functional competencies given levels of literacy imply.

To obtain a more useful estimate of the extent of illiteracy some recent work has been done to define what literacy-related tasks adult members of this society must perform, and to build assessment instruments that measure performance on those tasks. Certainly the best publicized of these attempts was made by the Harris survey team, who were commissioned by the National Reading Center to conduct a study of adult functional illiteracy.⁶ They asked respondents to read and fill in the appropriate information on five forms--Application for Public Assistance, Application for Medicaid, application for a driver's license, personal identification form, and a personal loan application. Using the criterion of 90 percent correct responses on

these forms, Harris reports that 13 percent of their sample, or an estimated 18.5 million Americans, fell below that level--that is, were marginally literate to functionally illiterate in terms of ability to perform these tasks. (It has been asserted by some that these data are statistically incorrect, and that the correct estimate, based on the 1970 Harris survey, should have been 6.5 percent below the literacy level.)⁷ While the range of literacy tasks employed in this limited, the tasks do represent some of the common literacy tasks which adults are required to perform.

A second survey, conducted by Harris in 1971, explored respondents' ability to successfully answer straightforward questions about newspaper employment advertisements. Ninety-two percent of the total sample got all nine of the questions correct, although only 70 percent of all Blacks tested got nine correct. Survey personnel obtained similar results using classified housing advertisements; 88 percent of those surveyed got all items correct. Blacks averaged 67 percent correct.⁸

Thus, it appears that, using several literacy tasks chosen simply as examples, the national level of marginal to complete illiteracy might encompass around ten percent of the population, and might be much higher among some minorities. These data also show higher illiteracy rates for low income and low education groups. As instruments for assessing literacy, however, neither the representativeness of the tasks nor the performance levels used, have any empirical support.

In order to produce a valid set of tasks for assessing adult competencies, Norvell Northcutt of the Adult Performance Level (APL) Project conducted an extensive literature search, surveying governmental agencies and foundations to determine the characteristics of successful and unsuccessful adults; and interviewed adults who were under-educated and underemployed, employers, and

personnel specialists. The necessary skills identified during this 1975 search can be grouped into the following four areas: (a) communication skills, (b) computational skills, (c) problem solving skills, and (d) interpersonal skills. Northcutt also identifies five general knowledge areas: (a) occupational knowledge, (b) consumer economics, (c) community resources, (d) government and law, and (e) health.⁹

Because these skills demand much more than the ability to use or comprehend written material, they do not fit comfortably within the concept of literacy. Therefore, the APL staff substituted the term "functional competency" for "functional literacy."

Using national samples, the Adult Performance Level Project has determined that as many as 20 percent of the adult population are functionally incompetent. Indeed, in one of the skill areas, computation, it appears that one-third of U.S. adults may be functionally incompetent. Only 70 percent of those surveyed could indicate the proper number of exemptions on a W-4 form when given the number of dependents. On a task requiring the respondents to match personal characteristics with job requirements in an employment advertisement, only 62 percent succeeded. More than 20 percent of those surveyed could not draw the proper conclusions from a notice of a store's check cashing privileges. Overall, the APL project staff estimate that more than 20 percent of U.S. citizens are functionally incompetent at reading--a figure which contrasts sharply with the results of earlier surveys.

It would appear that as the tasks used in literacy assessment instruments become more like "real world" tasks in the sense of requiring composite skills, estimates of the extent of illiteracy increase proportionately. One might expect this. It simply indicates that the more marginal a person's

skills, the more likely he is to fail at tasks for which the requisite skills are interdependent.

Bormuth has stated that it is important to carefully derive both the literacy behaviors and the acceptable levels for success.¹⁰ The Northcutt study appears to have surpassed previous studies on the first item, but is still arbitrary in assigning criterion levels of success.

Bormuth's work includes an example of a different task which has been used to assess the extent of functional literacy in a particular population. In 1969, he prepared cloze tests on several newspaper passages and tested a sample of high school seniors.* He set a level of 35 percent correct as a criterion for adequate performance on the test. The 35 percent criterion is based on a conclusion Bormuth drew from earlier research: that people with cloze scores of 35 percent or less were able to extract very little meaning from the passage. Only 65 percent of the sample correctly answered 35 percent of the cloze terms.

Literacy Definitions

The preceding discussion offers a general perspective of literacy based on the efforts of those who sought to assess levels of literacy. The estimates of illiteracy given in that section vary because there is little consensus about what constitutes literacy. The purpose of this section is to further examine the differences among conceptions of reading and literacy by presenting some common definitions. Consider the following definitions of the reading process. Bower commented that

Reading is a sequential process in which ongoing processing is affected by prior processing and will determine future processing.¹²

*To prepare a cloze passage one deletes every nth word and it is the task of the reader to fill in the missing words.

In a similar statement, Goodman said that the reader

...concentrates his total prior experience and learning on the task, drawing on his experiences and the concepts he has attained as well as the language competence he has achieved.¹³

Both emphasize the role of prior knowledge in facilitating the reading process, and couch their definitions in descriptions of what an individual does.

Gibson offers a similar description of reading:

There are several ways to characterize the behavior we call reading. It is receiving communication; it is making discriminative responses to graphic symbols; it is decoding graphic symbols to speech; and it is getting meaning from the printed page.¹⁴

These definitions of reading refer primarily to information processing mechanisms that the reader must or may employ, and say little about the nature of reading itself.

Literacy, in contrast to reading, implies both basic reading skills and socially appropriate reading behavior, and any definition of literacy must incorporate both. Bormuth offers the following comprehensive definition:

In the broadest sense of the word, *literacy* is the ability to exhibit all of the behaviors a person needs in order to respond appropriately to all possible reading tasks.¹⁵

Of course, no one is literate to this extent. If literacy is to be a realistic goal of an educational program, it must be defined as some subset of the total set of reading tasks and the behaviors required to accomplish those tasks.

Bormuth suggests that this subset be selected on the basis of economic, social, cultural, and political benefits to the individual or his society--that is, for pragmatic reasons.

In recent assessments, tasks assessing literacy have been chosen more for their social utility than for their relationship to presumed underlying dimensions of reading. This is consistent with the theory that literacy involves more than reading skills alone.

Functional Literacy

The term "functional literacy" connotes reading for a purpose--a purpose in some way related to social utility. William S. Gray defines functional literacy as "the ability to engage effectively in all those reading activities normally expected of a literate adult in his community."¹⁶ This definition, while circular, does emphasize the fact that certain tasks are required of adults by members of their community. The U. S. Office of Education has defined a literate person as

...one who has acquired the essential knowledge and skills in reading, writing, and computation required for effective functioning in society, and whose attainment in such skills makes it possible for him to develop new aptitudes and to participate actively in the life of his times.¹⁷

U.S.O.E. has operationalized this definition by suggesting that adults be able to perform the following tasks:

- o Read and understand all sections of a newspaper, with particular emphasis on the classified and advertisement sections
- o Read the drivers license test in any state
- o Read and understand voter registration instructions
- o Read and comprehend the key features of popular business contracts such as those issued by used car dealers, furniture stores, clothing shops, and auto repair dealers
- o Read labels on such household items as groceries, medicines, recipes, machine instructions, etc.
- o Read the materials necessary to perform jobs classified as entry level
- o Read personal letters, bills
- o Read and follow public instructions such as road and building signs
- o Read and use the telephone directory
- o Read and complete job application forms
- o Read and comprehend business letters from debtors and creditors¹⁸

Sticht defines functional literacy as "a possession of those literacy skills needed to successfully perform some reading task imposed by an external agent between a reader and a goal the reader wishes to obtain."¹⁹ He points out that this excludes such reading activities as reading for pleasure. Also he differentiates between reading to learn a job and reading to do a job. As a rule, the former requires a higher level of literacy than the latter.

From these definitions and operationalizations of the concept of functional literacy, one can infer that some of the major assessment problems relate to creating instruments which reflect special concerns and help establish the importance of certain tasks.

Choosing Tasks to be Measured

One difficulty in choosing tasks to assess functional literacy lies in accurately identifying the skills involved. Carver has argued that some of the higher order comprehension items in reading inventories may relate more to thinking than to reading itself.²⁰ Furthermore, successfully completing some comprehension items might also relate to one's general knowledge of the subject matter.

Carver also suggests that if it is actually the ability to reason that is being assessed, the evaluative judgment one makes about a reading program may be distinctly unfair.²¹ The same argument may be advanced regarding external knowledge or experience and their relationship to reading. One may choose to broaden a reading program's educational goals, basing them on performance tasks used in functional literacy assessments. Bormuth warns, however, that such an approach may commit a program to a much more difficult undertaking than anyone realizes.

Though traditional norm referenced reading tests--particularly the comprehension sections--may be measuring intelligence rather than reading skills, that problem cannot be categorically solved simply by shifting to criterion-

referenced tests. The tasks themselves determine what is being measured regardless of whether norms are constructed. MacGinitie argues that:

Giving a score that refers to some criterion rather than to a norm group does not absolve the test maker from showing that separate component scores index meaningful skill levels or separately measureable skills.²³

Unless one is very careful, one may be actually assessing language skills, intelligence, or general knowledge--even when using a criterion-referenced instrument.

Summary

Because no standard definition of literacy exists, estimates of the extent of illiteracy in the United States vary widely. Recent use of the term literacy connotes the ability to perform functional reading tasks--i.e., tasks which are important for successful participation in society. Because these definitions concern the attainment of a set of minimal skills, they imply the need for criterion-referenced tests that will measure the attainment of such skills. One consideration in constructing criterion-referenced tests of functional literacy is selecting tasks that are important for adequately functioning adults. A primary purpose of this report is to examine the extent to which measures of adult functional literacy meet this and other considerations.

TEST IDENTIFICATION

A major part of preparing this report was identifying tests to be included. Tests and background information were gathered in three ways. First, a literature search was conducted to identify tests, test reference books, and articles on current test development efforts. Many tests and some major test references, such as those noted earlier, were identified during this activity.

Second, requests for information were mailed to publishers of adult literacy materials and to professionals active in adult education. All known publishers of tests of adult education materials were contacted. They were asked if all tests they had available for measuring functional adult literacy could be purchased. Requests for tests and information from professionals were sent to state Right-to-Read Directors, State Directors of Adult Basic Education (ABE), USOE Staff Development Directors and Program Officers for ABE, and directors of programs for adult educators in colleges and universities.

One hundred twenty-eight (60 percent) of the 212 professionals contacted returned questionnaires. In addition, several individuals made copies of the questionnaire so that other members of their staff could respond as well. Forty-four (56 percent) of the 79 publishers contacted responded to the solicitation letters. Follow-up letters were sent as a part of this solicitation effort.

Third, telephone and personal interviews were conducted with individuals active in teaching and measuring adult reading. Those interviewed included developers of measurement instruments, coordinators of adult education programs, teachers of adults, and specialists in reading measurement. These interviews were conducted for varying reasons--to help identify tests, to gain more

information about tests already identified, and to obtain information about criteria to be used in evaluating the tests.

As a result of these activities, approximately 150 tests used in measuring adult reading ability were identified. Most were designed for elementary and secondary school students; less than 30 of the tests collected had been designed specifically for use with adults.

By contractual mandate the project focus was on tests developed for adults. Therefore, many commonly used tests were excluded because they were designed for children rather than adults. The Gray Oral Reading Test and the Stanford Diagnostic Reading Test are examples of widely used tests which were excluded from this report because they were originally designed for children.

EVALUATIVE CRITERIA

Numerous sources were consulted to identify or develop criteria for test evaluation. The criteria adopted for this report relied heavily on the criteria used by the Center for the Study of Evaluation (CSE) at UCLA in their comprehensive test evaluations. The CSE criteria offered two major advantages. First, they represented a complete compilation of generally accepted test standards. Second, they had been extensively used by CSE in evaluating tests; weaknesses and ambiguities had, therefore, been largely eliminated.

Even so, the CSE criteria presented some problems with respect to measuring adult literacy. For example, the CSE criteria included one judgment that favors tests which are group administered. But for test-anxious adults, a group administered test may not always be the better choice. Therefore, the criterion awarding a point for group administration of tests was dropped.

Furthermore, since the CSE criteria were designed for application to a wide range of tests, certain specific concerns in measuring functional literacy could not be addressed. Thus, it was necessary to add questions such as: "Are there scales of performance on real-life skills (e.g., map reading, understanding want ads, etc.)?"

Like CSE's criteria, our criteria focused on four major areas: measurement validity, examinee appropriateness, technical excellence, and administrative usability. Each of these areas consisted of several individual criteria. Tests were assigned points indicating the extent to which they met each criterion; then the points were totaled for each of the four areas. Finally, an area grade of good, fair, or poor was assigned, based upon the total points obtained for the criteria within the area. Within each area the numbers of

points designating the total grade (i.e., good, fair, poor) were chosen in such a way that most of the criteria would have to be attained at the maximum level in order for the test to obtain a high grade for the area.

The criteria were applied to each test as a whole, or subtest by subtest. Each test was independently evaluated by at least two people; differences in ratings were adjudicated by a third person. The evaluators all had previous experience or training in educational measurement. They were trained in the use of the criteria, and their judgments were checked for consistency and accuracy during the training. In addition to the evaluations, a descriptive review was prepared for each test. These reviews describe the tests, and summarize the administration, scoring and interpretation procedures and the available technical data.

On the following pages, the criteria used to evaluate the tests are described. Evaluative decisions were based on information presented in the manuals and supplements accompanying published tests, or on information concerning unpublished tests supplied by test authors at our request. No attempt was made to verify available information. When needed information was not available and was not readily inferrable, a test was credited with 0 points (the lowest rating) on the relevant criterion.

Credit and appreciation are due to Ralph Hoepfner and others at the Center for the Study of Evaluation, whose pioneering work we have freely borrowed and adapted in arriving at the criteria which follow.²⁴ Of course, we accept sole responsibility for the final set of criteria used in this project and for their application.

Measurement Validity

- a. Is information provided to indicate a rigorous selection of items and careful sampling of the behavior domain?

Such information was considered adequate, provided references on the construction of the test were included. If the procedures used in developing test specifications and items were described in some detail, the test was credited with 2 points; if reference was made indicating the use of a specific, rigorous item selection procedure, the test was credited with 1 point; if no information was provided on item selection, the test was credited with 0 points.

b. Were any empirical procedures used for screening or selecting the items?

Empirical procedures include item analyses, juries of experts, item difficulties, criterion-group analyses, or factor analyses. If more than one method was reported in some detail, the test was credited with 3 points; if it was stated that more than one method had been used, or if one method was reported in some detail, the test was credited with 2 points; if it was stated that one method had been used, the test was credited with 1 point; and if no information was given, the test was credited with 0 points.

c. Are the items tied into specified objectives or criteria?

If the test items were generally related to specified objectives or criteria (such as tasks from a task analysis), the test was credited with 1 point. If items were not generally so related, or if objectives or criteria were lacking, the test was credited with 0 points.

d. Does the construct or type of behavior that the test purports to measure have a supportive base in linguistic, educational, psychological, or learning theory?

This criterion was applied to statements describing the theoretical basis of the test or to statements justifying the existence of the test

(e.g., "oral reading scores correlate only slightly with silent reading test results; therefore, we felt the need for a separate oral reading test, which could possibly be used as part of a more comprehensive testing effort"). If the test included such a statement, it was credited with 1 point; if not, it was credited with 0 points.

e. Has the test been employed in experiments or evaluations?

If the test scores in such experiments appeared to have yielded meaningful results, the test was credited with 1 point; if not, the test was credited with 0 points.

f. Are any concurrent validity studies (demonstrating correlation with some criterion measures obtained at the same time as the test) reported or specifically referred to in which the criteria (not other scores of the same test) are related in a meaningful way to the goal behavior to which the test was assigned?

If the criterion behavior was relevant and the coefficient was .70 or more, the test was credited with 2 points; if the coefficient was between .30 and .70, or the criterion behavior was not convincingly relevant, the test was credited with 1 point; if no study was reported, coefficients were low, or the criterion was clearly irrelevant to the nature of the test, 0 points were credited.

g. Are any predictive validity studies (the criterion behavior--usually success in some area--is obtained after a stated time interval) reported or specifically referred to in which the criteria was related in a meaningful way to the goal behavior to which the test was assigned?

If coefficients at or above .70 were reported with relevant criteria and a time interval of one month or more, the test was credited with 2

points. If only moderate coefficients (.30 to .70) were reported, or the criteria were of questionable meaningfulness, the test was credited with 1 point. If no study was reported or referenced, or the study was patently irrelevant, the test was credited with 0 points.

The Measurement Validity ratings were summed for a total rating, varying from 0 to 12 points. These ratings were translated into letter grades of G (good, 10 to 12 points), F (fair, 6 to 9 points), and P (poor, 0 to 5 points).

Examinee Appropriateness

- a. Does the test justify itself by explaining to the examinee in an honest manner its purpose, intent, or recommended use?

Misuse of test scores was not considered here, since such misuse is impossible to control. If the test (usually the test instructions) specifically stated the real or suggested purpose, intent, or use of the test, or if the manual suggested that such a justification be given in each situation, the test was credited with 1 point. If no purpose, intent, or use was specified, if the purpose or intent was disguised or concealed, or if examinees were led to adopt ineffective test-taking strategies, the test was credited with 0 points. This criterion was evaluated rather liberally in most cases, so that a test whose instructions began, "this is a test of your ability to spell..." was given credit for justification.

- b. Are the test items personally inoffensive and appropriate in terms of difficulty for adults in basic education or similar settings?

If all the items appeared inoffensive and reasonably appropriate in difficulty level, the test was credited with 2 points. If most items appeared appropriate or there were few serious typographical errors, the

test was credited with 1 point. If many of the items were judged inappropriate because (1) they were ambiguous or misleading, they lacked demonstrably correct or incorrect alternatives, they were stated in unnecessarily complex language, or (2) they were personally offensive, inappropriate or offensive to special groups, too simple, or intellectually insulting in simplicity, the test was credited with 0 points.

c. Are the items relevant and interesting for adult examinees?

This rating was made somewhat independently of test content so that inherently interesting subject matter did not necessarily profit from this rating. One way to rephrase this rating would be to ask: Given the nature of the subject matter, have the items been made as relevant and interesting as they could be? If the items were judged relevant and interesting, the test was credited with 1 point. If they were judged irrelevant or dull, the test was credited with 0 points.

d. Are test instructions oral or written?

The issue here was whether successful performance on the test required only the behaviors being measured by the test items or whether competency in reading test instructions was confounded with the behaviors purportedly being measured. If instructions were either completely oral or were supposed to be read aloud in addition to being written out for examinees, the test received 1 point; if not, it received 0 points.

e. Are test instructions appropriate and comprehensible?

The instructions, either read by or to the examinees, were inspected for appropriateness of orientation, tone, syntax, and vocabulary. If the instructions exhibited appropriateness and comprehensibility on all counts, the test was credited with 1 point; if not, it was credited with 0 points.

f. Are the instructions comprehensive in their description of task aspects?

The question addressed was whether the instructions clearly and precisely described all aspects of the tasks the test measured, or left necessary issues unanswered or unaddressed. If all aspects were described clearly and precisely, the test was credited with 1 point. If descriptions were unclear or incomplete or left issues unanswered, the test was credited with 0 points.

g. Do the test instructions provide illustrative sample items?

If the instructions included sample items that effectively clarified and accurately illustrated the task(s) involved in the test in such a way that they were truly representative of the format and difficulty of test items the test was credited with 1 point. If there were no sample items, or if sample items presented were not representative in format or difficulty, the test was credited with 0 points.

h. Do the test pages (or materials) exhibit good layout designed to facilitate perception?

Test layout was examined for effective use of perceptual organizers, such as adequate white space, regularity of item form, symmetry, clarity, and continuity. If the test page layout was clear and helpful, the test was credited with 1 point. If the layout was unclear or confusing, the test was credited with 0 points.

i. Is the physical appearance of the test of high quality?

For this rating, attention was given to the quality (bold, up-to-date) of the print and illustrations in printed tests, and the quality of sound in auditory or taped tests. If the quality was judged high, the test was credited with 1 point; if not, the test was credited with 0 points.

j. Are oral instructions or items standardized?

To meet this criterion, tests with oral instructions or oral items (such as "language potential" items) needed a standardized script designed to be read aloud, or a recorded version, such as a cassette tape. If the test had one or the other, it was credited with 1 point; anything short of this--such as merely suggesting topics to be mentioned to the examinee--was deemed insufficient and the test was credited with 0 points.

k. Is there coherence between item stems and answers?

If item stems, their alternatives, and their answers appeared as a unit, in some way adjacent or "belonging to each other," the test was credited with 1 point. If the separate components of any item(s) appeared not to belong to each other, and therefore demanded untangling, the test was credited with 0 points.

l. Are the time and pacing of the test appropriate?

Tests purporting to be power tests either had to be untimed, or had to furnish evidence that 90% or more of the validating group attempted all items. If a test met these conditions, or was appropriately paced, it was credited with 1 point; if not, the test was credited with 0 points.

m. What is the mode of examinee response?

No points were assigned for this information, but the test or subtest being evaluated was described as requiring oral (Or), written (Wr) or mixed (Mi) responses.

n. Is there a simple and direct connection between the item stem and the examinee's recording of a response?

If the mode of responding was especially simple for the examinee, such as oral responses, or marking or writing directly on the test form,

the test was credited with 2 points; if the test used standard separate answer sheets, it was credited with 1 point; if the test was complicated by the need for more than one step to get from item to answer, it was credited with 0 points.

The Examinee Appropriateness ratings were summed for a total rating, varying from 0 to 15 points. These ratings were translated into letter grades of G (good, 12 to 15 points), F (fair, 8 to 11 points), and P (poor, 0 to 7 points).

Technical Excellence

a. Does the test have alternative-form reliability?

The correlation between alternate forms of a test is the subject of this evaluation. If the appropriate coefficient was .90 or above, the test was credited with 3 points; if .80 to .90, the test was credited with 2 points; if .70 to .80, the test was credited with 1 point; and if less than .70, the test was credited with 0 points.

b. Does the test exhibit stability?

The consistency of scores over time spans of one month or more, as measured by test-retest reliability, is the subject of this criterion. If the appropriate coefficient was .90 or more, the test was credited with 3 points; if .80 to .90, the test was credited with 2 points; if .70 to .80, the test was credited with 1 point; and if below .70, the test was credited with 0 points.

c. Does the test exhibit internal consistency?

The consistency of items or parts within a part as measured by split-half or Kuder-Richardson formulas was the focus of this criterion.

If the appropriate coefficient was .90 or more, the test was credited with 2 points; if .80 to .90, the test was credited with 1 point; and if below .80 the test was credited with 0 points.

d. Can the testing procedures be duplicated?

A test was deemed more desirable if the procedures for administration, scoring and interpretation were sufficiently standardized so that procedures could be duplicated or replicated from the validating group. If the test provided uniformity of procedure for administration and scoring, the gross characteristics of the standardization group were replicable, and the materials, time limits (where applicable), oral instructions, and preliminary demonstrations were precisely delineated, the test was credited with 1 point; if not, the test was credited with 0 points.

The Technical Excellence ratings were summed for a total rating, varying from 0 to 9 points. These ratings were translated into letter grades of G (good, 6 to 9 points), F (fair, 3 to 5 points), and P (poor, 0 to 2 points).

Administrative Usability

a. Who should administer the test?

If regular program personnel, such as a teacher or aide, could read the instructions, establish rapport, and conduct the pacing, the test was credited with 1 point; if special personnel--such as a reading specialist--were required, the test was credited with 0 points.

b. How long does it take to administer the test?

If the test could be given in twenty minutes or less, including instructions, it was credited with 1 point; if not, the test was credited with 0 points.

c. Is the manual clear and complete?

This criterion focused on three aspects of the test manual: discussion of the purpose, uses and limitations of the test; clear administering and scoring directions; and description of test development and validation. If the manual's discussion, directions, and descriptions were clear and complete, the test was credited with 1 point; if not, it was credited with 0 points.

d. How many administrators or observers are needed to administer the test?

If not more than one administrator or observer was needed, the test was credited with 1 point; if more than one was needed, the test was credited with 0 points.

e. How easy and objective is the scoring?

If the scoring was objective and simple, using a scoring guide, stencil, or template, or other straightforward process such as answer sheet or matching stencils, or if machine scoring was available, the test was credited with 2 points. If the scoring was objective but difficult, involving more than a stencil or template, such as scoring a passage written by the examinee for specified content, the test was credited with 1 point. If the scoring was subjective, requiring the scorer to make a non-trivial judgment, the test was credited with 0 points.

f. Who can interpret the test scores?

This rating examined whether regular teaching staff could interpret the test. The answer to this question was either found in an explicit statement in the test manual, or else was implied from the common and simple conversion system for the scores. If the score could be interpreted

by teaching staff, the test was credited with 1 point; if not, the test was credited with 0 points.

g. How great is the range of complexity or difficulty of the test?

Tests using some kind of grade equivalent scheme of reporting or organizing content, and having a spread of three years or more from the lowest to highest scores obtainable, or from the easiest to most difficult materials, were judged to have a reasonably extensive range, and were credited with 1 point. For tests not using grade equivalent schemes, if the validating group had a spread of three years or more on an external criterion task, or if the material in the test was organized around an extensive hierarchy (or hierarchies) of tasks, the test was credited with 1 point. Otherwise, the test was credited with 0 points.

h. How diverse are the skills measured by the entire test?

If the test had more than one separately reported, interpretable subtest, it was credited with 1 point for diversity. If not, it was credited with 0 points. *Although this judgment was made for the test as a whole, it is reported subtest by subtest. Thus, the 1 point for diversity reported under the "Oral Paragraph Reading" subtest of the Individualized Reading Placement Inventory refers to the entire Inventory--not to the "Oral Paragraph Reading" subtest per se.*

i. How clear and simple is the process of converting the raw score to the interpreted score?

If the score conversion procedure was simple, involving one easy-to-understand step--such as a clear chart or table--or if no conversion was necessary because the raw scores were interpretable, the test was credited with 2 points. If the score conversion was complicated by lack of clear

or simple tables or graphs, or if it required two or more steps to get from the raw to the converted scores, (e.g., using one table to get into another table), the test was credited with 1 point. If the score conversion was necessary but complicated and lacked tables or graphs, required many or complicated steps (e.g., computing scores), or was not explicitly provided, the test was credited with 0 points.

j. How interpretable are the scores?

This evaluation procedure looked for scores that were common and simple and could not readily be misunderstood or misused by program personnel. If the scores were pass/fail (or some other binary judgment), grade equivalents, percentiles, or meaningful raw scores (e.g., a words-per-minute reading rate or a precise report of letters for which the examinee could not give the sound), the test was credited with 1 point. If the scores were any other less common, novel, or ambiguous conversion, or conversion was lacking for raw scores not meaningful in themselves, the test was credited with 0 points.

k. Are there scales of performance on real-life skills?

If the test included such scales (e.g., map reading, following directions, reading classified ads), it was credited with 1 point; if judgments on such skills were not included, the test was credited with 0 points.

l. Is the validating group representative of the national population of adults for whom the test was designed?

Five considerations were included in the evaluation of the representativeness of the groups used to norm the test: (1) Was the sample obtained through cluster, stratified, or random rather than incidental sampling?

(2) Was the validating done less than five years ago? (3) Was there geographic representation? (4) Was the validating group composed of adults at the appropriate educational level (e.g., adult education, students or people of similar characteristics)? (5) Were various population density characteristics (e.g., urban, suburban, rural, etc.) represented? If the answers to four or five of these questions, based upon convincing tabulation for the third, fourth, and fifth ones, was "yes", the test was credited with 1 point. If there were fewer than four "yes" answers, the test was credited with 0 points.

m. Is racial, ethnic, and sex representation reported in the validation?

If representation on more than one of these characteristics was reported, the test was credited with 2 points; if representation on only one characterization was reported, the test was credited with 1 point; if no representation was reported, the test was credited with 0 points.

n. Are alternate forms available?

If alternate forms, developed according to the same specifications to measure the same attributes were available, the test was credited with 1 point; if not, the test was credited with 0 points.

o. Are alternate forms comparable?

Alternate forms of instruments can be comparable in many ways; there are considerations of content, approach or method, validities, similarity of descriptive statistics, and reliabilities. If available information indicated that the alternate forms were similar on these criteria, then the test was credited with 1 point. If a test's alternate forms exhibited low or no comparability, or the test had no alternate forms, it was credited with 0 points.

p. Can decisions be made?

This final aspect of administrative usability focused upon whether the test provided information useful in making decisions concerning individual examinees. If the test manual established definite relationships between scores and specific decisions through the use of graphs, charts, cut-off scores, or other means which encouraged fairly specific decisions (e.g., "a score below this point means the examinee needs remediation to strengthen his word attack skills"), the test was credited with 2 points. If the test indicated interpretations of scores that could lead to specific decisions, or merely presented interpretations or definitions rather than decisions (e.g., "a high score indicates the need for testing with a standardized reading test for more accurate information"), the test was credited with 1 point. If the test provided vague or poor guidelines, leading to highly intuitive, subjective judgments, or presented no information useful in making decisions, it was credited with 0 points.

The Administrative Usability ratings were summed for a total rating, varying from 0 to 19 points. These ratings were translated into letter grades of G (good, 16 to 19 points), F (fair, 12 to 15 points), and P (poor, 0 to 11 points).

TEST REVIEWS

Descriptive information on the individual tests is included in the following test reviews. The reviews are organized into three general categories similar to those suggested by Otto:²⁵ criterion-referenced functional literacy tests, standardized tests, and informal tests. Criterion-referenced functional literacy tests measure an examinee's performance on real-life skills (e.g., reading maps, reading bills and applications) against a predetermined standard of acceptable performance. Such tests intend to provide information which is very task-oriented and immediately relevant to the examinee's everyday activities.

Standardized tests measure an examinee's performance relative to the performance of others who have taken the test. Although these tests may use functional literacy tasks for content, they typically measure such traditional reading behaviors as vocabulary, comprehension of a reading passage, or spelling.

Informal tests may be designed to provide information about an examinee's general reading level, or about more specific reading abilities, such as letter or word recognition. They are often individually administered and seek to convey to the examinee a feeling of informality meant to reduce anxiety in the testing situation. Usually the directions for administering, scoring, and interpreting such tests are very short and suggestive, if present at all.

Within each of the three categories, the test reviews are arranged alphabetically by test name. All entries follow a standard format, as outlined below.

Test Name

Publisher: The name and address of the firm or individual making the test available are given here.

Description: This section indicates what the test is intended to measure, describes any subscales included in the test, and notes available alternate levels.

Availability of Alternate Forms: This section describes what alternate (parallel) forms are available to test users.

Administration Time: This section indicates the time necessary for the examinee to take the test, including the time spent in receiving initial instructions.

Administration Procedures: This section indicates whether the test is administered to individuals or to groups. It further details the activities of the examiner and the examinees during the test administration.

Materials Used: Materials needed by the examiner and also those needed by the examinee are listed here.

Scoring Procedures: Procedures for scoring the test are described in this section.

Interpretation Procedures: This section notes what interpretable scores the test provides, and specifies the types of conclusions that can be drawn or decisions that can be made on the basis of test results.

Validity: This section presents the evidence for validity offered by the test developer.

Reliability: This section presents the evidence for reliability provided by the test developer.

Field Tryouts:

This section describes the nature of field tryouts conducted with the test. The characteristics of the tryout population are included if the test developers reported them.

Ratings:

This section specifies the pages on which evaluations relating to the test may be found.

CRITERION-REFERENCED FUNCTIONAL LITERACY TESTS

Adult Performance Level Functional Literacy Test (APL)

- Publisher: Dr. Norvell Northcutt
Division of Extension
103 Extension Building
University of Texas at Austin
Austin, Texas 78712
- Description: The APL is a test of functional literacy for adults. There are 42 items, many of which involve more than one question. The items test an examinee's knowledge of consumer economics, law and health; his ability to perform real-life tasks; as well as his reading and writing ability.
- Availability of Alternate Forms: There are no alternate forms available.
- Administration Time: The test takes approximately 60 minutes to administer.
- Administration Procedures: The test is individually administered in an interview format. The examiner reads the questions aloud while the examinee follows along in his booklet. The examinee then responds, either by reading orally or calling out the correct answer from several choices. The examiner records the answer given and goes on. If the examinee is asked to do a task requiring writing (filling out a check, addressing a letter), the examiner gives the examinee the questionnaire in which to write his response. Thus all answers are recorded in the questionnaire.

Materials Used: Examiner: Questionnaire, pencil.
Examinee. Booklet, pencil, eraser.

Scoring Procedures:

The test is scored in two ways. Multiple choice items are scored by comparing the examinee's answer to the correct answer indicated in the questionnaire. Questions in which the examinee engages in a written task are scored according to a system of rules given in the handbook, indicating what answers are acceptable and what are not.

Interpretation Procedures:

For purposes of initial analysis, scores are grouped into quartiles according to the number of points achieved on the test. They are interpreted primarily, however, according to three APL levels; APL 1 (least competent), APL 2 (marginally competent) and APL 3 (most competent).

Validity:

Validity consists of research showing the relationship of items, groups of items, and levels of competence to various criteria such as income, education level, and job status. There was also a technical review conducted by experts, and several cycles of field testing and redesigning. These data are too extensive to summarize here.

Reliability:

Item difficulty levels comparing earlier surveys and the final survey are provided as a measure of reliability.

Field Tryouts:

The field tryouts were conducted on a random sampling of geographically stratified counties. Three-hundred sixty counties were chosen and divided into 6 blocks. Each block was an independent subsample representing the continental

U. S. A starting point within each county was randomly chosen, and interviewers visited individual residences and administered the test. The weighted sample compared very closely with the universe in sex, age, education, urban distribution, geographical distribution, family income, and race.

Ratings:

See pages 100-101.

Basic Reading Skills Mastery Test

- Publisher: Services for Educational Evaluation, Inc.
P.O. Box 261
Bloomington, Indiana 47401
- Description: This test is an objective measure of comprehension in functional reading. The test consists of four scored subscales: Following Directions, Locating References, Gaining Information, and Understanding Forms. There is also a non-scored subscale designed to indicate the examinee's attitudes and habits in reading for personal development. Three levels of the test are available: Level A for 12 year olds, Level B for 15 year olds, and Level C for 18 year olds. Level C is used for adults.
- Availability of Alternate Forms: There are no alternate forms available.
- Administration Time: Two 50-minute administrations are required for the test. All students are to be given time to finish the test.
- Administration Procedures: The test is group administered. The examiner provides testing materials and reads instructions to the students. The examinee reads passages or forms and answers comprehension questions on an answer sheet.
- Materials Used: Examiner: Examiner's manual, test booklet.
Examinee: Test booklet, pencil, eraser, answer sheet.
- Scoring Procedures: The answer sheets are computer scored, and the results returned on a print-out sheet.

Interpretation
Procedures:

Eighty percent correct or better is considered mastery on this test.

Validity:

Content validity was based on the conclusions of a committee of reading specialists regarding functional reading skills. There were also student reviews of the items, experts' reviews, and field tryouts.

Reliability:

The K-R 20 yielded an estimate of internal consistency of .98 for the total test. For the four subscales the K-R 20 values were .87, .91, .93, and .93.

Field Tryouts:

A sample of 2700 Maryland students, including minority groups, representing urban, suburban, and rural areas throughout the state, was used to test the three levels of this test.

Ratings:

See pages 100-101.

Reading/Everyday Activities in Life (R/EAL)

Publisher: CAL Press, Inc.
76 Madison Avenue
New York, New Yo 10016

Description: The test is an objective assessment of functional literacy presented in nine selected categories of common printed materials encountered in daily living. English and Spanish versions are available.

Availability of Alternate Forms: There are no alternate forms available.

Administration Time: The test requires approximately 20-30 minutes; an examinee works at his own pace.

Administration Procedures: The test may be individually or group administered. The examiner provides testing materials (i.e., test answer booklet and cassette tape recorder with R/EAL cassette). The examinee listens to taped questions which correspond to material in the test booklet and records answers in the test booklet.

Materials Used: Examiner: Examiner's manual.
Examinee: Test booklet, cassette recorder with R/EAL cassette tape, pencil, eraser.

Scoring Procedures: Scoring is done by hand, referring to pre-established correct responses. Raw scores are totaled for the nine categories and the total raw score is then converted to percentage of items passed.

Interpretation Procedures:

Criterion-referenced - Test items are directly related to sets of objectives associated with each of the nine reading activities. Functional literacy is defined as passing 80% or more of the test items (or achieving a raw score of 36).

Interpretation of Individual Subtests - Following a review of the examinee's performance on individual subtests, the interpreter can recommend prescriptive programs to meet areas of need indicated through detailed task analyses outlined for each subtest.

Validity:

Criterion-related validity was investigated by computing the correlation between this test and the Stanford Achievement Test; the correlation between the two tests was .74 (n=434). Content validity relies on the selection of questions from the task analyses which specified test objectives.

Reliability:

The internal (inter-item) consistency estimate of reliability, based on K-R 20, was $r=.93$; the target group for the reliability sample included a specified sex distribution, and a majority of minority individuals who had completed an average of nine years of school and who had a reading grade equivalent of 5.2 on the Stanford Achievement Test. No breakdown was provided, however, for sex, or individual minority representation.

Field Tryouts:

The testing manual indicates the subjects used to standardize the test included 169 males and 265 females,

aged 16-21. The subjects were all low income individuals, and a majority of them were Blacks, Spanish-surnamed or rural whites. Subjects had completed an average of nine years of school and had an average reading equivalent of 5.2 on the Stanford Achievement Test.

Ratings:

See pages 100-101.

Wisconsin Test of Adult Basic Education (WITABE)

Publisher: Rural Family Development Program
University Extension
University of Wisconsin
Madison, Wisconsin 53706

Description: This test was especially designed to monitor the basic skills achievement of persons enrolled in the Wisconsin Rural Family Development Program. The test appears appropriate for general use with adults who read below high school level.

Availability of Alternate Forms: There are no alternate forms available.

Administration Time: The test is generally untimed; however, the maximum administration time for the two reading sections combined should be less than one hour.

Administration Procedures: The testing conditions are very flexible. The examinee works at his or her own pace; the examiner's only responsibility is to ensure that the written instructions are understood. The test may be administered individually or to groups. The WITABE consists of verbal and coping skills sections, both of which might loosely be considered "reading" tests. The skills required to complete the coping skills subtest include using a road map, ordering by mail, filling out a tax return, using a phone book, and a variety of comparable tasks. A numerical subtest is also part of the WITABE. Any of the sections may be given separately.

Materials Used: Examiner: Test booklet.
Examinee: Test booklet, pencils, eraser.

Scoring
Procedures:

Scoring is done by hand; responses are compared with pre-established correct answers. A few questions in the coping skills subtest have more than one point scoring but assignment of points is still objective and relatively simple. The raw score obtained is not converted.

Interpretation
Procedures:

The WITABE was developed to measure differences between treatment groups and control groups in the Wisconsin program. Raw scores were adequate for this purpose and thus no score interpretation process exists. Test scores cannot at this time be converted into grade equivalents, percentiles or other norm-comparisons; nor is any criterion-referenced diagnostic information given.

Validity:

Without giving numerical information, the authors state that the test data item analysis conducted by the University's Psychometric Laboratory, which involved field test results from 120 rural Wisconsin 6th, 7th and 8th graders, led to rejection of unsuitable items. The modified instrument was administered to 37 adults to determine the psychometric quality of the items.

Reliability:

The authors report that the Hoyt reliability index for the 20-item verbal subtest was .90. The reliability for the 29 item coping skills subtest was also reported as .90.

Field Tryouts:

The WITABE has been used by the Wisconsin Rural Family Development Program with the 120 public school students and 37 adults mentioned above, and with treatment and control

groups chosen for Rural Family Development Program evaluation. The makeup of the latter two groups was specified by age, sex, and geographic location; however, no scoring or norming data was provided.

Ratings:

See pages 100-101.

STANDARDIZED TESTS

Adult Basic Learning Examination (ABLE), Level I

- Publisher: Harcourt Brace Jovanovich, Inc.
757 Third Avenue
New York, New York 10017
- Description: The test is designed to determine the general educational level of adults. It consists of three levels: Level I (Grades 1-4), Level II (Grades 5-8), and Level III (Grades 9-12). Each level includes vocabulary, reading, spelling and arithmetic tests. (The arithmetic test was not reviewed.)
- Availability of Alternate Forms: Alternate forms A and B are available.
- Administration Time: Estimated times for administration of the subtests are: vocabulary, 20 minutes; reading, 30 minutes; spelling, 15 minutes.
- Administration Procedures: The ABLE handbook recommends group administration. However, this test could be individually administered as well. The vocabulary and spelling tests are dictated to the examinee, who indicates his answers by shading in an oval in his test booklet under his word choice. The vocabulary section requires sentence completion; three word choices are given. Examinees complete the reading section independently, choosing the correct word to complete a thought.
- Materials Used: Examiner: Test handbook, scoring key and group scoring record.
Examinee: Test booklet and pencil.

Scoring
Procedures:

A key is provided in the packet for hand scoring, but scoring can be done by machine.

Interpretation
Procedures:

The number of items right for each test can be interpreted in terms of grade level equivalent. Grade level equivalents are the only conversion provided. The test developers also suggest that users develop local norms.

Validity:

Concurrent validity studies are reported, based on test administration of the ABLE and the Stanford Achievement Test (SAT) to a school group within a week's time. Correlations among appropriate scales ranged from .60 to .76. In addition, correlations were computed between the SAT Paragraph Meaning Scale and ABLE for a Job Corps group. These correlations ranged from .36 to .72.

Reliability:

Split-half (odd-even) reliability coefficients adjusted by the Spearman-Brown formula are reported for grade 3 of the school group (.87 for vocabulary, .93 for reading, .95 for spelling), grade 4 of the school group (.89 for vocabulary, .93 for reading, .95 for spelling), the Job Corps group (.85 for vocabulary, .96 for reading, .96 for spelling), and a group of adult basic education students (.91 for vocabulary, .98 for reading, .94 for spelling).

Field Tryouts:

ABLE was administered to three groups: 1) elementary and junior high school students, 2) Job Corps members, and 3) Hartford-New Haven adult students. The school group consisted of 1,000 pupils per grade (grades 2-7) from four school systems in four states. The Job Corps group consisted

of approximately 800 young men in both urban and conservation centers. The Hartford-New Haven group consisted of approximately 450 adults enrolled in basic education classes in those two cities. Statistics on ethnic composition and educational level are displayed in the test handbook.

Ratings:

Reading, see pages 102-103.

Spelling, see pages 102-103.

Vocabulary, see pages 102-103.

Basic Occupational Literacy Test (BOLT), Fundamental Level

Publisher: U. S. Department of Labor

Description: The test is designed to measure the basic reading and arithmetic skills of educationally disadvantaged adults. There are four subtests: reading vocabulary, reading comprehension, arithmetic computation, and arithmetic reasoning. Each test is available at four difficulty levels.

Availability of Alternate Forms: Three alternate forms are available for the first three levels. The advanced level offers two forms for each subtest.

Administration Time: Fifteen minutes for each subtest.

Administration Procedures: Before administering the subtests, each examinee is given the Wide Range Scale (included with the test) to determine the appropriate level of BOLT to administer. Directions are given orally to individuals or small groups. Each examinee records his answers on an answer sheet by marking the appropriate circle.

Materials Used: Examiner: Manual, scoring key, stop watch, test record cards.
Examinee: Test booklet, answer sheet, pencil, paper clips, scratch paper.

Scoring
Procedures:

Scoring can be done either by hand or by machine. Hand scoring is done by placing a stencil over the answer sheet and counting the number of visible marks. The total number of correct responses can then be converted to a standard score or General Evaluational Development (GED) level using conversion tables contained in the User's Manual.

Interpretation
Procedures:

Once scores are converted to GED levels they can be compared to the GED levels for occupations listed in the Dictionary of Occupational Titles. One must be familiar with GED scores as well as standard scores in order to interpret scores for the BOLT.

Validity:

To establish content validity, directions, test items, and time limits were given a preliminary tryout. Following revision, extensive field testing was conducted and an intricate set of item analysis rules led to development of the final forms. Construct validity research was conducted to answer general questions about testing disadvantaged adults.

Reliability:

Internal consistency of the subtests was judged according to K-R 20, and computed for each subtest as it was administered to each subgroup. K-R 20 coefficients for the final forms (fundamental level) were: vocabulary .79 (form A) and .80 (B), and comprehension .77 (A) and .76 (B).

Field Tryouts:

A preliminary tryout was conducted on 453 persons. The sample, from 10 states, was stratified by geographic area,

sex, age, education, and minority group status. A similarly stratified sample of more than 8,000 subjects from 33 states took part in the major field testing of 11 experimental forms of the reading tests. Some 1600 of these subjects were given various forms of what became the fundamental level tests. Extensive breakdowns of subjects by geographic area and by minority group status were presented.

Ratings: Comprehension, see pages 102-103.
Vocabulary, see pages 102-103.

General Educational Performance Index (GEPI)

- Publisher: Steck-Vaughn Company
807 Brazos
P.O. Box 208
Austin, Texas 78767
- Description: This test of high school equivalency was designed to predict success on the General Educational Development Test. Although it is divided into five subscales, this evaluation is concerned only with tests 1 (Correctness and Effectiveness of Expression) and 2 (Literature Interpretation).
- Availability of Alternate Forms: Alternate forms A and B are available.
- Administration Time: Although the test is untimed, it is estimated that tests 1 and 2 each require from 20 to 40 minutes.
- Administration Procedures: The test is group administered. The examiner distributes test booklets and answer sheets; the examinee reads instructions for each subscale, then does the items and records his answers.
- Materials Used: Examiner: Copy of test booklet and the Manual of Directions.
Examinee: Copy of test booklet, a score sheet, a pencil, an eraser.
- Scoring Procedure: Scoring the test is a simple, objective process. The examiner uses a template to mark the examinee's answer sheet, then counts the number of correct answers per subscale.

This raw score is then converted to a standard score using a table in the manual.

Interpretation
Procedures:

The standard scores in the GEPI subscales should give the examinee and examiner an idea of the examinee's readiness for the GED. There are tables comparing the standard GEPI scores and GED scores. The GED is a pass/fail test, with a specified cut-off standard score (40 in some states). The GEPI subscales can also be used to isolate areas of weakness, to determine groupings for instruction, or, (in a re-test situation) to see if the examinee has made progress.

Validity:

Content validity consists of the authors' statements that the test was prepared and reviewed by experts who had thoroughly researched the field. Also, the literacy related scales of the GEPI correlate with the appropriate GED scales in the range of .62 to .70.

Reliability:

The alternate form reliability coefficients were .73 for Correctness and Effectiveness of Expression and .68 for Literature Interpretation.

Field Tryouts:

Field tryouts were conducted in 1974 and 181 adult students randomly chosen from all parts of Texas, and enrolled in a variety of GED programs. Data were provided on sex, racial and age composition of the group.

Ratings:

Correctness and Effectiveness of Expression, see pages
100-101.

Literary Interpretation, see pages 100-101.

SRA Reading Index

Publisher: Science Research Associates, Inc.
259 Erie Street
Chicago, Illinois 60611

Description: This test has five parts. Part (Level) I, Picture-Word Association, tests the student's ability to associate a word with a picture of an object. Part (Level) II, Word Decoding, tests a student's ability to choose the right word to complete a sentence. Part (Level) III, Phrase Comprehension, requires the student to choose the appropriate word or phrase to complete a sentence. Part (Level) IV, Sentence Comprehension, requires the student to choose an accurate paraphrase of a given sentence. Part (Level) V, Paragraph Comprehension, has the student read a paragraph and answer comprehension questions.

Availability of Alternate Forms: There are no alternate forms available.

Administration Time: The test can be given in one 25-minute timed session, but timing is not required.

Administration Procedures: The test is group administered. The examiner first reads instructions orally to the group; examinees then read the questions and mark the appropriate answers.

Materials Used: Examiner: Examiner's manual, test booklet.
Examinee: Test booklet, two lead pencils.

Scoring Procedures: The test booklets are self-scoring; a student's marks are transferred through carbon paper to a key. The examiner counts the correct responses (those within boxes on the key) and records them for each part (level). He then records this number on the cover of the booklet.

Interpretation Procedures: Two sets of norms are given for the test: special education norms and industrial norms. Also, the test booklet includes a chart indicating the number of correct items needed to pass a given level. This number is based on an 80 percent proficiency criterion. The Examiner's Manual discusses use of these scores in relation to job analysis and minimum proficiency needs for certain jobs.

Validity: Content validity relies on the method of choosing items for the test. A pool of items was developed for the test, and then screened by the language department of a Job Corps center for appropriateness and ambiguity. A concurrent validity study was conducted in which the Reading Index scores were correlated with overall job ranking of people in twenty-one occupations. The largest coefficient was .32, and ten coefficients were significant at the .05 level. Also, correlations between the Reading Index and the Flanagan

Industrial test for each of the occupation groups are shown in the manual.

Reliability:

The K-R 20 reliability coefficient was .87. The Raju-Guttman Homogeneity Index was .93. The group tested consisted of 87 men and women from a combination on-the-job training and basic education program in Chicago.

Field Tryouts:

This test was pre-tested on a total of 675 males and females enrolled in special- and adult-education programs in Colorado and South Carolina. It was given to a group of 87 men and women in a combination on-the-job training and basic education program in Chicago in order to establish special education norms. Also, the test was given to 3274 workers to establish industrial norms for whites and nonwhites.

Ratings:

See pages 100-101.

Tests of Adult Basic Education (TABE), Level E

Publisher: CTB/McGraw-Hill
Del Monte Research Park
Monterey, California 93940

Description: The test provides a system for measuring the reading achievement level of adults, based upon a corresponding level of the California Achievement Tests. Level E (Easy) is intended for adults with severe educational limitations or for those from culturally disadvantaged backgrounds. It is intended for the "upper primary" levels, or Grade 2 to beginning Grade 4 level. Level M (Medium) is adapted from the elementary level of the CAT, and Level D (Difficult) is adapted from the junior high school level of the CAT.

Availability of Alternate Forms: Alternate forms 1 and 2 are available.

Administration Time: The Reading Vocabulary section takes 9 minutes; Reading Comprehension takes 31 minutes (total time - 40 minutes). It is permissible to provide a break or rest period after any of the test sections.

Administration Procedures: The test is group administered. The examiner reads general test directions to students before each section, and then reads section directions, which are also printed in the test booklet. Examinees record their answers in test booklets. Each test section has an

established time limit. A set of practice exercises and a locator test are available; these are designed specifically for pre-testing. The practice exercises familiarize examinees with the mechanics of the test, while the locator test, a short vocabulary test, provides a basis for determining the level of TABE best suited for a particular individual.

Materials Used:

Examiner: Examiner's manual, blackboard, stopwatch.
Examinee: TABE test booklet, pencil, eraser.

Scoring
Procedures:

Scoring consists of matching a hand-scoring key to corresponding pages in the examinee's test booklet. The score for each test is the number right; this is recorded on the bottom right hand corner of the last page of each section in the test booklet. The total right for each section is transferred to an appropriate box on the Profile Sheet. Total Section raw scores are added together to obtain the total raw score for the test. Total raw scores are then plotted according to a grade placement level on the Profile Sheet.

Interpretation
Procedures:

In addition to the grade-placement level, the Profile Sheet provides an "Analysis of Learning Difficulties." The analysis is completed by recording a student's errors in each section; the items in each section are listed according to skill areas. The resulting learning

profile becomes a basis for planning remedial or developmental work and individualized instruction needed by the student.

Validity: Claims for content validity are based upon item selection procedures for the California Achievement Test, from which the test has been adapted.

Reliability: No information is available on reliability.

Field Tryouts: No information is available on field tryouts.

Ratings: Comprehension, see pages 102-103.
Vocabulary, see pages 102-103.

INFORMAL TESTS

Adult Basic Reading Inventory

Publisher:

Scholastic Testing Service
480 Meyer
Bensenville, Illinois 60106

Description:

This test has five parts. Part I tests the student's ability to associate a word with a picture. Part II tests the student's sound and letter discrimination. Part III tests the student's ability to associate synonyms (or related words) as he reads the words. Part IV is similar to Part III, except that the student hears the words read orally. Part V requires the student to read paragraphs and answer comprehension questions.

Availability of
Alternate Forms:

There are no alternate forms available.

Administration
Time:

The test can be administered in one session; Parts I and II each require 5 minutes. Parts III and IV each require 10 minutes. Part V requires 15 minutes.

Administration
Procedures:

The test is group administered. In Part I, the examiner reads instructions and examinees underline words associated with pictures. In Part II the examiner reads words to the examinees, and examinees underline words beginning with the same sound as the word read by the examiner. In Part III, the examiner reads

instructions, and examinees underline the word in a list which has about the same meaning as a word written to the side. In Part IV, the examinee performs the same task; however, the words are read orally by the examiner. In Part V, the examiner reads the instructions, and examinees read paragraphs and choose the correct answer to comprehension questions.

Materials Used:

Examiner: Manual of directions.

Examinee: Test booklet, line marker, two colored pencils, eraser.

Scoring Procedures:

Scoring is objective and simple. The examiner simply compares the student's answers in the test booklet to a scoring key. For each part he indicates the number of correct answers. Each raw score is then converted to a percentage score according to instructions provided in the Manual.

Interpretation Procedures:

The Manual indicates how to assess an examinee's reading ability by approximate grade levels, or in terms of functional or absolute illiteracy as defined in the Manual. It also offers some general suggestions on assessing areas of weakness and aspects of remediation.

Validity:

Concurrent validity studies have been done with the Gates Advanced Primary Reading Test and with teacher ratings of student abilities from pre-primer to fifth

grade using a 9-point scale. Correlations with the Gates test ranged from .82 to .88 and with the teacher estimates from .67 to .76.

Reliability:

Reliability studies were conducted on 38 adults in an adult literary project in an urban area of Northern Illinois. The K-R 21 coefficient was .98.

Field Tryouts:

Small scale tryouts of the test were conducted with 38 adults and 17 juvenile male retarded readers. The adults were involved in an adult literacy project in an urban area of Northern Illinois. No sex, ethnic, or racial breakdown was given.

Ratings:

Part I, Sight Words, see pages 104-105.

Part II, Sound and Letter Discrimination, see pages 104-105.

Part III, Word Meaning (Reading), see pages 106-107.

Part IV, Word Meaning (Listening), see pages 106-107.

Part V, Context Reading, see pages 104-105.

Czyzk Pre-Reading Inventory

Publisher:

Janet L. Czyzk (Author).
Adult Reading Specialist
Baltimore County Board of Education
6901 N. Charles Street
Towson, Maryland 21204

Description:

The Inventory consists of various activities designed to help a teacher recognize deficiencies within discriminatory and perceptual skills in the visual, auditory, and perceptual motor areas that must be dealt with before an adult non-reader can begin learning to read.

Availability of
Alternate Forms:

There are no alternate forms available.

Administration
Time:

There are nine separate short sections to the test. Examinees may be given any number in a single session. The test are untimed; no estimate is given of the testing time required.

Administration
Procedures:

The Inventory may be individually or group administered. Each examinee receives a test booklet in which to underline the correct answers. Instructions are given orally by the examiner. Examinees do some of the activities independently, and in the remaining activities respond to lists of words read by the examiner.

Materials Used:

Examiner. Test directions.
Examinee: Test booklet, pencil.

Scoring
Procedures:

The test is hand-scored by the examiner who determines the adequacy of each response. In its present form it serves only to provide diagnostic information to the teacher who seeks, through personal evaluation of test results, to identify students' deficiencies.

Interpretation
Procedures:

The test activities measure examinee abilities in motor skills, reading functional words, perception of letter forms, order and sequence of letters and digits, handwriting speed, auditory discrimination, word perception and word discrimination. Poor examinee performance on any of the sections suggests that the teacher should conduct additional testing on an individual basis.

Validity:

No information is available on validity.

Reliability:

No information is available on reliability

Field Tryouts:

No information is available on field tryouts.

Ratings:

See pages 106-107.

Harris Graded Word List and
the Informal Textbook Test

Publisher: Adult Continuing Education Resource Center
Montclair State College
Upper Montclair, New Jersey 07043

Description: These two tests are used together. The Harris Graded Word List consists of seven lists of words representative of varying reading levels. The Informal Textbook Test, given to applicants who score above grade level 2.0, involves a series of seven passages (at reading levels 2-8), each followed by a list of comprehension questions.

Availability of Alternate Forms: There are no alternate forms available.

Administration Time: The Harris Graded Word List requires only one minute for each examinee. The administration time for the Informal Textbook Test (group administered) is not known.

Administration Procedures: The Harris Test is individually administered. The examiner has the examinee read each list of words, noting mentally the level at which he makes 3 or 4 errors. This level is later entered on the registration form. Examinees who score above 2.0 reading level take the group administered Informal Textbook Test. The examinee reads seven passages and answers the comprehension questions in the booklet.

Materials Used: Examiner: Harris Graded Word List, pencil.
Examinee: Informal Textbook Test booklet, pencil,
eraser.

Scoring Procedures: Harris Graded Word List: The examiner mentally notes at which level the examinee makes 3 or 4 errors in reading words. Informal Textbook Test: The examiner compares the examinee's responses with pre-established correct responses.

Interpretation Procedures: Harris Grade Word List: If the examinee does not read above 2.0 reading level, he is classified as a beginning reader. Informal Textbook Test: The examinee's instructional level is determined by noting at which reading level he scores 2-3 (out of a possible 4). Any score below 2 indicates he should be in a beginning group.

Validity: Validity consists of the author's statement that the Harris Graded Word List is "scientifically organized," and that the standards for reading levels are based upon the Dale-Chall Formula and ratings given in a combined word list by Buckingham and Dolch.

Reliability: No information is available on reliability.

Field Tryouts: No information is available on field tryouts.

Ratings:

Harris, see pages 106-107.

Informal, see pages 104-105.



Idaho State Penitentiary Informal Reading
Inventory

Publisher: The Reading Education Center
Boise State University
Boise, Idaho 83720

Description: The Inventory is designed to provide a reading teacher with a student's estimated independent reading level, estimated instructional level, estimated frustration level, estimated listening level, specific word recognition deficiencies, and specific comprehension deficiencies. The test is applicable specifically to penal adult populations, and particularly to those persons who have difficulty learning to read.

Availability of Alternate Forms: Alternate forms A and B are available. Each is divided into two major sections, Word Lists and Stories. The two forms are bound in one booklet to facilitate repeated administration.

Administration Time: The word lists require approximately 10 minutes. Each of the eight stories (corresponding to grade levels in difficulty) takes 5-10 minutes to read aloud. The estimated time for administration of comprehension tests following each story is five minutes per story. All of the stories need not be administered at one sitting.

Administration
Procedures:

The test is individually administered by a reading teacher. The examinee reads words selected from each of the stories aloud while the examiner codes errors on a copy of the word lists, beginning with the first grade level story. The examinee continues pronouncing words until three words within one list have been missed. For the oral stories, the examinee reads each story aloud while the examiner codes errors. The coding procedures suggested are somewhat complex and not standardized. After the examinee has finished the oral reading, the examiner asks comprehension questions on each of the stories, recording correct and incorrect responses.

Materials Used:

Examiner: Pencil, teacher's copy of Student Word List and Student Stories, recapitulation sheet, manual of directions.

Examinee: Student's copy of Word List and Stories.

Scoring
Procedures:

Scoring consists of a complex and highly detailed system of coding to note student errors in oral reading. Scoring of comprehension questions is done using a guide for acceptable answers. Percentage scores are used to determine achievement level (roughly corresponding to grade levels 1-8) on the word list portion of the test. On the oral reading portion of the test, word recognition and comprehension errors are recorded following each story. The examiner

then transfers the errors in each story (grade level) into the terms Independent, Instructional, Frustration and Listening to indicate the examinee's ability, in each category, in correspondence to a grade level. All scores are recorded on the Recapitulation Sheet, which provided an estimated picture of the examinee's composite reading ability.

Interpretation Procedures:

Information recorded on the Recapitulation Sheet is intended to establish the examinee's estimated Independent, Instructional, Frustration and Listening levels in a manner roughly corresponding to grade levels. It also shows specific strengths and weaknesses in word recognition and comprehension as well as in pronunciation. The interpretation procedures are subjective, with judgments and estimates left to the examiner's discretion.

Validity:

Content validity consists of the authors' claim that the stories are designed to appeal to the penitentiary inmate-student. All stories were subjected to readability formulas (Botel, Dale-Chall and Flesche) to coincide with other graded materials.

Reliability:

No information is available on reliability.

Field Tryouts:

No information is available on field tryouts.

Ratings:

See pages 104-105.

An Informal Reading Inventory for Use by
Teachers of Adult Basic Education

Publisher:

Office of Adult Basic Education
State Department of Education
Concord, New Hampshire 03301

Description:

This test measures reading performance from level 1 through level 6. These levels correspond with the levels in graded readers. The inventory has four parts: Part I, Word Recognition (testing word attack skills and vocabulary level); Part II, Oral Reading and Comprehension questions; Part III, Listening Ability (present potential level); and Part IV, Visual and Auditory Perception and Discrimination (used for examinees who cannot function at the introductory level of Part I).

Availability of
Alternate Forms:

There are no alternate forms available. /

Administration
Time:

The time required for the test is not specifically indicated, though administration probably requires from 20 to 30 minutes, depending on how soon a student reaches his frustration level.

Administration
Procedures:

The test is individually administered. In Part I the examiner exposes words for one second for the examinee's flash recognition. If the examinee misses the word, he is allowed to analyze it. In

Part II the examinee reads paragraphs orally and answers comprehension questions. In Part III the examiner reads paragraphs orally to the examinee and the examinee responds to comprehension questions. Part IV is administered to examinees who cannot function at the introductory level of word recognition. The examinee names letters pointed out to him, gives the sounds of blends, and writes the initial, final, or middle sounds of words read to him.

Materials Used: Examiner: Informal Reading Inventory Booklet, pencil, two 3x5 cards.

Examinee: Paper, pencil, eraser.

Scoring Procedures:

The scoring of this test is objective, but fairly complicated. The examiner must record each error the student makes, using a system of notations. The number of words correctly recognized in Part I is totalled. In Part II, the examiner computes the number of reading errors and percentage of comprehension questions answered correctly. In Part III, the examiner computes the number of comprehension questions answered correctly. In Part IV, the examiner records the examinee's oral errors to letter recognition and blending tasks and hand scores the written responses to the auditory discrimination tasks.

Interpretation
Procedures:

Based on the scores, the examiner computes the examinee's independent level, instructional level and frustration level. These levels correspond closely with comparable levels in a graded reader.

Validity:

No information is available on validity.

Reliability:

No information is available on reliability.

Field Tryouts:

No information is available on field tryouts.

Ratings:

Part I, Word Recognition, see pages 106-107.

Part II, Oral Reading, see pages 104-105.

Part III, Present Potential Level, see pages 104-105.

Part IV, Visual and Auditory Perception and

Discrimination, see pages 106-107.

Individual Reading Placement Inventory

Publisher:

Follett Educational Corporation
1010 West Washington Blvd.
Chicago, Illinois 60607

Description:

This test is divided into five parts. Part I, Word Recognition and Analysis tests a student's knowledge of sight words and his ability to decode words he cannot immediately recognize. Part II, Oral Paragraph Reading tests the student's oral reading skills and comprehension. Part III, Present Language Potential tests the student's comprehension of paragraphs read to him. Part IV tests the student's auditory discrimination. Part V, which is not scored, tests the student's ability to name letters of the alphabet and their sounds. This test is used only if the student scores 1.0 on Part I.

Availability of
Alternate Forms:

Alternate forms A and B are available

Administration
Time:

The test has four parts each of which require approximately 10-20 minutes, depending on how many items a student is able to complete before reaching his frustration level.

Administration
Procedures:

The test is individually administered. In Part I the examiner asks the examinee to read words aloud, either by recognition or word analysis. In Part II the examinee reads paragraphs orally and answers comprehension questions.

In Part III the examinee listens to paragraphs read orally by the examiner and answers comprehension questions. In Part IV the examiner reads lists of words orally and the examinee identifies the word in each list that begins or ends differently or has a different vowel sound in the middle. In Part V (used only if examinee scores 1.0 on Part I) the examiner points to letters of the alphabet, and the examinee names each letter and gives one sound of the letter.

Materials Used:

Examiner: Student's Test and Scoring Booklet, pencil, word recognition wheels, paragraphs on cards.

Examinee: No equipment needed.

Scoring Procedures:

The examiner records the student's errors on each part of test using an objective, but (for Parts I and II) quite complicated system of notations. The errors are then totalled.

Interpretation Procedures:

On the basis of the number of items missed per level, the student's Independent Level, Instructional Level, and Frustration Level are computed. Each level of the test is apparently comparable to a grade level. The Student's Test and Scoring Manual also has places for the examiner to indicate a student's specific reading problems--word analysis, recitation, rate difficulties, etc.

Validity:

Content validity consists of the author's reliance upon the researchers' formulas in determining levels of reading difficulty. A concurrent validity study correlated three tests of silent reading ability to the Individual Reading Placement Inventory. The tests used were the Rasof-Neff ($r = .89$, $N = 146$), the Stanford Achievement ($r = .78$, $N = 75$) and the California Achievement ($r = .87$, $N = 104$).

Reliability:

Reliability coefficients were obtained by using alternate forms in pre- and post-test situations. Coefficients ranged from .91 to .98 for overall performance on the inventory in six different reliability studies.

Field Tryouts:

The User's Manual indicates that the field tryouts incorporated 410 students, including 124 adult basic education students from Florida, 69 junior-senior high school retarded readers from Florida, 111 junior-senior high school retarded readers from Illinois, 86 junior-senior high school retarded readers from Georgia, and 20 adult basic education students from a federal prison in Florida. No sex, ethnic, or racial breakdown was included.

Ratings:

Part I, Word Recognition and Analysis, see pages 106-107.
Part II, Oral Paragraph Reading, see pages 104-105.
Part III, Present Language Potential, see pages 104-105.
Part IV, Auditory Discrimination, see pages 104-105.

Initial Testing Locator Tests

Publisher:

Adult Continuing Education Resource Center
Montclair State College
Upper Montclair, New Jersey 07043

Description:

The reading test includes three passages of varying difficulty, each followed by comprehension questions. It is a preliminary screening test, designed to help the instructor tentatively assign students to different instructional levels or classes within General Educational Development (GED) programs. This test is given in conjunction with the Slosson Oral Reading Test.

Availability of
Alternate Forms:

There are no alternate forms available.

Administration
Time:

Although the time required for the test varies according to an examinee's performance, it would seem the test would require less than 20 minutes.

Administration
Procedures:

The test is individually administered. The examiner asks the examinee to read Passage A orally and answer the comprehension questions orally. If he is unable to do this, the test ends. If he is able to do it easily, he is given Passage B and asked to read and answer questions in the booklet by himself. If he can do this, he is given level C and asked to read it and respond to questions. After

he has reached his highest level--B or C--he is given the CTB/McGraw-Hill Test of Adult Basic Education (TABE), levels M or D for further diagnostic testing.

Materials Used: Examiner: Test Booklet, pencil
Examinee: Test Booklet, pencil, eraser

Scoring Procedures: The examiner compares the examinee's answers with pre-established correct answers.

Interpretation Procedures: If the student cannot read Passage A, he is probably a low-level ABE student. If he can read Passage A and Passage B, but not Passage C, he is probably higher-level ABE or Pre-GED. If he can also read Passage C, he is at least low-level GED. In all but the first situation, use the TABE level M or D for further diagnostic testing.

Validity: No information is available on validity.

Reliability: No information is available on reliability.

Field Tryouts: No information is available on field tryouts.

Ratings: See pages 104-105.

Reading Evaluation - Adult Diagnosis (READ)

Publisher: Follett Publishing Company
1010 West Washington Blvd.
Chicago, Illinois 60607

or

Literacy Volunteers of America, Inc.
222 West Onondaga Street
Syracuse, New York 13203

Description: The test has three parts. Part I, Word Recognition, tests the student's knowledge of sight words. Part II, Word Analysis, tests the student's decoding skills. Part III, Reading Inventory, tests the student's oral reading and comprehension.

Availability of Alternate Forms: Alternate Forms 1 and 2 are under one cover for the Reading Inventory (Part III) of the test.

Administration Time: The three parts of the test do not need to be administered at the same time. Administration times for Parts I and II are estimated at five and ten minutes respectively; estimated administration time for completion of all levels (B-J) of Part III is half an hour.

Administration Procedures: The test is individually administered. In Parts I and II, the examinee reads words and sounds aloud while the examiner records errors for each list. In Part III, the examinee reads stories and answers questions aloud while the examiner records errors for each story.

Materials Used:

Examiner: Testing/Record Booklet, pencil,
Examinee: Reading Lists and passages from test
booklet.

Scoring
Procedures:

Scoring is accomplished through an objective and fairly simple process of recording student scores for each of the test's three parts on a Summary Sheet. Correct scores are converted to percentages for Part I (Word Recognition); in Part II, specific diagnostic information is recorded on a variety of reading subskills, such as knowledge of alphabet and letter sounds. The difficulty of reading and listening comprehension selections in Part III corresponds roughly to grade levels, and passing any selection depends upon not exceeding a specified error count. The total passing score is converted to equivalent grade level. The test is intended for administration on a pre-post basis.

Interpretation
Procedures:

The Test Summary Sheet provides a detailed reading profile for use in planning a specific instructional program for the examinee. The test booklet also provides suggestions for analyzing and using the test scores for individualized prescriptive programs.

Validity:

Content validity relies on the acceptance of the test items by teachers in adult education.

Reliability: No information is available on reliability.

Field Tryouts: No information is available on field tryouts.

Ratings: Part I, Word Recognition, see pages 106-107.

Part II, Word Analysis, see pages 106-107.

Part III, Reading Inventory, see pages 104-105.

TEST EVALUATIONS

The test evaluations on the following pages are divided into the three major categories used in the Test Review section: criterion-referenced functional literacy tests, standardized tests, and informal tests. The standardized tests and informal tests are further categorized under subheadings indicating the specific behaviors being tested (e.g., oral reading, spelling, vocabulary).

The behaviors or skills listed under standardized tests are as follows:

- o General Educational Development Performance Tests. These tests predict examinee performance on the General Educational Development Test.
- o Multiple Reading Skills Tests. These tests yield results of a composite nature (e.g., word meaning and passage comprehension) not readily assigned to a single category.
- o Reading Comprehension Tests. These tests measure the ability to comprehend material read silently.
- o Spelling Tests. These tests measure spelling ability.
- o Vocabulary Tests. These tests measure knowledge of word meanings.

Behaviors or skills listed under informal tests include the following:

- o Oral Reading Tests. These tests measure the ability to read passages aloud and to understand what was read, and are sometimes used to measure the level of listening comprehension as well.
- o Reading Comprehension Tests. These tests measure the ability to comprehend material read silently.
- o Recognition or Discrimination Tests. These tests measure the ability to discriminate between sounds, to pronounce the sounds made by letters and blends, or to recognize sight words.

- o Vocabulary Tests. These tests measure knowledge of word meanings.

The tables containing ratings for each test appear on facing pages.

The left page includes three major headings: Test Name, Measurement Validity, and Examinee Appropriateness. The right page also includes three headings: Technical Excellence, Administrative Usability, and Total Grades. These headings function as follows:

- o Test Name. The test name appears entirely in upper case letters. Subtest names are in upper and lower case.
- o Measurement Validity, Examinee Appropriateness, Technical Excellence, and Administrative Usability. Individual criteria are listed underneath each of these headings; following the criteria are the possible ranges of points assigned on each criterion. The actual entries for each test (or subtest) are listed in the body of the table.
- o Total Grades. Grades of good, fair, or poor are assigned to each test (or subtest) summarizing the ratings in the four major criterion areas (measurement validity, examinee appropriateness, technical excellence, and administrative usability).

A detailed discussion of the criteria used to evaluate the tests is presented in the section Evaluative Criteria.

TEST NAME	MEASUREMENT VALIDITY		EXAMINEE APPROPRIATENESS											
	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Comprehension		Format					Response Mode	Recording Response	Power Test
					Content	Instructions	Layout	Illustration/Print Quality	Auditory Presentation	Coherence				
Rating Range →	0-8	0-4	0-1	0-1	0-2	0-4	0-1	0-1	0-1	0-1		0-2	0-1	
I. CRITERION-REFERENCED FUNCTIONAL LITERACY TESTS														
ADULT PERFORMANCE LEVEL	7	0	1	1	2	3	1	1	1	1	Mi	2	1	
BASIC SKILLS READING MASTERY TEST	7	0	0	1	2	3	1	1	1	1	Wr	1	1	
READING/EVERYDAY ACTIVITIES IN LIFE	7	2	1	1	2	4	1	1	1	1	Wr	2	1	
WISCONSIN TEST OF ADULT BASIC EDUCATION, Life Coping Skills	6	0	0	1	2	2	1	1	0	1	Wr	2	1	
II. STANDARDIZED TESTS														
A. General Educational Development Performance Tests														
GENERAL EDUCATIONAL PERFORMANCE INDEX, Correctness & Effectiveness of Expression	5	1	1	1	1	3	1	1	0	1	Wr	1	1	
GENERAL EDUCATIONAL PERFORMANCE INDEX, Literary Interpretation	5	1	1	0	1	2	1	1	0	1	Wr	1	1	
B. Multiple Reading Skills Tests														
SRA READING INDEX	6	1	0	1	1	4	1	1	1	1	Wr	2	1	
WISCONSIN TEST OF ADULT BASIC EDUCATION, Word Meaning and Reading	6	0	0	0	1	2	1	1	0	1	Wr	2	1	

Note: The body of the table includes the ratings assigned to each test for individual criteria. A figure of zero on any criterion indicates non-compliance or lack of information.

The meanings of the symbols under "Response Mode" are as follows:
 "Or" = Oral; "Wr" = Written; and "Mi" = Mixed.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY																TOTAL GRADES	
Reliability			Repliability	Administration					Interpretation							Alternate Forms	Form Comparability				
Alternate Form	Test Retest	Internal Consistency		Training of Administrator	Administration Time	Clarity of Manual	Number of Administrators	Scoring	Training	Range	Diversity	Score Conversion	Interpretation	Real Life Skills	Validating Group			Racial/Ethnic/Sex Representation	Can Decisions Be Made?		
0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-2	0-1	0-1			
																			Good-Fair-Poor		
0	0	0	1	1	0	1	1	1	1	1	1	0	0	1	1	2	0	0	0	FGPP	
0	0	2	1	1	0	1	1	2	1	1	1	2	1	1	0	0	0	0	0	FGFF	
0	0	2	1	1	0	1	1	1	1	1	1	2	1	1	0	2	2	0	0	FGFF	
0	0	2	1	1	0	0	1	1	1	1	1	0	0	0	0	1	0	0	0	FFFP	
1	0	0	1	1	0	1	1	2	1	0	1	1	1	0	1	2	1	1	1	FFPF	
0	0	0	1	1	0	1	1	2	1	0	1	1	1	0	1	2	1	1	1	FFPF	
0	0	1	1	1	0	1	1	2	1	1	1	2	1	0	0	2	2	0	0	FGPF	
0	0	2	1	1	1	0	1	2	1	1	1	0	0	0	0	1	0	0	0	FFFP	

The entries under Total Grades summarize test performance on the four major criterion areas, in this order: 1. Measurement Validity, 2. Examinee Appropriateness, 3. Technical Excellence, and 4. Administrative Usability. Thus, the entry "PGFF" is to be interpreted:

- Poor for Measurement Validity
- Good for Examinee Appropriateness
- Fair for Technical Excellence
- Fair for Administrative Usability



TEST NAME	MEASUREMENT VALIDITY		EXAMINEE APPROPRIATENESS													
	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Comprehension		Format						Response Mode	Recording Response	Power Test	
					Content	Instructions	Layout	Illustration/Print Quality	Auditory Presentation	Other	Other	Other				
Rating Range →	0-2	0-4	0-1	0-1	0-2	0-4	0-1	0-1	0-1	0-1	0-1					
<i>C. Reading Comprehension Tests</i>																
ADULT BASIC LEARNING EXAMINATION, Reading	5	1	1	1	1	4	1	1	1	1	1	Wr	2	1		
BASIC OCCUPATIONAL LITERACY TEST, Comprehension	4	0	1	1	1	4	1	1	1	1	1	Wr	2	1		
TESTS OF ADULT BASIC EDUCATION, Comprehension	4	0	1	1	1	4	1	1	1	1	1	Wr	1	0		
<i>D. Spelling Tests</i>																
ADULT BASIC LEARNING EXAMINATION, Spelling	6	2	1	1	2	3	1	1	1	1	1	Wr	2	1		
<i>F. Vocabulary Tests</i>																
ADULT BASIC LEARNING EXAMINATION, Vocabulary	5	1	1	1	2	4	1	1	1	1	1	Wr	2	1		
BASIC OCCUPATIONAL LITERACY TEST, Vocabulary	4	0	1	1	1	4	1	1	1	1	1	Wr	2	1		
TESTS OF ADULT BASIC EDUCATION, Vocabulary	4	0	1	1	1	4	1	1	1	1	1	Wr	1	0		

Note. The body of the table includes the ratings assigned to each test for individual criteria. A figure of zero on any criterion indicates non-compliance or lack of information.

The meanings of the symbols under "Response Mode" are as follows:
 "Or" = Oral; "Wr" = Written; and "Mi" = Mixed.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY																	TOTAL GRADES
Reliability				Administration					Scoring	Interpretation								Alternate Forms	Form Comparability		
Alternate Form	Test-Retest	Internal Consistency	Reliability	Training of Administrator	Administration Time	Clarity of Material	Number of Administrators	Training		Range	Diversity	Score Conversion	Interpretation	Real Life Skills	Validating Group	Racial/Ethnic/Sex Representation	Can Decisions Be Made?				
0-3	0-3	0-1	0-1	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-2	0-1	0-1	Good-Fair-Poor	
0	0	2	1	1	0	1	1	2	1	1	1	2	1	0	0	2	1	1	1	FGFG	
0	0	0	1	1	1	1	1	2	1	0	1	2	0	0	0	2	2	1	1	PGPG	
0	0	0	0	1	0	1	1	2	1	1	1	2	1	0	0	0	2	1	1	PGPF	
0	0	2	1	1	1	1	1	2	1	1	1	2	1	0	0	2	1	1	1	FGFG	
0	0	1	1	1	1	1	1	2	1	1	1	2	1	0	0	2	2	1	1	FGPG	
0	0	0	1	1	1	1	1	2	1	0	1	2	0	0	0	2	2	1	1	PGPG	
0	0	0	0	1	1	1	1	2	1	1	1	2	1	0	0	0	2	1	1	PGPG	

The entries under Total Grades summarize test performance on the four major criterion areas, in this order: 1. Measurement Validity, 2. Examinee Appropriateness, 3. Technical Excellence, and 4. Administrative Usability. Thus, the entry "PGPF" is to be interpreted:

Poor for Measurement Validity
 Good for Examinee Appropriateness
 Fair for Technical Excellence
 Fair for Administrative Usability

TEST NAME	MEASUREMENT VALIDITY		EXAMINEE APPROPRIATENESS												
	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Comprehension		Format						Response Mode	Response	Power Test
					Content	Instructions	Layout	Illustration/Print Quality	Auditory Presentation	Coherence	Response Mode	Response			
Rating Range →	1-8	0-4	0-1	0-1	0-2	0-4	0-1	0-1	0-1	0-1	0-1		0-2	0-1	
<i>III. INFORMAL TESTS</i>															
<i>A. Oral Reading Tests</i>															
INFORMAL TEXTBOOK TEST	1	0	0	1	1	1	1	0	0	1	Wr	2	1		
IDAHO STATE PENITENTIARY INFORMAL READING INVENTORY	3	0	0	1	2	3	1	1	0	1	Or	2	1		
INDIVIDUAL READING PLACEMENT INVENTORY, Oral Paragraph Reading	4	2	0	1	1	1	1	1	0	1	Or	2	1		
INDIVIDUAL READING PLACEMENT INVENTORY, Present Language Potential	4	2	0	1	1	1	1	1	0	1	Or	2	1		
INFORMAL READING INVENTORY, Oral Reading	0	0	0	1	1	1	0	0	0	1	Or	2	1		
INFORMAL READING INVENTORY, Present Potential Level	0	0	0	1	1	1	1	0	0	1	Or	2	1		
<i>INITIAL TESTING LOCATOR TESTS</i>															
READING EVALUATION--ADULT DIAGNOSIS, Reading Inventory	1	0	1	1	1	3	1	1	1	1	Or	2	1		
<i>B. Reading Comprehension Tests</i>															
ADULT BASIC READING INVENTORY, Context Reading	0	2	1	1	1	3	1	1	1	1	Wr	2	1		
<i>C. Recognition or Discrimination Tests</i>															
ADULT BASIC READING INVENTORY, Sight Words	0	0	1	1	1	4	1	1	1	1	Wr	2	1		
ADULT BASIC READING INVENTORY, Sound and Letter Discrimination	0	0	1	1	1	4	1	1	1	1	Wr	2	1		

Note: The body of the table includes the ratings assigned to each test for individual criteria. A figure of zero on any criterion indicates non-compliance or lack of information.

The meanings of the symbols under "Response Mode" are as follows:
 "Or" = Oral; "Wr" = Written; and "Mi" = Mixed.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY																TOTAL GRADES	
Reliability			Repeatability	Administration						Interpretation						Alternate Forms	Form Comparability				
Alternate Form	Test-Retest	Internal Consistency		Training of Administrator	Administration Time	Clarity of Manual	Number of Administrators	Scoring	Training	Range	Diversity	Score Conversion	Interpretation	Real Life Skills	Validating Group			Racial/Ethnic/Sex Representation	Can Decisions Be Made?		
0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-2	0-1	0-1	Good-Fair-Poor	
0	0	0	0	1	1	0	1	1	1	1	0	2	1	0	0	0	1	0	0	PFPP	
0	0	0	0	1	0	1	1	1	1	1	0	1	1	0	0	0	2	1	1	PGPF	
3	0	0	1	1	1	1	1	1	1	1	1	2	1	0	0	0	2	1	1	FFFF	
3	0	0	1	1	1	1	1	1	1	1	1	2	1	0	0	0	2	1	1	FFFF	
0	0	0	0	1	0	0	1	1	1	1	1	0	1	0	0	0	1	0	0	PPPP	
0	0	0	0	1	0	0	1	1	1	1	1	2	1	0	0	0	1	0	0	PFPP	
0	0	0	0	1	1	0	1	1	1	1	0	0	1	0	0	0	1	0	0	PPPP	
0	0	0	0	1	0	0	1	1	1	1	1	2	1	0	0	0	1	1	1	PGPF	
0	0	2	1	1	1	0	1	2	1	1	1	2	1	0	0	0	2	0	0	PGFF	
0	0	2	1	1	1	0	1	2	1	0	1	2	1	0	0	0	0	0	0	PGFP	
0	0	2	1	1	1	0	1	2	1	0	1	2	0	0	0	0	1	0	0	PGFP	

The entries under Total Grades summarize test performance on the four major criterion areas, in this order, 1. Measurement Validity, 2. Examinee Appropriateness, 3. Technical Excellence, and 4. Administrative Usability. Thus, the entry "PGFF" is to be interpreted:

Poor for Measurement Validity
 Good for Examinee Appropriateness
 Fair for Technical Excellence
 Fair for Administrative Usability

TEST NAME	MEASUREMENT VALIDITY		EXAMINEE APPROPRIATENESS												
	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Comprehension		Format						Response Mode	Recording Response	Power To: 1
					Content	Instructions	Layout	Illustration's Print Quality	Auditory Presentation	Color	Response Mode	Recording Response			
Rating Range →	1-8	0-4	0-1	0-1	0-2	0-4	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-1
<i>C. Recognition or Discrimination Tests (Continued)</i>															
CYZYK PRE-READING INVENTORY	0	0	0	1	1	3	0	0	1	1	Wr	2	1		
HARRIS GRADED WORD TEST	0	0	0	1	1	1	0	0	0	1	Or	2	1		
INDIVIDUAL READING PLACEMENT INVENTORY, Auditory Discrimination	4	2	0	1	2	1	1	1	0	1	Or	2	1		
INDIVIDUAL READING PLACEMENT INVENTORY, Word Recognition	4	2	0	1	1	1	1	1	0	1	Or	2	1		
INFORMAL READING INVENTORY, Visual and Auditory Perception and Discrimination	0	0	0	1	2	1	1	0	0	1	Or	2	1		
INFORMAL READING INVENTORY, Word Recognition and Analysis	0	0	0	1	1	1	1	0	0	1	Or	2	1		
READING EVALUATION--ADULT DIAGNOSIS, Word Analysis	2	0	0	1	2	1	1	1	0	1	Or	2	1		
READING EVALUATION--ADULT DIAGNOSIS, Word Recognition	2	0	0	1	2	1	1	1	0	1	Or	2	1		
<i>D. Vocabulary Tests</i>															
ADULT BASIC READING INVENTORY, Word Meaning (Listening)	0	2	1	1	1	4	1	1	1	1	Wr	2	1		
ADULT BASIC READING INVENTORY, Word Meaning (Reading)	1	2	1	1	1	4	1	1	1	1	Wr	2	1		

Note: The body of the table includes the ratings assigned to each test for individual criteria. A figure of zero on any criterion indicates non-compliance or lack of information.

The meanings of the symbols under "Response Mode" are as follows: "Or" = Oral; "Wr" = Written; and "Mi" = Mixed.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY																TOTAL GRADES
Reliability				Administration						Interpretation						Alternate Forms	Form Comparability			
Alternate Form	Test Retest	Internal Consistency	Replicability	Training of Administrator	Administration Time	Clarity of Manual	Number of Administrators	Scoring	Training	Range	Diversity	Score Conversion	Interpretation	Real Life Skills	Validating Group			Racial/Ethnic/Sex Representation	Can Decisions Be Made?	
0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-2	0-1	0-1	
																				Good-Fair-Poor
0	0	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	PFPP
0	0	0	0	1	1	0	1	2	1	1	0	2	1	0	0	0	1	0	0	PPPP
3	0	0	1	1	1	1	1	2	1	0	1	2	1	0	0	0	2	1	1	FFFF
3	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	2	1	1	FFFF
0	0	0	0	1	0	0	1	2	1	0	1	2	1	0	0	0	1	0	0	PFPP
0	0	0	0	1	0	0	1	2	1	0	1	2	1	0	0	0	1	0	0	PFPP
0	0	0	0	1	1	0	1	2	1	0	1	2	1	0	0	0	2	0	0	PFPP
0	0	0	0	1	1	0	1	2	1	0	1	2	1	0	0	0	2	0	0	PFPP
0	0	2	1	1	1	0	1	2	1	1	1	2	1	0	0	0	1	0	0	PGFF
0	0	2	1	1	1	0	1	2	1	1	1	2	1	0	0	0	2	0	0	PGFF

The entries under Total Grades summarize test performance on the four major criterion areas, in this order: 1. Measurement Validity, 2. Examinee Appropriateness, 3. Technical Excellence, and 4. Administrative Usability. Thus, the entry "PGFF" is to be interpreted:

- Poor for Measurement Validity
- Good for Examinee Appropriateness
- Fair for Technical Excellence
- Fair for Administrative Usability

SUMMARY

In summarizing the results of these evaluations, it is useful to examine the different groups of tests--criterion-referenced tests of functional literacy, standardized tests, and informal tests--for strengths and weaknesses. Examining these specific groups of tests reveals several trends.

Criterion-referenced tests of functional literacy are the newest type of tests on the adult literacy testing market. Their availability is a result of the recent interest in teaching and measuring functional skills for adults. The newness of these tests is reflected by the fact that only one of the four tests has been made available for general dissemination by a commercial publisher although some of the others can be obtained from the authors.

The primary strength of the criterion-referenced tests lay in their appropriateness for the examinees. In general, these tests demonstrated sensitivity to the testing requirements of their target group. However, these tests were generally not rated highly otherwise; most of their ratings in the other three areas were "Fair." Particular points to which the test developers have not attended are establishing concurrent or predictive relationships, developing and testing alternate forms--including alternate form reliability--and determining test-retest reliability. Doubtless, failure to attend to these points stems from the problem of obtaining adequate time and money to accomplish all important tasks. For tests in the early stages of development, considerations of time and money would probably prevail regardless of the proclivities of test developers. Nevertheless, the need for adequate data remains, and should become the focus of subsequent efforts.

The standardized tests evaluated were generally accompanied by the most complete information concerning their development and use. The major strengths of these tests lay in their high appropriateness for examinees, and, to a lesser extent, their administrative usability. Also, they were often accompanied by extensive data describing results from field tryouts and other studies. On the negative side, much of the data presented did not strongly support the measurement validity or technical excellence of the tests. The psychometric quality of some tests clearly called for improvement.

Informal tests were found to vary the most in quality, and were the weakest overall. They presented a definite advantage in that they could be easily and quickly administered in a low threat environment. However, they entailed many problems. Most lacked adequate directions for administration, scoring, and interpretation; many included no description of design or development procedures. The inadequacy of this information was evidenced by the many "Poor" ratings the tests received. However, some informal tests had undergone substantially more testing than the others and therefore stood in contrast to the others in terms of quality.

Even though informal tests are typically used for a fairly limited purpose--the initial placement of students or the diagnosis of specific reading problems--their psychometric quality should not be ignored. In fact, the diagnosis of reading problems is so important that it ought to be done with thoroughly tested instruments. Omission of essential information limits the utility of any test and opens its results to question.

Continuing Development of Functional Literacy Measures

In addition to the tests listed earlier in this report, three other developmental efforts represent the continuing interest in developing measures of functional literacy. These efforts were not noted elsewhere in this report

because no tests associated with these efforts are available. They are briefly noted here to provide information about potential sources of literacy measures.

Current work is ongoing at the National Assessment of Educational Progress (NAEP). NAEP provides continual, direct assessment of educational outcomes nationwide in several learning areas for four age groups, including young adults.²⁶ Although NAEP does not publish its tests, it does periodically release some test items with attendant iter data. The measurement of reading is organized around nine themes: (1) words and word relationships, (2) visual aids, (3) written directions, (4) reference materials, (5) significant facts, (6) main ideas and organization, (7) inference, (8) critical reading, and (9) reading rate. Some of these themes are obviously more closely related to measuring functional skills than others. The items for each theme, as well as available data, can be found in several NAEP publications.²⁷

Work conducted by the Human Resources Research Organization (HumRRO) on the measurement of work-related literacy for military occupations provides a second source of information on measuring literacy.²⁸ The primary value of the HumRRO work lies not so much in the tests themselves--since even if available, they would be applicable only to military specialties; but rather in the comprehensive methodologies that HumRRO has established for determining the reading requirements of occupations. It would be particularly productive to apply their methodologies to determine what literacy skills are needed to function adequately in typical daily tasks and selected occupations or groups of occupations. Only by first determining such skills can educators of adults provide training in functional literacy.

A third effort involves the Adult Functional Reading Study conducted by the Educational Testing Service (ETS).²⁹ Initiated in 1970, the study began with a national survey to determine typical tasks of adults, construction of

a measurement instrument for determining the ability of adults in performing these tasks, and a national survey assessing the attainment of skills for such tasks. More recently, project staff examined the relationship of performance on functional reading tasks to decoding skills and performance on cloze tests, and developed an experimental test for assessing reading competency in schools.³⁰ The results of these studies are available, although to date no tests of functional literacy from the project have become accessible.

Conclusion

The reviews and evaluations in this report indicate that adult literacy testing is still a developing field marked by broad variety in the quality of available instruments. And despite the recent emphasis on reducing adult illiteracy in the United States, very few instruments have been developed and tested specifically for use with adults.

Much recent work in test development has concentrated on identifying important functional skills and constructing instruments to measure these skills. Further test development using the criteria suggested in this report can help make these tests highly appropriate for use with adult students. While much has been done, test users and developers must continue to combine their competence and efforts to produce instruments responsive to the testing needs of adults.

FOOTNOTES

- 1 Request for proposal to collect and evaluate tests of functional adult literacy (Office of Planning, Budget, and Evaluation, U. S. Office of Education, September 1974).
- 2 Winthrop R. Adkins, "Life Skills Education for the Adult Learner," *Adult Leadership*, 22, No. 2 (1973), 55-58, 82-84. Also Louis Harris, et al., "Survival Literacy Study." (New York, 1970) ED 068 813.
- 3 Such sources concerned with adult basic education tests include Kathleen Vanderhaar, Donald W. Mocker, Robert E. Leibert, and Vera Maass, *Tests for Adult Basic Education Teachers: "28 Suggestions for Classroom Teachers"* (Kansas City, Missouri: Center for Resource Development in Adult Education, University of Missouri-Kansas City, 1975); and Joan Fischer, Jane F. Flaherty, and Robert H. Arents, *Testing Guidelines for Adult Basic Education and High School Equivalency Programs* (Trenton, New Jersey: The Office of Adult Basic Education, Department of Education, 1973). For a review of all tests in print, see Oscar Krisen Buros, *Tests in Print II: An Index to Tests, Test Reviews, and the Literature on Specific Tests* (Highland Park, New Jersey: The Gryphon Press, 1974).
- 4 *Statistical Abstract of the United States* (Washington, D.C.: U. S. Government Printing Office, 1974).
- 5 *Literacy Among Youths 12-17 Years, United States* (DHEW Publication # (HRA) 74-1613, Baltimore, Maryland: National Center for Health Statistics, 1973).
- 6 Louis Harris, et al., "Survival Literacy Study."
- 7 Alex M. Caughran and John A. Lindlof, "Should the 'Survival Literacy Study' Survive?" *Journal of Reading*, 15, No. 6 (1972), 429-435.
- 8 Louis Harris, et al., "The 1971 Reading Difficulty Index: A Study of Functional Reading Ability in the U.S." (New York, 1971), ED 057 312.
- 9 Norvell Northcutt, Charles Kelso, and W. E. Barron, "Adult Functional Competency in Texas" (Austin, Texas: University of Texas Press, 1975).
- 10 John R. Bormuth, "Reading Literacy: Its Definition and Assessment," *Reading Research Quarterly*, 9, No. 1 (1974), 7-66.
- 11 John R. Bormuth, "Development of Readability Analysis," Report No. 7-0052 (Chicago: University of Chicago Press, 1969).
- 12 T. G. R. Bower, "Reading by Eye," *Basic Studies in Reading*, Ed. H. Levin and J. P. Williams (New York: Basic Books, 1970), pp. 134-146.

FOOTNOTES (Continued)

- ¹³ K. S. Goodman, "Analysis of Oral Reading Miscues: Applied Psycholinguistics," *Reading Research Quarterly*, 5, No. 1 (1969), 119-130.
- ¹⁴ Eleanor J. Gibson, "Learning to Read," *Theoretical Models and Processes of Reading*, Ed. Harry Singer and Robert B. Buddell (Newark, Delaware: International Reading Association, 1970), pp. 315-334.
- ¹⁵ Bormuth, 1974, pp. 7-66.
- ¹⁶ William S. Gray, *The Teaching of Reading and Writing* (Switzerland: United Nations Educational, Scientific, and Cultural Organization, 1969).
- ¹⁷ Request for proposal.
- ¹⁸ Request for proposal.
- ¹⁹ Thomas G. Sticht, Ed., *Reading for Working: A Functional Literacy Anthology* (Alexandria, Virginia: Human Resources Research Organization, 1975).
- ²⁰ Ronald P. Carver, "Reading as Reasoning: Implications for Measurement," *Assessment Problems in Reading*, Ed. Walter H. MacGinitie (Newark, Delaware: International Reading Association, 1973), pp. 44-56.
- ²¹ Carver, pp. 44-56.
- ²² Bormuth, 1974, pp. 7-66.
- ²³ Walter H. MacGinitie, "An Introduction to Some Measurement Problems in Reading," *Assessment Problems in Reading*, Ed. Walter H. MacGinitie (Newark, Delaware: International Reading Association, 1973), pp. 1-7.
- ²⁴ Ralph Hoepfner, et al., *CSE Secondary School Test Evaluations*, (Los Angeles: Center for the Study of Evaluation, Graduate School of Education, University of California, 1973); and Ralph Hoepfner, et al., *CSE-RBE Test Evaluations: Tests of Higher-Order Cognitive, Affective, and Interpersonal Skills* (Los Angeles: Center for the Study of Evaluation, Graduate School of Education, University of California, 1972).
- ²⁵ Wayne Otto, Evaluation Instruments for Assessing Needs and Growth in Reading," *Assessment Problems in Reading*, Ed. Walter H. MacGinitie (Newark, Delaware: International Reading Association, 1973), pp. 14-20.

FOOTNOTES (Continued)

26 *National Assessment of Educational Progress: General Information Yearbook* (Report #03/04-GIY, Washington, D.C.: U. S. Government Printing Office, 1974).

27 *National Assessment of Educational Progress, Reading and Literature: General Information Yearbook* (Report #02-GIY, Washington, D.C.: U. S. Government Printing Office, 1972).

28 John S. Caylor, Thomas G. Sticht, Lynn C. Fox, and J. Patrick Ford, *Methodologies for Determining Reading Requirements of Military Occupational Specialties* (Report # HumRRO-TR-73-5, Alexandria, Virginia: Human Resources Research Organization, 1973). Also Sticht, 1975.

29 Richard T. Murphy, *Final Report: Adult Functional Reading Study* (Grant #OEC-0-70-4791(508), National Institute of Education, U. S. Department of Health, Education, and Welfare, 1973).

30 Richard T. Murphy, *Supplement to Final Report: Adult Functional Reading Study* (Grant #OEC-0-70-4791 (508), National Institute of Education, U.S. Department of Health, Education and Welfare, 1975).

BIBLIOGRAPHY

- Adkins, Winthrop R. "Life Skills Education for the Adult Learner." *Adult Leadership*, 22, No. 2 (1973), 55-58, 82-84.
- Bormuth, John R. "Development of Readability Analyses." Report No. 7-0052. Chicago, Illinois: University of Chicago Press, 1969.
- "Development of Standards of Readability." Project No. 9-0237. Chicago, Illinois: University of Chicago Press, 1971.
- "Reading Literacy: Its Definition and Assessment." *Reading Research Quarterly*, 9, No. 1 (1974), 7-66.
- Bower, T. G. R. "Reading by Eye." *Basic Studies in Reading*. H. Levin and J. P. Williams, Eds. New York: Basic Books, 1970.
- Buros, Oscar Krisen. *Tests in Print II: An Index to Tests, Test Reviews, and the Literature on Specific Tests*. Highland Park, New Jersey: The Gryphon Press, 1974.
- Carver, Ronald P. "Reading as Reasoning: Implications for Measurement." *Assessment Problems in Reading*. Walter H. MacGinitie, Ed. Newark, Delaware: International Reading Association, 1973.
- Caughran, Alex M. and John A. Lindlof. "Should the 'Survival Literacy Study' Survive?" *Journal of Reading*, 15, No. 6 (1972), 429-435.
- Caylor, John S., Thomas G. Sticht, Lynn C. Fox, and J. Patrick Ford. *Methodologies for Determining Reading Requirements of Military Occupational Specialties*. Report # HumRRO-TR-73-5, Alexandria, Virginia: Human Resources Research Organization, 1973.
- Fischer, Joan, Jane F. Flaherty, and Robert H. Arents. *Testing Guidelines for Adult Basic Education and High School Equivalency Programs*. Trenton, New Jersey: The Office of Adult Basic Education, Department of Education, 1973.
- Gibson, Eleanor J. "Learning to Read." *Theoretical Models and Processes of Reading*. Harry Singer and Robert B. Buddell, Eds. Newark, Delaware: International Reading Association, 1970.
- Goodman, K. S. "Analysis of Oral Reading Miscues: Applied Psycholinguistics." *Reading Research Quarterly*, 5, No. 1 (1969), 119-130.
- Gray, William S. *The Teaching of Reading and Writing*. Switzerland: United Nations Educational, Scientific and Cultural Organization, 1969.
- Harris, Louis, et al., "Survival Literacy Study." New York, 1970. ED 069 813.

BIBLIOGRAPHY (Continued)

- , "The 1971 Reading Difficulty Index: A Study of Functional Reading Ability in the U. S." New York, 1971. ED 057 312.
- Hoepfner, Ralph, et al., *CSE Secondary School Test Evaluations*. Los Angeles: Center for the Study of Evaluation, Graduate School of Education, University of California, 1973.
- , *CSE-RBS Test Evaluations: Tests of Higher-Order Cognitive, Affective, and Interpersonal Skills*. Los Angeles: Center for the Study of Evaluation, Graduate School of Education, University of California, 1972.
- Literacy Among Youths, 12-17 Years, United States*. DHEW Publication #(HRA)74-1613, Baltimore, Maryland: National Center for Health Statistics, 1973.
- MacGinitie, Walter H. "An Introduction to Some Measurement Problems in Reading." *Assessment Problems in Reading*. Walter H. MacGinitie, Ed. Newark, Delaware: International Reading Association, 1973.
- Murphy, Richard T. *Final Report: Adult Functional Reading Study*. Grant # OEC-0-70-4791(508), National Institute of Education, U. S. Department of Health, Education and Welfare, 1973.
- , *Supplement to Final Report: Adult Functional Reading Study*. Grant # OEC-0-70-4791(508), National Institute of Education, U. S. Department of Health, Education, and Welfare, 1975.
- National Assessment of Educational Progress: General Information Yearbook*. Report #03/04-GIY, Washington, D.C.: U. S. Government Printing Office, 1974.
- National Assessment of Educational Progress, Reading and Literature: General Information Yearbook*. Report #02-GIY, Washington, D.C.: U. S. Government Printing Office, 1972.
- Northcutt, Norvell, Charles Kelso, and W. E. Barron. "Adult Functional Competency in Texas." Austin, Texas: University of Texas Press, 1975.
- Otto, Wayne. "Evaluation Instruments for Assessing Needs and Growth in Reading." *Assessment Problems in Reading*. Walter H. MacGinitie, Ed., Newark, Delaware: International Reading Association, 1973.
- Request for proposal to collect and evaluate tests of functional adult literacy. Office of Planning, Budget and Evaluation, U. S. Office of Education, September 1974.
- Statistical Abstract of the United States*. Washington, D.C.: U. S. Government Printing Office, 1974.
- Sticht, Thomas G., et al. *Auditing and Reading: A Developmental Model*. Alexandria, Virginia: Human Resources Research Organization, 1974.

BIBLIOGRAPHY (Continued)

----- . *Reading for Working: A Functional Literacy Anthology.* Alexandria, Virginia: Human Resources Research Organization, 1975.

Vonderhaar, Kathleen, Donald W. Mocker, Robert E. Leibert, and Vera Maass.

• *Tests for Adult Basic Education Teachers: "28 Suggestions for Classroom Teachers."* Kansas City, Missouri: Center for Resource Development in Adult Education, University of Missouri-Kansas City, 1975.