DOCUMENT RESUME

ED 210 317 TM 820 008

AUTHOR Anderson, Beverly L.
TITLE Guide to Adult Functional Literacy Assessment Using Existing Tests.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE Jun 81
GRANT 400-80-0105
NOTE 113p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Adult Counseling; *Adult Literacy; Adults; Basic Skills; *Functional Literacy; Minimum Competencies; *Personnel Evaluation; Reading Tests; *Test Selection; Writing Evaluation

ABSTRACT

This is a guide designed for professionals who need to assess the literacy ability of adults. The three general categories of literacy skills which are distinguished include generic literacy skills, everyday skills, and job-related skills. Functional literacy has been defined in various ways, therefore, the assessment of it is very difficult. The purpose of literacy assessment must be established before selecting the appropriate instrument. The criteria to be considered are usability, validity, and reliability. The decisions to be made before a specific test or assessment approach are identified relate to the categories of literacy, the purpose for testing, the uses and users of test results, examinee characteristics and logistics. General guidelines for reviewing existing published or unpublished tests are examined. These include preliminary screening and technical quality review. Assessment of everyday literacy activities may be determined through the use of one of several published tests designed for this purpose. Assessment of on-the-job literacy is usually limited to tests of clerical ability. Appendices to the guide include suggested readings on definitions of functional literacy, published tests of basic writing and reading skills, and procedures for scoring writing samples. (DWH)

GUIDE TO ADULT FUNCTIONAL LITERACY ASSESSMENT Using Existing Tests

Functional Literacy Project

Northwest Regional Educational Laboratory

GUIDE TO ADULT FUNCTIONAL LITERACY ASSESSMENT
Using Existing Tests

Beverly L. Anderson

June 1981

Functional Literacy Project

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue Portland, Oregon 97204

June 1981

This work is published by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work contained herein has been developed under grant 400-80-0105 with the National Institute of Education (NIE), Department of Education. The opinions expressed in this publication do not necessarily reflect the position of the National Institute of Education, and no official endorsement by the Institute should be inferred. Mention of trade names, commercial products or organizations does not imply endorsement by the U.S. Government.

NIE Project Officer: Monte Penney

Acknowledgements

Appreciation is expressed to many people for their contribution to this Guide. Stephen Reder and Karen Reed Green of the Functional Literacy Project were heavily involved in designing the structure of the Guide as well as in reviewing and contributing to early versions. Jean Greene conducted interviews with practitioners to help determine what content should be included. Sharon Tamura and Nancy Bridgeford assisted in identifying and reviewing tests. Barbara Hejtmanek and Laura Hopkins provided secretarial support. Ann Adelman and Nikki Sullivan, adult educators in Portland, provided helpful input through review of drafts of the Guide. And Vicki Spandel once again demonstrated her excellent skills in the final editing of the Guide. To all of you, thanks.

Beverly Anderson

Table of Contents

List of Figures

Preface

If you are a career counselor, community college staff member, employment program director, teacher of adult basic education or personnel staff member needing to assess the literacy ability of adults, then this Guide was written with you in mind. This Guide--the first in a projected series—is meant to help you determine when to use existing tests to assess adult literacy. This is the most economical approach to testing. We recognize, however, that extant literacy tests often do not adequately reflect the multiple functions and social meanings of literacy. Therefore, future publications will show how to develop one's own literacy assessment methods to accommodate special needs.

For purposes of this discussion, three general categories of literacy skills are distinguished: basic (generic) literacy skills, everyday skills, and job-related skills. The Guide offers specific examples of appropriate existing tests for measuring skills in each category. The status of literacy assessment in these three situations is also reviewed.

The following examples illustrate the kinds of situations for which each chapter can offer guidance.

CHAPTER	EXAMPLE SITUATION
I. Introduction	
II. Decisions Prior to Test Selection	You are beginning to establish or are about to revise your general procedure and philosophy of assessment.
III. Test Review Procedures	You have determined that an existing test is likely to be useful in your situation but you are not sure how to select from among the available ones.
IV. After Test Selection	You have selected a test but are not sure how to determine what level of performance is acceptable and/or how to interpret results.
V. Basic Literacy Assessment	You are a counselor at a community college responsible for assessment of potential enrollees. You want to assess their literacy skills to assist them in selecting courses.

CHAPTER		EXAMPLE SITUATION
VI.	Everyday Literacy Activities Assessment	You are an instructor for a community college English as a Second Language Program. You need to assess the extent to which new students have the literacy skills necessary to perform everyday life tasks.
VII.	On-the-Job Literacy Assessment	You are a CETA intake officer or an industrial personnel officer who needs to determine if a person can perform literacy tasks required for a specific job.

The impetus for this Guide comes from research being conducted by the Functional Literacy Project at the Northwest Regional Educational Laboratory. In 1975, the Laboratory published Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments.* That review highlighted the diverse definitions of functional literacy and the consequent confusion about how to assess it. Although it seemed likely that different definitions of literacy would be appropriate in various settings, there was little rationale available for selecting one particular definition or measurement method over another in a given situation. To make that selection easier, the Functional Literacy Project began to study how adults use literacy in their everyday lives. To get an insider's perspective, the project relies on ethnographic methods: actually living in a community, observing and participating in as many daily activities as possible, and interviewing not only community members, but public service agency and private business representatives. As staff gain insight regarding the social and contextual variations in functional literacy, they intend to prepare additional, practical assessment guidelines.

---

*Nafziger, D., et al. Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments. Portland, OR: Northwest Regional Educational Laboratory, 1975.

Chapter I

Introduction

CHAPTER I

INTRODUCTION

Individuals, groups and institutions define functional literacy in various ways depending on their special needs and interests. For example, public schools often view functional literacy in terms of students' demonstrating established competencies required for everyday life. These competencies may incorporate basic reading, writing and computation skills. Examples include writing letters, locating government agencies, following voting procedures, preparing tax forms, completing job applications and giving oral directions.

The definition of functional literacy used in the National Health Survey of 1973, on the other hand, refers to the average performance of fourth graders as measured on the Brief Test of Literacy. And a national survey of employers considered literacy as the integration of mathematical and linguistic skills necessary for filling out a job application, doing filing, conducting routine correspondence, monitoring inventories and writing articulately. The Adult Performance Level Project took yet another tack, establishing functional literacy levels according to the test performance of people at various levels of income, occupational prestige and educational attainment. And the National Census characterizes a literate individual as one who has completed six or more grades in school and has "the ability to read and write a simple message in any language."*

This wide variety of definitions makes assessing functional literacy particularly difficult. As might be expected, there are as many assessment techniques and performance criteria as there are definitions, and the task of matching the appropriate technique to a particular assessment situation is complex. For example, a standardized reading comprehension test designed for use in high school is often used to determine if a person has the necessary literacy skills to work in an industrial setting where the only literacy task is reading an instructional manual. This is clearly a mismatch between situation and approach.

To further complicate matters, the variety of purposes and meanings ascribed to literacy are not carefully considered in many instructional and assessment situations. For example, it is often assumed that adults who have literacy skills will apply them as needed. Research suggests that this is not at all the case; for many complex reasons, an adult will apply literacy skills in some situations but not in others. Also, "literacy problems" arise from mismatches among individuals' values or expectations and those underlying literacy tests and assessments. For example, personal ideas about who should read in church or when one should write a memo rather than make a phone call can affect performance on or reaction to a literacy assessment that touches on these matters.

---

*The reader interested in further discussion of alternative functional literacy definitions is referred to Appendix A for an overview of articles which highlight definitional problems, examine the issues involved and discuss the benefits and limitations of various definitions.

Value systems are often deeply established, the result of social concepts and practices that have evolved throughout the history of a community or a group of people.*

Moreover, many individuals use alternative means to cope with the lack of literacy skills: use of nonprint media, for instance, or assistance from more literate people. Thus, literacy assessment is not always a true reflection of competence. Persons who have found a means of circumventing their lack of literacy skills may perform certain tasks at a higher level than expected.

Presently, literacy assessment research is not sufficiently advanced for us to give specific guidelines on handling these issues. However, this Guide, based on our research findings to date, provides the reader with our best information on selecting and using existing literacy assessment materials. We wish to emphasize the importance of clarifying the purpose for assessment first. Criteria for "best available instrument" include usability, validity and reliability for a given purpose.

---

*Reder, S., & K. R. Green. Comparative aspects of the community structure of literacy: Annual report of the Functional Literacy Project. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Reder, S., & K. R. Green. Literacy as a functional component of social structure in an Alaska fishing village. Paper presented at the 78th Annual Meeting of the American Anthropological Association. Cincinnati, December 1979.

Reder, S., & K. R. Green. Social meanings of literacy in an Alaska fishing village. Paper presented at the First Annual Ethnography in Education Research Forum, University of Pennsylvania. Philadelphia, March 1980.

Chapter II

Decisions Prior to Test Selection

CHAPTER II

DECISIONS PRIOR TO TEST SELECTION

This chapter addresses five important categories of decisions that must precede test selection. These decisions relate to

- Category of literacy assessment
- Purpose for testing
- Uses and users of test results
- Examinee characteristics
- Logistics

What General Category of Literacy Assessment is of Concern?

In identifying the general nature of literacy to be assessed, one is essentially setting forth an operational definition of literacy for the situation. The latter chapters of this Guide distinguish three categories of literacy skills: basic (generic) literacy skills, everyday literacy and on-the-job literacy. One must first decide which of these situations is appropriate.

Tests designed to measure basic (generic) literacy skills cover skills and knowledge which are not specific to a given context, e.g., general vocabulary; phonetic analysis skills, including the ability to sound out new words; recognition of antonyms and synonyms; and the ability to write and comprehend simple sentences. Generally, these tests involve reading to gather information (reading to learn) rather than reading to perform a task (reading to do).

Underlying the concept of basic skills measurement is the assumption that people can transfer skills from one situation to another. For certain people, such transfer does not seem to occur. For this and other reasons, the literacy assessment is best conducted within a restricted context or for a restricted purpose.

This leads to our other two literacy assessment categories: everyday literacy and on-the-job literacy. Certain tests have been designed specifically to measure literacy within one of these two contexts. Such tests are far more useful and appropriate in these contexts than tests intended to measure a broad array of general literacy skills. By identifying the testing purpose and context in advance, one avoids being burdened with excess or irrelevant information: e.g., learning that a person can read at the fourth grade level when what really matters is whether the person can understand the instruction manual for operating a sewing machine.

Specific skills within each literacy assessment category (generic, everyday, on-the-job) are further delineated in Chapters V through VII.

What Are the Purposes for Assessing?

Nearly all literacy assessment is done to facilitate some educational decision, each of which requires different kinds of information. For example, it would be inefficient to measure basic reading skills in detail if an administrator needed only a general idea of whether students could comprehend written materials. In other words, testing is only efficient when one knows in advance what kinds of decisions are to be made.

Purposes for assessment can be divided into three broad categories: instructional management, entry-exit decisions and program planning and evaluation decisions. Let's take a closer look at each category.

Instructional Management

Instructional management includes (1) the diagnosis of individual learner strengths and weaknesses, (2) student placement and (3) educational and vocational student guidance.

Diagnosis. The most frequent reason teachers test is to diagnose the strengths and weaknesses of individual students. Tests and other performance indicators can help pinpoint a student's current level of development and assist the teacher in selecting the next appropriate instructional unit. A test used for diagnosis must relate closely to course content and should be very detailed. Results must be readily comprehensible to teachers and students.

Placement. Whereas diagnosis helps determine which individual instructional units are most appropriate for a student, within a course, placement tests determine which course or class best meets a student's needs. Placement generally demands less detailed information than diagnosis. While such information is of greatest value to program adminstrators, instructors may also find placement test results useful in determining the general level of students' skills.

Guidance. Guidance tests help predict the chances of success and satisfaction in various educational and vocational programs. Such tests need to measure skills which are predictive of success. Tests which give the examinee an indication of his/her standing relative to other examinees is often useful. In guidance testing, it is the individual being assessed who makes decisions—with assistance from a guidance counselor.

Entry-Exit Decisions

Very carefully controlled testing can assist in entry or exit decisions in an educational or employment context. For example, literacy tests may be administered for the purposes of (1) selection for employment or (2) certification of competencies for occupational licensing or course completion.

Selection. Selection tests help an employer determine the best qualified person for a job or help an educator determine what applicants are most likely to succeed in or benefit from an educational program. Selection decisions have serious implications for a person's future and are under close legal scrutiny. Therefore, a test used for selection purposes must (1) clearly focus on these skills and knowledge essential for future success, and (2) rank order students in terms of relevant skills and knowledge to identify those most likely to succeed.

Certification. Tests often play an important role in certifying those who have attained a minimum acceptable level of educational development. For example, a test may be used to certify attainment of minimally acceptable skills for receipt of a high school diploma. Certain professions also use a test to certify competence to practice that profession. In either case, examinees must pass the test to be certified minimally competent. Therefore, the test must thoroughly and accurately cover all relevant competencies.

Programmatic Decisions

Tests also facilitate programmatic decisions. For instance, test data that reflect the needs of a group can serve as the basis for developing a new program, allocating resources or evaluating existing programs. Testing for these purposes falls into three categories: (1) survey assessment, (2) formative program evaluation and (3) summative program evaluation.

Survey assessment. One established use of testing in education is surveying student achievement and exploring trends over time to make programmatic planning decisions. This kind of testing is usually designed to raise questions rather than answer them. For example, why are writing scores gradually declining among entering community college students? Why are the national trends in literacy among adults changing? A survey assessment should encompass the most predominant or significant elements of instruction for the population being surveyed; it cannot measure all the skills and knowledge taught in any one location or class. Survey information, though necessarily broad in nature, helps educational administrators set policy and allocate resources.

Formative evaluation. The goal of formative evaluation is to determine which instructional units or features of a program (e.g., remedial reading) are effective, and which need revision. The decision to revise is made by program teachers and administrators. Tests covering specific interim and long-term program outcomes are administered during the program's operation to help shape the program during its formative stages.

Summative evaluation. Summative evaluation reveals a program's overall merit, thus suggesting to teachers and administrators whether that program should be modified, continued as is or terminated. Tests designed to assess students' performance on final learning outcomes of a program are typically part of such an evaluation. Such tests are generally given prior to and following instruction, so that results can be compared and the impact of instruction determined.

How and By Whom Will Test Results Be Used?

Before selecting a test, it is important to decide who will use the test scores. The kind of information needed by students may differ markedly from that needed by teachers, administrators or counselors. These groups differ not only in their interest, but also in their capability and experience in interpreting test results. For example, a student or teacher may be interested only in using an individual student's score to determine what should be taught next. An administrator, on the other hand, may want only group scores to justify financial support for further literacy instruction.

Deciding who will use the results is easy if the testing purpose has been carefully delineated. Deciding what type of test score to report, however, may not be so easy. Test scores fall into two general categories: norm referenced and criterion referenced. Norm referenced test scores reflect how one examinee compares to another examinee or group. The most common types of norm referenced scores are percentiles, grade equivalent scores and stanines. Criterion referenced scores compare an examinee's performance to prespecified criteria, without regard to the performance of others. Criterion referenced scores are frequently expressed as "percent of objectives mastered" and "percent of items correct for each learning objective."

Criterion referenced test scores are particularly useful for placing adults in a literacy program or an employment position. These scores illustrate the extent to which a person has acquired certain specific skills. The utility of norm referenced scores depends on the purpose for testing and on how well defined and appropriate the norming group is.

Most published tests are normed on a national sample of people selected by grade level, age or occupation. Scores may be reported for subsets of the total norming group, e.g., rural vs. urban or Northeast vs. other regional groups. Norm referenced scores of this type may be useful in broad program assessment. For example, community college administrators may want to know how their students compare to community college students nationwide.

In many situations, however, normative information is of limited use in testing adults. For example, suppose a survey showed that 60 percent of the army personnel tested received scores below a grade equivalent of 7.0.* What would this mean? Very little, because this score tells us nothing about what tasks the examinees can or cannot do, nor how their performance compares to that of other adults.

---

*The implicit norm groups used for grade equivalent scores are children in elementary and secondary schools.

Norm referenced scores are frequently thought to imply certain standards. In particular, it is often assumed that an above average score is satisfactory, and a below average one unsatisfactory. By definition, half the people tested must be above average and half below. Whether "average" performance is satisfactory, below satisfactory or superior is another issue entirely. This decision requires determining what specific tasks a person needs to be able to do, and what levels of performance within each task are acceptable.

Grade equivalent scores present a particular problem. On many standardized tests the grade equivalent score scale is constructed in such a way that a score as high as 6.0 may be equivalent to the chance score.** On such a test, results showing that an adult is reading at the fifth grade level only indicate that the test was too difficult. There is no way to know whether the person has some reading skills or was simply guessing in response to test questions.

In short, national norm referenced scores may be of little use in adult literacy assessment. However, local norms may be worth consideration. Local norms, which can be developed on a criterion referenced or nationally normed test, allow you to compare individuals' test performance to that of a group you select. For example, you may want to determine local norms for the incoming students in your particular community college.***

Various norm referenced and criterion referenced scores are described in more detail in Appendix B, along with the advantages and disadvantages of each. Additionally, many test manuals and measurement textbooks describe the differences among these scores.

What Examinee Characteristics Influence Test Selection?

Too many tests attempt to measure technical literacy without regard for the functions and social meanings associated with literacy in the examinee's world. If an examinee is accustomed to reading only the newspaper and novels, a reading comprehension test using reading passages from scholarly articles may not allow the examinee to display his or her true ability to comprehend written material. Test materials unrelated to examinees' past experience, or lacking cultural or social significance, often significantly mask ability.

**The chance score is the number of items one would expect a person to answer correctly on a multiple choice test merely by guessing.

***Helpful information on developing local norms can be found in Chapter 5 (Local Norms) of the Evaluator's References, Vol. II, ESEA Title I Evaluation and Reporting System. Mountain View, CA: RMC Research Corporation. August 1980.

Test format is also crucial. For example, a recent Indochinese immigrant may never have seen a multiple-choice test. He or she may be able to read a paragraph orally and respond to oral questions, but not do well on a multiple-choice test simply because the format is unfamiliar. An alert examiner can sometimes sense when certain test characteristics may distort results. In some cases, an examinee may be taught the mechanics of test taking prior to testing, thus eliminating this problem. In other cases, it may be best to use another assessment approach.

What Logistical Considerations Are Most Important?

Certain practical logistical factors must be considered. Probably the most important of these are test length, type of administration and cost.

First, the test must be of an appropriate length for a given situation. A series of tests designed for frequent classroom diagnostic use, for example, must be fairly short, whereas a test used for a once-a-year survey assessment may be longer.

Second, some tests must be individually administered. The additional costs of individual (vs. group) administration need to be carefully weighed against the increased value of the resulting scores.

Third, when considering cost, one must first note whether test manuals, booklets and answer sheets may be purchased separately. Some items may be reusable (e.g., test booklets) whereas others are not (e.g., answer sheets). Further, having the test scored by the publisher may be costly, depending on the type of reports desired. Usually a variety of reporting options are available.

Summary

This chapter has presented key decisions a person must make before identifying a specific test or assessment approach to use to assess literacy. Those decisions relate to (a) the categories of literacy, (b) the purpose for testing, (c) the uses and users of test results, (d) examinee characteristics, and (e) logistics.

Chapter III

Test Review Procedures

CHAPTER III

TEST REVIEW PROCEDURES

Once the issues raised in Chapter II are resolved, one is ready to select or develop an assessment approach which will provide the desired information. Basically there are two choices: using an existing test or developing a new one. The most economical approach is, of course, to use an existing test--provided a high quality, appropriate instrument can be obtained. Existing tests are available from publishers or from people who have developed their own for a specific purpose but have never published them. While published tests are generally extensively reviewed, pilot tested and revised before publishing, they may not have all the characteristics important for a particular situation. One must carefully review either published or unpublished tests, therefore, to be sure they can provide useful results.

This chapter provides general guidelines for reviewing existing tests. Chapters V-VII focus on the application of these guidelines in specific situations.

Preliminary Screening

The first step in test selection is to locate existing tests and identify those that come close to measuring the specific skills one desires to measure. The appendices referenced in Chapters V-VII of this handbook list tests which will serve as a starting point for this search. Several reference books are also useful:

The Mental Measurement Yearbook (Buros, 1978) reviews hundreds of existing tests. It is updated about every six years. Tests in Print (Buros, 1974) provides descriptive information on all published tests. It is updated about every ten years. A recent publication entitled A Consumer's Guide to Writing Assessment (Bridgeford & Stiggins, 1981) provides very helpful descriptions of existing writing tests, and Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments (Nafziger, et al., 1975) provides comprehensive technical reviews of many existing instruments.

Periodically, documents in the ERIC* system contain test reviews. Another helpful source is the monthly publication entitled News on Tests, available from the Educational Testing Service (ETS).** ETS also houses a test collection. They will provide, upon request, summaries of existing tests categorized by content area. Helpful suggestions can also be obtained from community college program personnel (e.g., Adult Basic Education, high school equivalency, career counseling, refugee or tutoring programs), university libraries, CETA offices and state departments of education or directly from publishers.

---

*ERIC is Educational Resources Information Center, an educational information data base providing access to fugitive documents.

**For further information on this publication and Test Collection Service, contact Test Collection, Educational Testing Service, Princeton, NJ 08541.

It is not easy to locate unpublished tests since no central source exists. They are best located by word of mouth and checking sources such as ERIC. Often the unpublished tests are associated with a research study reported in the ERIC system.

Before reviewing a test itself, one should first review the descriptive information provided through the publisher or one of the reference books mentioned in the preceding paragraphs. Once a small number of relevant tests have been identified, the more time consuming review for technical quality can begin.

When reviewing test references, keep in mind that tests are usually revised every five to ten years. In almost all cases, the most recent edition should be considered. Also, many tests are part of a series spanning a wide age range (e.g., K-adult). Only the level appropriate for the group with which it is to be used need be reviewed. In some cases, though, you may want to review several levels to determine which is best for your situation. For example, in an adult basic education class, a diagnostic reading test designed for junior high students may cover precisely the correct level of skills.

An important consideration is whether the test is norm and/or criterion referenced. One may be more useful. If a norm referenced score is desired, it is particularly important to determine if the norming group is relevant for the situation.

Information on the intended uses of the test, skills and knowledge measured, response format (e.g., multiple choice), administration time and available scoring procedures (hand scored vs. machine scored) will also help you determine quickly whether the test merits further consideration.

This preliminary screening should yield a handful of potentially useful tests. The next step is a careful technical review.

Technical Quality Review

A technical quality review of existing tests demands a set of criteria. Three general categories of criteria should be used when reviewing functional literacy tests: validity, usability and reliability. Validity refers to the extent to which a test measures the skills and knowledge the user intends it to measure. Usability refers to the extent to which a test is suitable for the context in which it will be used. Reliability refers to the extent to which a test measures a trait consistently.

When reviewing adult functional literacy tests, one should ask specific questions relating to each of these three categories. Suggested questions are given on the rating form in Figure 1. Following the form are instructions for assigning points (Figure 2). Extensive rationales for these questions are not provided. In cases where the reasons are not intuitively obvious, the reader is referred to Guidelines for Selecting Basic Skills and Life Skills Tests (Anderson, Stiggins & Hiscox, 1980) or college level measurement textbooks.

Once tests have been rated on each of the technical quality questions, an overall judgment must be made. In almost all cases, the two critical questions will be:

- Do the test items measure the specific learning objectives or skills that need to be assessed? (Question 1 under Validity)

- Is the appropriate type of score reported for the intended use? (Question 6 under Usability)

If a test does not meet these criteria, the other aspects of its quality will be of little consequence. Scores of tests which meet these criteria should be summed, and the test with the highest score selected (assuming all other criteria are considered of equal importance).

FIGURE 1
TECHNICAL REVIEW OF TESTS

Rate each test of interest using the rating system described in the instructions in Figure 2.

Validity

1. Do the test items measure the specific learning objectives or skills that need to be assessed?

   List learning objectives or skills to be assessed in the left hand column of the chart below. Put the name of each test being reviewed in a numbered box at the top of one of the right hand columns. (Repeat on each page of the test review chart.) Under each test indicate the number of items measuring each objective. Review the test manual or the actual test to determine if each objective or skill is measured. The publisher may indicate in the manual what skill each test item measures. If not, examine each test item to determine what skill or objective it measures.

	Test Name			
Learning objectives or skills to be assessed	1	2	3	4
1.				
2.				
3.				
4.				
5.				
6.				

For the remainder of the questions, assign points as described in the Instructions for Assigning Ratings in the Technical Review of Tests Form (Figure 2). The number in parentheses after each question indicates the number of possible points.

FIGURE 1 (continued)

	Test Name			
	1	2	3	4
2. Were the test items developed in a systematic and rigorous manner so that the content is adequate and bias is minimized? (6)				
3. Were any empirical procedures used for screening or selecting items to ensure that items are measuring what they were designed to measure, are understandable, contain reasonable answers, and are free of ambiguous alternative answers and unnecessarily complex language? (3)				
4. Is the validating group representative of the population with which the test is to be used? (2)				
5. Are any special validity studies reported or specifically referenced? (2)				

Usability

1. Are the test items suitable for adults with limited literacy skills? (2)				
2. Are instructions to the test administrator clear and complete? (2)				

FIGURE 1 (continued)

	Test 1	2	3	4
3. Are instructions to the examinee written in clear and understandable terms? (2)				
4. Is the test formatted clearly? (2)				
5. Is there a simple way for examinees to record their responses? (2)				
6. Is the type of score reported useful for my situation? (2)				
7. Is the process of converting raw scores simple and does it yield scores which are easily interpreted? (2)				
8. Is the amount of time required for testing appropriate? (2)				

Reliability

1. Is reported reliability for major subtests and/or total test scores sufficiently high? (3)				
2. Are the scoring procedures clear and complete, thus ensuring reliable scores? (2)				

FIGURE 2
INSTRUCTIONS FOR ASSIGNING RATINGS
ON THE TECHNICAL REVIEW OF TESTS FORM

I. Validity--the extent to which the test measures what the user intends it to measure.

1. Do the test items measure the specific learning objectives or skills that need to be assessed?

To answer this question, list the learning objectives to be assessed. Then go through the test item by item and determine which objective, if any, each item measures. Adequate coverage requires at least three items measuring a given objective. If the test measures 75 percent of your objectives, the coverage may be adequate for placement decisions. It would not necessarily be adequate for diagnostic decisions.

2. Were the test items developed in a systematic and rigorous manner so that the content is adequate and bias is minimized? (6 possible points)

Award points for this criterion on the following basis:

1) Relation of items to specific objectives:

4 points if test items generally relate to specific objectives or criteria (tasks from a task analysis)

1 point if items relate to general content areas

0 points if items do not generally relate or if objectives or criteria are lacking

2) Item development procedure:

2 points if procedures for developing test specification and items are described in detail

1 point if reference is made to use of a specific rigorous item development procedure but details are not given

0 points if no information is provided on item selection

3) Bias review procedures:

2 points if statistical analyses and reviews for racial, sexual and cultural perspectives were conducted

1 point if either statistical anaylses or technical reviews were conducted

0 points if neither bias review procedure was used

3. Were any empirical procedures used for screening or selecting items to ensure that items are measuring what they were designed to measure, are understandable, contain reasonable answers, and are free of ambiguous alternative answers and unnecessarily complex language? (3 points possible)

Empirical procedures include item analysis, review by juries of experts, review of item difficulties, criterion-group analyses or factor analyses. Award points on the following basis:

3 points if more than one appropriate method was conducted and reported in detail

2 points if more than one method was used but insufficient reporting is done to assess its appropriateness, or if one appropriate method is reported in some detail

1 point if it is stated that one method was used

0 points if no information is given

4. Is the validating group representative of the population with which the test is to be used?* (2 points possible)

Include the following considerations in the evaluation of validating group representativeness:

(1) Were both males and females included in the validating group?

(2) Were all major ethnic groups represented in the validating group?

---

*The rating system used on this criterion assumes that the test user is working with a broad cross section of adult learners and wants assurance that it is appropriate for such a range of people. There may be cases where a more parochial test is desired. This criterion would then be modified to fit that population.

FIGURE 2 (continued)

(3) Was the validating group a nationally representative sample in terms of population density characteristics (e.g., urban, suburban, rural, etc.?), and geographic representation of the age group for which the test was designed?

(4) Was the sample obtained through cluster, stratified or random rather than incidental sampling?

(5) Was the validating done less than five years ago?

Award points based on the following:

2 points if answers to at least four of these questions were "yes"

1 point if answers to two or three questions were "yes"

0 points if fewer than two answers were "yes' or insufficient information was provided to determine the answers.

5. Are any special validity studies reported or specifically referenced? (2 points possible)

Include these two questions here in the consideration of this criterion:

(1) Has anyone examined the relationship between this test and other measures of adult literacy?

(2) Are any studies referenced or reported which examine how this test predicts success in other educational programs, life survival tasks or jobs?

Award points as follows:

2 points if at least one study is referenced for both questions 1 and 2 above

1 point if only one study is referenced for either question 1 or 2 above

0 points if no studies are referenced for either question.

FIGURE 2 (continued)

II. Usability--the extent to which a test is suitable for the context in which it will be used.

1. Are the test items suitable for adults with limited literacy skills? (2 points possible)

Award points on the following basis:

2 points if all items appear inoffensive, reasonably appropriate in difficulty and intellectually stimulating (regardless of test content)

1 point if most items appear appropriate

0 points if many items are judged inappropriate because (1) they are inappropriate or offensive to special groups or (2) they are too dull or insultingly simplistic.

2. Are instructions to the test administrator clear and complete? (2 points possible)

Award points on the following basis:

2 points if the instructions to the administrator clearly and completely describe (1) the materials to have available, (2) the conditions in the room where testing was to occur, and (3) time allotments for testing, if any

1 point if any of the above is unclear

0 points if more than one of the above is unclear.

3. Are instructions to the examinee written in clear and understandable terms? (2 points possible)

Award points on the following basis:

2 points if (1) instructions clearly and precisely describe the examinee's task and (2) sample items are included which effectively illustrate the task

1 point if instructions clearly and precisely describe the examinee's task but no sample items are provided

0 points if instructions are unclear, incomplete or nonexistent.

FIGURE 2 (continued)

4. Is the test formatted clearly? (2 points possible)

Test layout should be examined for effective use of perceptual organizers, such as adequate white space, regularity of item form, symmetry, clarity and continuity.

Award points on the following basis:

2 points if (1) test page layout is clear and helpful and (2) print and illustrations in printed tests and sound in auditory or taped tests are high quality

1 point if only (1) or (2) above apply

0 points if layout is unclear or confusing or quality of print or tapes is low.

5. Is there a simple way for examinees to record their responses? (2 points possible)

Award points on the following basis:

2 points if response is especially simple for examinee--e.g., oral responses, or marking or writing directly on test form

1 point if test uses standard separate answer sheets

0 points if test is complicated by the need for more than one step to get from item to answers.

6. Is the type of score reported useful for my situation?

Award points on the following basis:

2 points if the score reported is precisely the type needed

1 point if a usable score can be obtained even though it is not exactly what is desired

0 points if the desirable score is unattainable.

FIGURE 2 (continued)

7. Is the process of converting raw scores simple and does it yield scores which are easily interpreted? (2 points possible)

Award points on the following basis:

1 point if the scores are reported in reference to a clearly identified norm group, a level of competency on clearly identified skills or learning objectives or in terms of 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

0 points if converted scores are ambiguous or conversion is lacking for raw scores not meaningful in themselves.

And:

1 point if the score conversion procedure is simple, involving one easy-to-understand step—such as a clear chart or table--or no conversion is necessary because the raw scores are interpretable.

0 points if the process of achieving final scores is complicated by lack of clear or simple tables or graphs or if it requires two or more steps to get from the raw to the converted scores--e.g., using one table to get into another table.

8. Is the amount of time required for testing appropriate? (2 points possible)

Award points on the following basis:

2 points if the amount of time matches well with the available time

1 point if adjustments can be made by either using only part of the test or adjusting the testing time limits

0 points if it will be extremely difficult to accommodate the required testing time.

FIGURE 2 (continued)

III. Reliability--the extent to which the test measures a trait with consistency.

1. Is reported reliability for major subtests and/or total test scores sufficiently high? (3 points possible)

One or more of three different types of reliability may be reported--alternate form reliability, test-retest reliability, and internal consistency estimates.

Assign points on the following basis:

Reliability	Points Awarded
.90 or above	3
.75 to .89	2
.65 to .74	1
.64 or below	0

When more than one type of reliability is reported, use the most frequent rating across the different reliabilities.

2. Are the scoring procedures clear and complete, thus ensuring reliable scores? (2 points possible)

Award points on the following basis:

2 points if (1) scoring procedures are clear and complete and (2) scoring of objective items is done using a scoring guide, template stencil or other straightforward process, or machine scoring is available and scoring of subjective items is done using rigorous training and scoring guides.

1 point if only (1) or (2) above applies

0 points if neither (1) or (2) above applies.

Chapter IV

After Test Selection

CHAPTER IV

AFTER TEST SELECTION

Once a test is selected, one must determine acceptable levels of performance, and decide how to report results to ensure they are used for their intended purpose(s).

Setting Performance Levels

Keep in mind that setting performance levels is secondary to determining test content. If the content tested is inappropriate (e.g., using a standardized reading comprehension test for making decisions about job placement), then performance levels are of little importance.

Setting a level of accepta. performance on a test is separate from obtaining a test score. Whethe. scores are norm or criterion referenced, the level of satisfactory performance for a specific situation must be determined.

Setting standards of acceptable performance is always an arbitrary decision, but it should at least be an informed arbitrary decision. There is no magical answer concerning what level of performance is satisfactory; the decision must rest on the opinions of experienced, knowledgeable people who understand the nature of the skills or knowledge being tested and the general capabilities of the examinees in question.

There are two methods of setting performance levels: reviewing actual student performance, and reviewing test items.* Either approach calls for human judgment.

One method, which relies on inspection of the test, is described in Figure 3. In this method, judges are provided with a set of all the items on a test. Looking at each item, they determine how many wrong alternatives a minimally qualified examinee should be able to identify. The number of remaining alternatives is used to determine item difficulty. The score indicating acceptable performance is based on the average difficulty of all test items.

Other standard setting methods call for judges to review the scores of selected groups of students. Knowing how certain students performed, the judges then decide what level of performance is acceptable for a given purpose or situation.

---

*See Handbook for proficiency assessment: Section VI-passing scores. Berkeley, CA: Educational Testing Service, December 1979 for further information on the methods described on the next pages as well as additional methods.

The simplest method of this type is the Borderline Method. A group of students is identified as being so close to the borderline between mastery and nonmastery that a teacher or counselor cannot be certain if they need supplemental instruction or not. Approximately 100 such borderline students are then tested, and the median score attained by these students is selected as the passing score.

It is possible to combine these two standard setting methods. Figure 4 describes one way this might be done.

The previous example focused on minimal competency levels. It is possible to have more than one competency level: e.g., one level to distinguish superior from average performance, another to distinguish average from incompetent performance. The methods discussed earlier could be easily adapted for setting multiple levels.

In setting performance levels, the crucial question is: Whose opinions should guide what's termed "acceptable"? For example, suppose test results will determine whether a person is ready to undertake a job that requires reading a manual. Then the opinions of people who supervise that position may be particularly relevant. In some cases, groups or individuals may set quite different standards for acceptable performance. For example, suppose that incoming freshmen in all community colleges in a state are tested. Though all students take the same test, staff at each college may set their own standards for acceptability.

FIGURE 3

SETTING PASSING SCORES BASED ON TEST INSPECTION*

DESCRIPTION OF THE PROCEDURE:	Judges are provided with a set of all items in the test. Looking at each item, they decide for each question the number of wrong alternatives a minimally qualified person would be expected to identify as obviously wrong. The number of alternatives remaining constitutes the set from which the examinee is expected to guess. The passing score of the test is the average judged difficulty of all items.
NUMBER OF JUDGES	Between seven and 25 judges representing perspectives deemed important should be used.
ADVANTAGES OF THE METHOD:	1. The procedure is independent of the number of students taking the test. A passing score can be calculated for even very small groups of students. 2. The procedure can accommodate participation by a broad cross-section of judges (e.g., teachers, administrators, parents, students, employers). 3. The procedure is based on close scrutiny of test items. 4. The procedure closely parallels the decision processes of test takers; each alternative for each item is individually considered.
DISADVANTAGES OF THE METHOD:	1. The procedure is blind to actual performance on the test. 2. The passing score can be too high or too low when there are a disproportionate number of judges with the same interest or bias. 3. The procedure must be repeated for each form or test when different forms or tests are used that are not equated. 4. More time and people are needed to make judgments than with performance-based procedures. 5. The procedure can only be used with multiple choice tests.

*This method is known as the Nedelsky method. The information in this figure is taken from the Handbook for proficiency assessment: Section IV-Passing Scores. Berkeley, CA: Educational Testing Service, December, 1979.

FIGURE 3 (continued)

PROCEDURES FOR TRAINING JUDGES

1. Orally review the purpose of the tests and standard-setting exercise, including presentation and discussion of the purpose of the standard. Be sure to distinguish between average performance and minimally acceptable performance.

2. Distribute practice questions.

3. Explain to judges that this method requires a group of knowledgeable judges to inspect each question and make a judgment about each wrong alternative. Each judge must decide whether a hypothetical examinee who just barely meets the definition of a minimally acceptable performance for the situation under consideration could be expected to eliminate the wrong alternative.

4. Review the practice questions and as a group decide which alternatives could be eliminated by a minimally competent person. Complete a judge's recording form for the practice questions.

5. After the task is clear to the judges, pass out the test booklets and as a group decide which alternatives could be eliminated. For each wrong alternative where there is not unanimous agreement, ask one judge from each viewpoint to give a brief explanation of his/her reasons. The purpose of such an exchange is not to force consensus but to allow different points of view to be heard. Use a judge's recording form to allow each judge to indicate his/her decision of the number of alternatives which can be omitted.

6. Go through each question on the test in the way described above.

7. When all test items have been reviewed, have judges tally each column of circled numbers. Transfer the column totals to the corresponding blanks by the probabilities and multiply. The sum of those multiplications is that judge's passing score estimate. Average all of the judges' estimates to obtain the group's recommended passing score.

FIGURE 3 (continued)

Example Test Items:

Directions: Look carefully at the following lists of words. Each row has one word that is spelled wrong. Mark your answer sheet to show the word in each row that is spelled wrong.

1. a. house
   b. woman
   c. suny
   d. have

2. a. wash
   b. second
   c. zipper
   d. rownd

3. a. lesson
   b. appel
   c. toys
   d. because

Judge's Recording Form:

Test Item Number	Circle Number of Choices Eliminated			
1	0	1	2	3
2	0	1	2	3
3	0	1	2	3

Total 0's = ____ x .25 = ____

Total 1's = ____ x .33 = ____

Total 2's = ____ x .50 = ____

Total 3's = ____ x 1.00 = ____

____ Sum = ____

FIGURE 4

SETTING PASSING SCORES USING THE COMBINATION METHOD*

DESCRIPTION OF THE PROCEDURE: Judges are provided with a description of each skill covered in the test, a sample item measuring that skill taken from the test, performance data on the skill gathered from examinees and the projected number of examinees who would fail given several different passing scores. Each judge considers this information and decides what percentage of the items in that skill area a "proficient" person should answer correctly.

JUDGES: Either a small group of judges representing various perspectives can be used to set passing scores or a larger sample of the constituencies can participate in the decision making by responding to a mailed survey.

ADVANTAGES OF THE METHOD:

1. The combination of representative items and local performance levels provides judges with the most comprehensive set of information possible.
2. The procedure is independent of the number of people taking the test. The passing score can be calculated for even very small groups.
3. The procedure can accommodate participation by a broad cross-section of people (e.g., teachers, administrators, parents, students, employers).
4. Judges' decisionmaking time is reduced because only one judgment per subtest rather than one per item is required.
5. The procedure can be adapted easily for mailing as a survey to receive many views.

DISADVANTAGES OF THE METHOD:

1. Item specifications must narrowly define the range of content and difficulty of items used in the test; otherwise the sample item may be misleading to the judges.
2. Performance data must be recent and must adequately represent all examinees for whom the performance level is being set.
3. The passing score can be too high or too low when there are a number of judges with the same interest or bias. Returns from mailed surveys are particularly vulnerable to special interests.
4. More time and people are needed to make judgments than with performance-based methods. This method is particularly time consuming if a mailed survey is used.

---

*Taken from Handbook for Proficiency Assessment: Section IV-Passing Scores. Berkeley, CA: Educational Testing Service, December 1979.

FIGURE 4 (continued)

EXAMPLE OF INFORMATION JUDGES MIGHT USE IN SETTING PASSING SCORES USING THE COMBINATION METHOD

Subtest Content: SPELLING

Sample Item: (One of six items) Mark your answer sheet to show the word that is spelled wrong. a. house b. woman c. suny d. have	Expected Peformance on Subtest (Adult Basic Education Class): 85 percent correctly answer item.

Projected Failure/Success Rate of Adult Basic Education Students:

If passing score is set at:	The % of students who would probably pass:	The % of students who would probably fail:
100%	38%	62%
90%	71%	29%
80%	87%	13%
70%	92%	8%
60%	95%	5%

SAMPLE INSTRUCTIONS FOR SETTING PASSING SCORES ON LANGUAGE ARTS TEST (COMBINATION METHOD)

Directions: After considering the information provided for you, please indicate in the corresponding space on the right what you consider to be an appropriate minimum passing score for Adult Basic Education students for each of the areas listed below. Use percents to show the passing score you think is appropriate.

Skill Areas	Percent
A. Spelling (6 items)	_____
B. Subject/Verb Agreement (4 items)	_____
C. Ending Punctuation (4 items)	_____
D. Complete Sentences (4 items)	_____
E. Capitalization (6 items)	_____

Now go back and average the passing scores assigned to each skill area so that you have a passing score on the total Language Arts test.

Minimum Passing Score for Total Test _____

Interpreting Test Results

Prior to reporting, test scores must be carefully interpreted. Scores in themselves are virtually meaningless unless the context and purpose for testing are explained. In addition, the following factors should be considered in interpreting the results:

a. Curriculum characteristics (e.g., discrepancies between what is tested and what is taught)
b. Staff characteristics (e.g., the teaching or administrative structure)
c. Student characteristics (e.g., students' ability, interests)
d. Technical characteristics of the test (e.g., the test's reliability and validity)
e. Sampling characteristics (e.g., the relationship of those tested to a larger group, to which the user may generalize)
f. Social factors (e.g., the social meanings of the skills tested for the examinees)

A wide range of perspectives enhances interpretation. Teachers, curriculum specialists, administrators, employers, guidance counselors, measurement specialists and students may each be able to contribute uniquely to interpretation of the test results.

Reporting Test Results

Total reliance on written reports assumes that those awaiting test results have the interest and time to read and interpret what is written. This is often not the case. Alternative approaches should be considered.

The best reporting method depends on what needs to be communicated. A major distinction can be made between (1) one-way provision of information (e.g., written reports) and (2) two-way interaction (e.g., community meetings). Written or televised reports may sometimes be useful; in other situations, personal discussions about the results can be much more appropriate.

Chapter V

Basic Literacy Assessment

CHAPTER V

BASIC LITERACY ASSESSMENT*

The term basic literacy skills is used here to mean those skills a person needs in order to learn and communicate via reading or writing. Though curriculum specialists have probably been most active in delineating these skills, persons from the fields of psychology, history, linguistics and library science have all contributed to their identification (Bormuth, 1975, p. 70).

Although exhaustive listings and definitions are not appropriate here, reading and writing taxonomies are presented to give the reader a general impression of the range of skills to consider when conducting basic literacy assessment.

Reading Skills**

1. Decoding skills: Decoding skills enable a person to recognize letters, letter groups and patterns in print and their associated oral sounds and meanings. Phonetic analysis (associating sounds with letters) and structural analysis (associating syllables, affixes or whole words with their corresponding sounds) are usually considered decoding skills.
2. Literal comprehension: Literal comprehension skills enable a person to understand information explicitly stated in a text. Included are vocabulary skills (assigning the correct meaning to words in context) and ability to combine the meanings of words in sentences, or sentences in paragraphs.
3. Inference skills: Inference skills allow a person to derive information not explicitly stated, i.e., to "read between the lines."
4. Critical reading skills: A person with critical reading skills can thoughtfully examine a text for logical consistency, as well as detect and evaluate propaganda techniques.
5. Aesthetic appreciation skills: These skills enable a person to evaluate the tone or mood of a story or the rhythm of prose.
6. Reading flexibility skills: These skills allow a person to read faster or slower, depending on the nature of the task, to focus selectively on parts of the text, and to switch attention to conform to a wide variety of instructions.
7. Study skills: These skills enable a person to use various reference materials (e.g., maps, graphs, charts, tables of contents and diagrams) to locate information and judge its relevance to a particular task.

---

*Chapters VI and VII focus on everyday literacy skills and on-the-job literacy skills, respectively.

**This reading taxonomy is from Bormuth (1975). Other taxonomies include that provided by Stiggins (1981), in which seven component skills of reading are presented. Stiggins' taxonomy assumes that reading is an interactive process in which the characteristics of the written text interact with the reader's knowledge.

Writing Skills*

1. Ideas--the quality, development, support and relevance of the arguments, opinions and thoughts expressed.
2. Mechanics--factors such as usage, sentence structure, punctuation and spelling.
3. Organization--a sense of order, ability to stay on topic and relate details to a central idea or argument.
4. Wording and phrasing--the choice and arrangement of words, including the deletion of unnecessary words.
5. Flavor--the personal qualities revealed by the writing--style, individuality, originality, interest and sincerity.

One could assess the kinds of reading or writing skills noted here using available tests for several of the testing purposes mentioned in Chapter II. The most likely purposes would be diagnosing specific needs for further instruction, placing a student in the appropriate level of an English or writing course, guiding a student into a general program of study, and conducting a survey assessment of literacy skills among a broad group of people.

In several situations, existing basic literacy skills tests are not likely to be useful. Most selection or certification decisions affecting adults are an example. Such decisions would be less likely to rest on basic literacy skills than on job-related or everyday life skills. Similarly, most formative or summative evaluation purposes would be best served by a test designed specifically to measure skills taught in the program.

Available Tests of Basic Literacy Skills--Reading

Reading tests are among the most common published tests. Most have been developed for elementary school students but a significant number exist for junior and senior high school students and adults. In recent years, an increasing number of criterion referenced tests of basic skills have become available. Although many of these tests are designed for elementary and secondary school students, they can also be used with adults who are acquiring basic skills. It is important to ensure at the onset that the test measures the desired skills, and that the test items do not appear childish to an adult. Most tests designed for junior high and higher are adaptable for use with adults.

---

*This writing taxonomy is based on the work of Diedrich. In an effort to identify the characteristics that most influence a writing expert's judgment of the quality of a piece of writing Diederich (1974) analyzed the results of ratings of a large number of student essays. He isolated the factors given in this taxonomy. This list is actually qualities of writing. The ability to achieve each of these factors is the skill to be assessed.

Available tests often provide separate subtest scores. The most common subtests on reading tests are decoding, vocabulary, comprehension and study skills. Published tests measuring oral reading and reading flexibility skills are less common, but some do exist.

Appendix C provides a fairly comprehensive list of available reading tests published in the last ten years. The reader should view this list as a starting point. For further detail on these and other tests, review the reference materials mentioned in Chapter II.

Available Tests of Basic Literacy Skills--Writing

Two general methods of assessing writing proficiency are used: direct and indirect. The direct approach involves gathering samples of writing and evaluating them according to specified criteria. This method simulates real-life writing circumstances and requires trained judges to apply the criteria. The indirect approach involves the use of objective tests (usually paper-and-pencil multiple choice tests) to measure language usage skills important in effective writing. This approach is often less costly because tests can generally be machine scored.

Although the results of indirect and direct assessments are generally highly related, indirect assessment should not be considered a substitute for direct. Each measures different skills. Direct measures focus on writing composition skills while indirect measures focus on prerequisites of good writing. The following lists of skills measured by the two methods were distilled from 18 statewide writing assessments being conducted in the United States:*

DIRECT ASSESSMENT	INDIRECT ASSESSMENT
Usage	Punctuation
Sentence sense	Grammar
Expression of feeling	Placing modifiers
Persuasiveness	Determining idea relevance
Organization	Diction
Format	Style
Cohesiveness	Transition
Revision skills	Logic
Transition	Organization
Overall writing proficiency	Overall usage proficiency
	Sentence structure

*This analysis is taken from Bridgeford, N., & R. Stiggins. A Consumer's Guide to Writing Assessment. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Each method has advantages and disadvantages. The major advantages of direct assessment are the extent of information gathered (i.e., when one has a chance to examine an actual sample of the examinee's writing); the flexibility or adaptability of the exercise to a variety of relevant real world writing circumstances; and users' positive response thanks to the high face validity of writing samples. The major advantages associated with indirect assessment, on the other hand, are high score reliability, relatively low test scoring costs and a high degree of control over the nature of the skills tested.

The potential disadvantages of direct assessment include the high cost of scoring and the potential lack of control over the skills tested (i.e., because every response is unique). The potential limitations of the indirect method are the lack of correspondence to real world writing tasks and heavy reliance of the assessment method on reading skills rather than writing proficiency.

There are numerous indirect measures of writing skill available. Since most tend to focus primarily on mechanics, the greatest choice is available in this area.

There are fewer direct measures available, but extensive work is underway to develop more, and to train people in the effective scoring of writing samples. A brief description of the basic approaches to scoring writing samples is given in Appendix D along with references to three excellent publications on the topic.

Appendix E provides a comprehensive list of available published tests of writing skills. Of the 47 tests listed, only nine have a component which involves a direct measure of writing.

Martha: A Case Study in the Selection of Available Tests of Basic Literacy Skills*

LITERACY SKILLS OF INTEREST

In January, Martha was hired as a community college counselor in a suburban city in a Northwestern state. Her responsibility was to help incoming students determine what courses they should take. Knowing that nearly all courses require reading and writing, she decided to make tests of basic literacy skills part of the entering assessment process. Martha was told that in the past reading skills had been measured with a norm referenced test. Scores were reported to students and teachers as grade equivalents. Writing skills had not been tested. To be sure that she was testing the most important literacy skills, Martha asked members of each department in the college to provide her with representative examples of the reading materials used in each course, and a list of required writing tasks.

After analyzing that information (with assistance from a reading and writing specialist), she decided to measure student reading skills in the areas of literal comprehension, inference and study skills. She also decided that whatever test she chose should rely on materials commonly found in school texts. She also decided to measure students' writing mechanics skills and their ability to write a short, well organized, and informative essay.

Although other skills were needed in some courses, Martha decided that it should be left to the instructors to measure those skills. She would be available to help them find appropriate tests.

PURPOSE FOR TESTING

Martha's purpose for testing was to identify students who would benefit from tutorial assistance, as well as those who should enroll in a special reading or writing class.

---

*The case studies in this and subsequent chapters reflect the important testing decisions outlined in Chapters II-IV.

USES AND USERS OF TEST RESULTS

Martha decided to report test scores to students and their instructors, as well as use the results herself to counsel students on how their reading or writing skills might affect their performance in various classes. She decided a criterion referenced score would be the most useful. However, to interpret scores properly she would need information on the performance of students in typical classes throughout the college.

EXAMINEE CHARACTERISTICS

In reviewing college records, Martha learned that generally about 20 percent of the students were of Spanish-American descent--and half of those were bilingual--that 10 percent were black, 5 percent from other cultural minority groups (many recent Indochinese refugees with little or no English skills), and the remainder were white.

Students whose native language was not English could indicate upon registration that they spoke, read or wrote little or no English. Martha decided these students should not be tested using the approach she was planning, since their unfamiliarity with the language would make the results useless. Instead, such students would be assessed informally by teachers in English as a Second Language (ESL) classes, using tests developed by ESL teachers. Such tests would be based on current instructional materials and information gained from seminars on the cultural characteristics of those for whom English was a second language.

LOGISTICAL CONSIDERATIONS

Martha had been told by her supervisor that she could test for no longer than two hours. Students would be tested in small groups throughout registration week and the first week of classes. They would sign up for the testing time most convenient for them.

To simplify administration, Martha wanted one reading test which will cover all three reading skills categories: literal comprehension, inference, and study skills.

The community college had access to test scoring services so she could have the tests machine scored.

TEST SELECTION PROCESS

Since Martha had a limited budget and short timeline, she considered only existing tests. If she could not find what she needed, she would have to postpone measurement until the following year.

In beginning her search for the right instrument, Martha decided to consider tests designed for junior and senior high school students, since there were few criterion referenced tests specifically designed for adults. Also, she knew from talking to the previous counselor that many students who had difficulty in classes had skills comparable to those of the typical eighth or ninth grader.

Knowing the kinds of reading and writing skills she wanted to measure, Martha went through the lists of tests in Appendices B and D. She looked up each test in the Mental Measurements Yearbook, Tests in Print, News on Tests, or the Consumer's Guide to Writing Assessment to learn as much as possible about each one.

Martha found four tests which appeared to be likely candidates for measuring reading skills. These were (1) Individualized Criterion Referenced Testing: Reading (ICRTR), (2) Mastery: An Evaluation Tool: Reading (SOBAR), (3) Criterion-Referenced: Reading Tactics, and (4) Analysis of Skills (ASK): Reading.

After reading the Consumer's Guide to Writing Assessment, Martha decided the best approach to measuring students' writing ability was to use a direct measure of writing (i.e., have students compose a writing sample following carefully designed instructions). Since she also wanted to measure writing mechanics, she decided to first review published tests which included writing samples as well as multiple choice items focused on writing mechanics.

Martha decided to review the Basic Skills Assessment Program: Writer's Skills Test, IOX Basic Skills Tests--Secondary Level (Writing), WRITE: Senior High and Writing Proficiency Program for possible use in measuring writing.

Martha reviewed the reading and writing tests using the review procedure specified in Chapter II. Her ratings are shown in Figure 5. After weighing the pros and cons of each she decided to use Mastery: An Evaluation Tool: Reading (SOBAR) and the Basic Skills Assessment Program: Writer's Skills Test. She selected the SOBAR because it included a wide variety of study skills items in addition to appropriate reading items. It could also be customized to match her objectives and time limitations, and included an optional section on reading in content areas.

FIGURE 5
TECHNICAL REVIEW OF BASIC READING SKILLS TESTS
(AN EXAMPLE)

Rate each test of interest using the rating system described in the instructions in Figure 2.

Validity

1. Do the test items measure the specific learning objectives or skills that need to be assessed?

   List learning objectives or skills to be assessed in the left hand column of the chart below. Put the name of each test being reviewed in a numbered box at the top of one of the right hand columns. (Repeat on each page of the test review chart.) Under each test indicate the number of items measuring each objective. Review the test manual or the actual test to determine if each objective or skill is measured. The publisher may indicate in the manual what skill each test item measures. If not, examine each test item to determine what skill or objective it measures.

Learning objectives or skills to be assessed	Test Name 1 ICRTR	2 SOBAR	3 CR-RT	4 ASK
1. Literal Comprehension	7 obj 14 items	12 obj.* 36 items	9 obj 6.5 items/obj.	4 obj 12 items
2. Inference	14 obj 28 items	10 obj.* 30 items	11 obj. 6.8 items/obj.	9 obj. 27 items
3. Study Skills	none	27 obj.* 81 items	11 obj. (distributed)	12 obj. 36 items
4.			13 obj. reading rate 6.8 items/obj	
5.				
6.				

For the remainder of the questions, assign points as described in the Instructions for Assigning Ratings in the Technical Review of Tests Form (Figure 2). The number in parentheses after each question indicates the number of possible points.

* Total number of objectives available from which to design customized test.

Figure 5 (continued)

	Test Name			
	1 ICRTR	2 SOBAR	3 CR-RT	4 ASK
2. Were the test items developed in a systematic and rigorous manner so that the content is adequate and bias is minimized? (6)	2	2	NR*	2
3. Were any empirical procedures used for screening or selecting items to ensure that items are measuring what they were designed to measure, are understandable, contain reasonable answers, and are free of ambiguous alternative answers and unnecessarily complex language? (3)	0	1	NR	2
4. Is the validating group representative of the population with which the test is to be used? (2)	1	0	NR	NR
5. Are any special validity studies reported or specifically referenced? (2)	1	0	NR	NR

Usability

1. Are the test items suitable for adults with limited literacy skills? (2)	2	2	2	2
2. Are instructions to the test administrator clear and complete? (2)	1	1	2	1

* Not reported in materials available to reviewer.

Figure 5 (continued)

	Test 1	2	3	4
3. Are instructions to the examinee written in clear and understandable terms? (2)	2	1	2	2
4. Is the test formatted clearly? (2)	2	2	2	1
5. Is there a simple way for examinees to record their responses? (2)	1	1	2	1
6. Is the type of score reported useful for my situation? (2)	2	2	2	2
7. Is the process of converting raw scores simple and does it yield scores which are easily interpreted? (2)	1	2	2	2
8. Is the amount of time required for testing appropriate? (2)	2	*	2	1

Reliability

1. Is reported reliability for major subtests and/or total test scores sufficiently high? (3)	2	NR	NR	NR
2. Are the scoring procedures clear and complete, thus ensuring reliable scores? (2)	2	1		1

* Time varies according to number of objectives included.

FIGURE 6
TECHNICAL REVIEW OF WRITING SKILLS TESTS
(AN EXAMPLE)

Rate each test of interest using the rating system described in the instructions in Figure 2.

Validity

1. **Do the test items measure the specific learning objectives or skills that need to be assessed?**

   List learning objectives or skills to be assessed in the left hand column of the chart below. Put the name of each test being reviewed in a numbered box at the top of one of the right hand columns. (Repeat on each page of the test review chart.) Under each test indicate the number of items measuring each objective. Review the test manual or the actual test to determine if each objective or skill is measured. The publisher may indicate in the manual what skill each test item measures. If not, examine each test item to determine what skill or objective it measures.

	Test Name			
Learning objectives or skills to be assessed	1 Writer's Skills Test	2 IOx Basic Skills Test	3 WRITE	4 Writing Proficiency Program
1. Write a well organized essay	2	4	2	2
2. Organize ideas logically	27		4	8
3. Capitalization	5	7	3	44
4. Spelling	14	7	4	12
5. Using words correctly (verbs, pronouns)	16	10	16	5
6. Punctuation	9	6	7	36
7. (other objectives)	4	10	15	24

For the remainder of the questions, assign points as described in the Instructions for Assigning Ratings in the Technical Review of Tests Form (Figure 2). The number in parentheses after each question indicates the number of possible points.

Figure 6 (continued)

	Test Name			
	1 WST	2 IOX	3 WRITE	4 WPP
2. Were the test items developed in a systematic and rigorous manner so that the content is adequate and bias is minimized? (6)	5	3	4	3
3. Were any empirical procedures used for screening or selecting items to ensure that items are measuring what they were designed to measure, are understandable, contain reasonable answers, and are free of ambiguous alternative answers and unnecessarily complex language? (3)	3	NR	NR	NR
4. Is the validating group representative of the population with which the test is to be used? (2)	2	NR	NR	NR
5. Are any special validity studies reported or specifically referenced? (2)	1	0	NR	NR

Usability

1. Are the test items suitable for adults with limited literacy skills? (2)	2	1	2	2
2. Are instructions to the test administrator clear and complete? (2)	2	0	2	2

Figure 6 (continued)

	Test			
	1 WST	2 IOX	3 WRITE	4 WPP
3. Are instructions to the examinee written in clear and understandable terms? (2)	2	1	2	2
4. Is the test formatted clearly? (2)	2	2	2	1
5. Is there a simple way for examinees to record their responses? (2)	2	2	2	2
6. Is the type of score reported useful for my situation? (2)	2	1	2	2
7. Is the process of converting raw scores simple and does it yield scores which are easily interpreted? (2)	2	NR*	NR	NR
8. Is the amount of time required for testing appropriate? (2)	2	NR	1	2

Reliability

1. Is reported reliability for major subtests and/or total test scores sufficiently high? (3)	3	NR	NR	NR
2. Are the scoring procedures clear and complete, thus ensuring reliable scores? (2)	2	NR	2	2

* Not reported in materials available to reviewers.

She selected the Writer's Skills Test because it had a relatively short administration time while maintaining a high level of reliability. She had the most complete information on this test and it closely matched her criteria.

SETTING PERFORMANCE LEVELS

Martha decided to use the Borderline Method (see p. 23) to set a level of performance that would indicate whether a student should have special reading or writing assistance. She asked teachers in all the freshman classes to identify students whom they considered "borderline" (i.e., possibly inadequate) in their reading and writing skills. These students would be tested in the spring, and the results used to set a standard Martha could refer to in counseling future students on the advisability of receiving extra help.

INTERPRETING AND REPORTING RESULTS

Once Martha receives the test scores, she will determine which students fall below the cutoff. She will meet with each student who scored below the cutoff to discuss possible problems. In many cases she may recommend special help in reading or writing.

Martha will make the test results available to other students upon request, and will confer with either students or teachers regarding test results.

Chapter VI

Assessment of Everyday Literacy Activities

CHAPTER VI

ASSESSMENT OF EVERYDAY LITERACY ACTIVITIES

There is growing concern that many adults lack the literacy skills to perform everyday tasks--such as reading road signs, maps, recipes and simple instructions. Whether the problem is actually due to inadequate reading skills or other conceptual limits remains a matter of debate. In any case, there are increasing numbers of tests that measure everyday skills involving reading and writing, and instructors of adults are frequently interested in assessing their students' abilities to perform these everyday tasks.

Not all such tests demand the same skills. For example, many everyday activities, such as filling out a job application, require that a person both read and write. Depending on what skills one wishes to measure, a test which requires the student to actually fill out an application may be preferable to one in which the student simply selects the correct answer from among several alternatives. Or, if one does not want to test both reading and writing on the same test item, a multiple choice test may be preferable.

Published tests of everyday activities best fulfill two testing purposes: diagnosis of students' skills and survey assessment (see pp. 4-5). Summative and formative evaluation and certification of minimal skills are also possible if the test is designed for a population very similar to that with which it will be used, and if the skills tested are clearly those being taught or those to be certified. Tests of everyday activities are not generally appropriate for use in selection.

Available Tests of Everyday Literacy Activities

Fourteen published tests of everyday literacy activities are listed in Appendix F. This list provides a starting point for locating useful tests of this type.

Nathan: A Case Study in the Selection of Available Tests of Everyday Literacy Activities

LITERACY SKILLS OF INTEREST

Nathan was a reading specialist in the Spring Park Community College Adult Basic Education program. One class there was designed specifically to help people acquire literacy skills essential for everyday life tasks (e.g., reading road signs, completing job applications, reading want ads, following simple directions).

PURPOSE FOR TESTING/ USES AND USERS OF TEST RESULTS	Nathan wanted to test students upon registration to obtain a general idea of what tasks they had difficulty with. He will use the results in conjunction with other information to plan instructional activities for each student. Nathan also wanted to administer the test again at the end of the year as part of a summative evaluation to determine how well skills were attained. The results would be used solely by Nathan in planning course changes for next year.
EXAMINEE CHARACTERISTICS	Nathan had taught this course before and was quite familiar with the type of students who signed up for the class. Many students were native English speakers who, for one reason or another, had not acquired many literacy skills during their previous schooling. They had problems with certain everyday tasks requiring literary skills. For another group of students, English was a second language. Some of these students had lived in the United States for a considerable time and were reasonably familiar with the predominant culture. Others were recent immigrants (including many Indochinese refugees) who had taken ESL classes for about one year. They now wanted to focus on literacy skills in everyday life.
LOGISTICAL CONSIDERATIONS	Nathan wanted a test which would take no longer than one hour to administer. Since he expected to have less than 40 students, he planned to hand score the test. He decided to divide the students into two groups for administration. While one group was taking the test, proctored by an assistant, he would interview the others to obtain additional information helpful in planning curriculum.
TEST SELECTION PROCESS	Nathan reviewed the list of tests presented in Appendix F. He found some of them described and reviewed in Tests of Adult Functional Literacy: A Review of Currently Available Instruments and further information on others in News on Tests. Based on this information, he identified the Adult Performance Level Functional Literacy Test, Senior High Assessment of Reading Performance (SHARP), and Reading/Everyday Activities in Life (R/EAL) tests as most appropriate for his needs. Next, he obtained copies of the tests from the publishers. He

looked at each item carefully to see whether it related to what he usually taught in the course, and whether the item content seemed appropriate for his students. He completed the technical quality rating form as shown in Figure 7.

None of the tests matched as well with his course content as Nathan would have liked. He felt that situations described in each test would be appropriate for his native English speaking students and most of the bilingual students who had been in the United States for several years. But the more recent immigrants, he suspected, would find some of the situations described in test items so unfamiliar that their responses would not reflect their true literacy skills. Nathan concluded that while a test would be useful, he would have to take these factors into account in using the results.

Nathan decided to use the R/EAL test because it matched his course content better than the others. It also required students to write answers; the other tests used a multiple choice format. Nathan thought written answers would be helpful in diagnosing student problems.

SETTING PERFORMANCE LEVELS

Once Nathan had selected the test, he thought more about the variety of meanings literacy had for his students. He realized even more than before that it would be inappropriate to use a particular performance level for making unilateral decisions. His students simply differed too much in background and experience.

INTERPRETATION AND REPORTING OF RESULTS

Nathan decided that the test would be very adequate for one of his purposes—initial placement and instructional planning for individual students. However, he modified his original plan to use the test for evaluation. The test did not reflect either the curriculum or his students' characteristics accurately enough to justify use in summative evaluation.

Nathan decided to meet with a test development specialist to find out what information to collect during the year so that they could develop a test suitable for course evaluation.

Once tests were scored, Nathan planned to meet with each student to discuss the kinds of everyday tasks on which the student had difficulty. This review would be only a first step toward learning more about what each student viewed as his/her difficulties, and what further learning was desired.

FIGURE 7
TECHNICAL REVIEW OF EVERYDAY LITERACY ACTIVITIES TESTS
(AN EXAMPLE)

Rate each test of interest using the rating system described in the instructions in Figure 2.

Validity

1. **Do the test items measure the specific learning objectives or skills that need to be assessed?**

   List learning objectives or skills to be assessed in the left hand column of the chart below. Put the name of each test being reviewed in a numbered box at the top of one of the right hand columns. (Repeat on each page of the test review chart.) Under each test indicate the number of items measuring each objective. Review the test manual or the actual test to determine if each objective or skill is measured. The publisher may indicate in the manual what skill each test item measures. If not, examine each test item to determine what skill or objective it measures.

	Test Name			
Learning objectives or skills to be assessed	1 APL	2 SHARP	3 REAL	4
1. Road Signs	—	4	5	
2. Time Schedule	1	4	5	
3. Do It Yourself Instructions	—	4	5	
4. Telephone Directory	—	4	—	
5. Magazine Articles	—	4	5	
6. Advertisements	—	—	5	
7. Lease	1	—	5	
8. Road Map	1	8	5	
9. Want Ad	1	4	5	
10. Job Application	1	4	5	
11. Other	76	35	—	

For the remainder of the questions assign points as described in the Instructions for Assigning Ratings in the Technical Review of Tests Form (Figure 2). The number in parentheses after each question indicates the number of possible points.

Figure 7 (continued)

	Test Name			
	1 APL	2 SHARP	3 REAL	4
2. Were the test items developed in a systematic and rigorous manner so that the content is adequate and bias is minimized? (6)	2	NR*	5	
3. Were any empirical procedures used for screening or selecting items to ensure that items are measuring what they were designed to measure, are understandable, contain reasonable answers, and are free of ambiguous alternative answers and unnecessarily complex language? (3)	2	NR	2	
4. Is the validating group representative of the population with which the test is to be used? (2)	2	NR	NR	
5. Are any special validity studies reported or specifically referenced? (2)	NR	NR		

Usability

1. Are the test items suitable for adults with limited literacy skills? (2)	2	2	1	
2. Are instructions to the test administrator clear and complete? (2)	1	2	2	

* Not reported in materials available to reviewer.

Figure 7 (continued)

	Test 1 APL	2 SHARP	3 R/EAL	4
3. Are instructions to the examinee written in clear and understandable terms? (2)	2	2	2	
4. Is the test formatted clearly? (2)	2	2	2	
5. Is there a simple way for examinees to record their responses? (2)	1	1	2	
6. Is the type of score reported useful for my situation? (2)	2	2	2	
7. Is the process of converting raw scores simple and does it yield scores which are easily interpreted? (2)	2	2	2	
8. Is the amount of time required for testing appropriate? (2)	1	1	1	

Reliability

1. Is reported reliability for major subtests and/or total test scores sufficiently high? (3)	1	NR	2	
2. Are the scoring procedures clear and complete, thus ensuring reliable scores? (2)	2	2	1	

Chapter VII

Assessment of On-the-Job Literacy

CHAPTER VII

ASSESSMENT OF ON-THE-JOB LITERACY

While literacy skills alone rarely determine job success, many jobs demand skill in reading and writing. Thus, assessment of such skills is useful in countless personnel decisions, including those involving staff development and job placement.

When Mikulecky (1980) studied documentation of work related literacy competencies, however, he found that fewer and fewer businesses were using or making available tests designed for employees. The possible reasons are numerous. Schultz (1975) and Hunt and Lindley (1977) found, for instance, that pre-employment tests are often extremely difficult to read and not commensurate with the reading levels of particular jobs.

A primary concern when assessing job competency is the legal acceptability of tests. In Griggs vs. Duke Power Company (U.S. Supreme Court, 1971, 3 FEP Cases 175), the court ruled that tests used to measure job competency must be validated as job related. Showing this relationship can be a complex and time consuming task. Job analysis is a critical first step in test selection or development.

One of the most enlightening studies of on-the-job literacy was done by Sticht (1975)*. For nearly a decade Sticht analyzed the role of literacy in selected military occupations. He concluded that there are two fundamental purposes for reading at work: reading-for-learning and reading-for-doing. A reading-to-do task involves looking up information necessary to accomplish a certain task--information which can then be forgotten. Following a recipe, locating parts in a supply catalog, completing an inventory checklist and studying a computer reference manual are examples of reading-to-do. Reading-for-learning requires understanding and assimilating information for later use.

Reading-to-do requires searching for needed information. This may involve using tables of contents and reference indexes. While memorization is not necessary, some may occur naturally through repetition and association. Knowledge of how and where to locate information efficiently must, of course, be acquired and retained.

Reading-for-learning predominates in the school setting and in many job training programs, especially for positions above entry level. It is appropriate when procedures must be learned because reference documents are not conveniently available (e.g., at a construction site) or because time requirements demand that complex tasks be completed quickly. Reading-for-learning also occurs when a job requires high order mental operations, such as evaluation and information synthesis to solve complex problems. For example, a law clerk may need to read several past court cases and evaluate and synthesize the information to determine implications for a current case. Or a computer programming supervisor may need to read a variety of articles to glean and synthesize information helpful in analyzing the reasons for certain recurrent programming problems.

---

*Other studies have been done by Thomas and Smotherman (1976) and Moe, Rush and Storlie (1979).

Stiggins (1981) analyzed the differences between reading-to-learn and reading-to-do in terms which have implications for testing. He states that reading-to-do requires identifying and constructing word meanings automatically, taking advantage of explicit structure provided by the author to quickly assess needed information. The required vocabulary is defined by the specific job context. The memory demands are not high since the material used is generally familiar, provided in an easily used form and readily available for reference. The information read does not have to be integrated with existing knowledge but does have t compared with existing knowledge to ensure accuracy.

On the other hand, Stiggins maintains that reading-to-learn, though also requiring the reader to automatically identify or construct word meanings, differs from reading-to-do. Materials designed for reading-to-learn are usually presented in narrative form with an implicit, not an explicit, structure. The reader must impose a summary structure on the material that facilitates integration of the material with existing knowledge. Effective learning is contingent on reading from the proper perspective. The memory demands are higher since the material learned is generally unfamiliar and must be compared with existing knowledge. In reading-to-learn, the reader relies predominantly on careful processing of written material, as opposed to rapid scanning. Information must be evaluated in terms of accuracy and appropriateness and assimilated into one's existing knowledge structure. In short, reading-to-learn requires more complex information processing.

The United States Army is currently conducting task analyses of over 120 basic army jobs to determine their literacy requirements (Mikulecky, 1980). Competencies and patterns of competencies across jobs are being determined to facilitate development of training programs. This approach assumes that seemingly similar competencies can be taught for several different jobs at the same time. Scribner and Cole s (1979) work suggests that this assumption may not be wholly accurate. What makes a reader competent may not be mastery of a few basic skills easily transferable to different settings, but rather, mastery of a wide variety of specific reading skills gained through numerous diverse experiences.

Moe, Rush and Storlie (1979) have investigated the literacy requirements of ten semi-skilled and skilled occupations and the corresponding requirements necessary to succeed in training programs for each of those occupations. These job specific literacy requirements are described to aid educators, counselors and administrators in providing services to adults who aspire to these occupations but have minimal literacy skills. Recommendations for instructional programs for each of the occupations are included. The following occupations were studied:

Account Clerk
Automotive Mechanic
Draftsman
Electrician
Heating and Air Conditioning Mechanic
Industrial Maintenance Mechanic
Licensed Practical Nurse
Machine Tool Operator
Secretary
Welder

Available Tests of On-the-Job Literacy

Available tests for on-the-job literacy are limited primarily to clerical positions (Mikulechy, 1980). Tests for other positions often measure factors other than literacy. Appendix G lists available clerical tests which may be useful in certain situations. These tests may be adequate for diagnosing students' skills, placing them in appropriate training programs and in some cases, for certification. In certification, careful task analysis and establishment of appropriate criterion competence levels required for each job are critical.

Phyllis: A Case Study in the Selection of Available Tests of On-the-Job Literacy Skills

LITERACY SKILLS OF INTEREST

Phyllis was a personnel officer for Aloha Computer Company, a large computer assembly industry. She was responsible for ensuring that new employees had adequate skills to perform the tasks required in their respective positions. Supervisors had been particularly concerned recently about the literacy skills of employees, and many had asked Phyllis to start routinely testing applicants' reading and writing skills.

Phyllis was concerned about both the potential legal problems associated with such testing and the usability of the results. She had read how important it was to be sure any reading and writing skills measured were clearly identified as those necessary for a given job.

Phyllis decided to begin by identifying tests developed for specific jobs. After reviewing the organizational positions and talking with a local university measurement professor, she decided that the clerical positions were the only ones for which existing tests might be useful.

Phyllis had job descriptions for all clerical positions on file. She also had samples of the types of letters and memos which clerical staff were asked to write, and samples of the manuals and other materials they had to read. For other areas she would initiate a task analysis and begin working with an assessment specialist to develop measures specifically designed for their situation. She thought it would take at least a year before any of these measures were ready for use in selection.

PURPOSES OF TESTING/ USES AND USERS OF TEST RESULTS	Phyllis wanted to use test results as part of the criterion for hiring. She also wanted to use results in determining current staff development needs. In addition to Phyllis, personnel selection and staff development committee members would have access to results.
EXAMINEE CHARACTERISTICS	Phyllis had worked at the Aloha Computer Company for seven years and was quite familiar with the type of applicants Aloha usually received for clerical positions, and the characteristics of successful and unsuccessful employees in these positions. Most applicants were recent high school graduates, though lately more women who had been housewives for 10 to 20 years were applying. The community was about 30 percent Black, 5 percent Spanish and 5 percent Asian and other minorities. Phyllis wanted to be sure the test did not unfairly discriminate among applicants based on race. Nearly all applicants were accustomed to taking standardized tests.
LOGISTICAL CONSIDERATIONS	Each applicant would be given the test when they applied for a position. It was important that the test take no longer than 30 minutes to administer.
TEST SELECTION	Phyllis reviewed the list of tests in Appendix G. She called each publisher and asked for sample copies. After reviewing several tests using the procedures given in Chapter II, she decided to use the Short Tests of Clerical Ability.
SETTING PERFORMANCE LEVELS	Phyllis planned to administer the test to a randomly selected sample of present employees, then use the results in establishing a criterion level of performance. She planned to use the combination method (p. 27) for setting the performance standard. She would carefully select judges from various positions (supervisors, office managers, clerical staff) to participate in standard setting. She would also share results from this test with the staff development committee.
INTERPRETING AND REPORTING RESULTS	Phyllis decided to report results to the selection committee using a graph which showed the criterion score and each examinee's score (see Figure 8). She would emphasize the importance of interpreting scores in light of other information obtained via interviews and references.

Figure 8
Short Tests of Clerical Ability Results

Applicant	0	10	20	30*	40	50	60	70
Andrews, P.					X			
Barclay, R.						X		
Davis, J.			X					
Levine, T.				X				
Lopez, B.							X	

The two-to-five-page report to the staff development committee would not give information on specific examinees. Rather it would discuss which skills appeared most deficient. Important concerns raised during the standard setting sessions would also be reported.

---

*Criterion score

Summary

This Guide has attempted to introduce the reader to the key issues in selecting existing tests to measure adult literacy. Extensive research and development is still needed to provide educators with adequate adult literacy measures.

The staff of the Functional Literacy Project welcome your comments on the utility of this Guide. Suggested additions or changes to be made when the Guide is revised should be directed to

Assessment and Measurement Program Director
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

References

Anderson, B., R. Stiggins, & S. Hiscox. Guidelines for Selecting Basic Skills and Life Skills Tests. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Bormuth, J. Toward a literate society (Eds. J. B. Carroll & J. S. Challs). New York: McGraw Hill Book Co., 1975.

Bridgeford N., & R. Stiggins. A Consumer's Guide to Writing Assessment. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Buros, O. K. (Ed.) Tests in Print II. Highland Park, NJ: The Gryphon Press, 1974.

Buros, O. K. (Ed.) The Eighth Mental Measurements Yearbook. Highland Park, NJ: The Gryphon Press, 1978.

Hunt, T., & C. Lindley. Documentation of selection and promotion test questions: Are your records sagging? Public Personnel Management, October-November, 1977, 6.

Mikulecky, L. Literacy competencies and youth employment, a paper prepared for the National Institute of Education and the Department of Labor, 1980.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of account clerk on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of draftsman on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of electrician on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of secretary on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of heating and air conditioning mechanic on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of welder on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Moe, A., R. Rush, & R. Storlie. The literacy requirements of licensed practical nurse on the job and in a vocational training program (Project Report). West Lafayette, IN: Department of Education, Purdue University, November 1979.

Nafziger, D., et.al. Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments. Portland, OR: Northwest Regional Educational Laboratory, 1975.

Schultz, C. An unnecessary chore for employment applicants. Public Personnel Management, 1975, 4.

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

Stiggins, R. An Analysis of the Dimensions and Testing of Job Related Reading. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Thomas, J., & E. Smotherman. Catalog of performance objectives, criterion referenced measures and performance guides for carpenters. Vocational/Technical Consortium of the States, State Department of Education and Department of Vocational Education, University of Kentucky, 1976.

Appendices

APPENDIX A

SUGGESTED READINGS ON DEFINITIONS OF FUNCTIONAL LITERACY*

Bormuth, J. R. Reading literacy: Its definition and assessment. Reading Research Quarterly, 1973-74, 9(1), 7-66. A revised and abridged version was subsequently published in J. B. Carroll & J. S. Chall (Eds.), Toward a literate society. New York: McGraw Hill Book Co., 1975.

This monograph presents a thorough discussion of the concept of literacy, oriented towards assessment and measurement issues. The analysis of literacy focuses on identifying the parameters which must be specified in any definition of literacy. Corresponding measurement issues and available literacy assessment procedures are described. The author is refreshingly aware that literacy cannot be parameterized only in terms of certain characteristics of the people being treated--performance depends critically on the characteristics of the written materials and the reading tasks.

Harman, D. Illiteracy: An overview. Harvard Educational Review, 1970, 40(2), 226-243.

The author reviews current definitions of illiteracy and functional literacy and discusses their relationship to estimates of the extent of illiteracy and to literacy education. Although the article is somewhat dated, Harman's anticipation of the need to measure literacy in relation to the functional requisites of particular societies is noteworthy. He argues that adult basic education efforts here and abroad should be planned on a situation-specific basis, with goals, content and evaluative components derived independently of the usual grade school equivalencies.

Hunter, C., with D. Harman. Adult literacy in the United States. New York: McGraw Hill Book Co., 1979.

Much of the material in this volume updates and elaborates on the earlier Harman (1970) article. Here we consider just the first two chapters. In Chapter I of this volume, Hunter and Harman differentiate two types of literacy: conventional and functional literacy. Conventional literacy is defined as the "ability to read, write and comprehend texts on familiar subjects and to understand

---

*For reference on other aspects of functional literacy, see Reder, S., M. F. Walton, and K. R. Green. A bibliographic guide to functional literacy. Portland, OR: Northwest Regional Educational Laboratory, 1979. (ED 189 197)

whatever signs, labels, instructions and directions are necessary to get along in one's environment." Functional literacy is "the possession of skills perceived as necessary by particular persons and groups to fulfill their own self-determined objectives as family and community members, citizens, job-holders..." This definition includes the ability to gain access to information they may want to use.

In Chapter II, the authors examine demographic data in an effort to delineate the illiterate population of the United States. For this purpose they take high school graduation as the criterion of literacy. Although this is in many ways an arbitrary and inappropriate measure, there is abundant data on high school completion. They then characterize the illiterate American population in terms of several social and economic variables. They indicate that although the exact number of illiterates is not known, and although more individuals are completing high school and achieving "acceptable" levels of literacy, well over one-third of the adult population suffers some educational disadvantage. Hunter and Harman stress that

Kirsch, I., & J. Guthrie. The concept and measurement of functional literacy. Reading Research Quarterly, 1977-78 13(4), 485-507.

This article reviews the literature on the concept and measurement of functional literacy. The authors differentiate the various meanings applied to the term functional literacy and their implications for its definition and measurement. Various measures used in past surveys to assess functional literacy are critically examined and specific limitations of each approach are discussed.

Peck, C., & M. Kling. Adult literacy in the Seventies: Its definition and measurement. Journal of Reading, 1977, 20(8), 677-682.

After summarizing definitions and estimates of functional literacy/illiteracy that emerged from several major studies, this article critically examines several instruments developed to assess functional literacy levels. Special attention is given to the assessment of real-life reading skills. The article concludes by noting that any definition and assessment is relevant only to a given subpopulation rather than the United States as a whole.

Smith, L. L. Literacy: Definitions and implications. Language Arts, 1977, 54(2), 135-138.

Although somewhat uncritically presented, this articles provides a lay-oriented summary of the literature on issues involved in defining functional literacy. Based on identification of these issues, the author outlines the elements needed in any definition of functional literacy.

APPENDIX B

SUMMARY OF COMMON TEST SCORES

SCORES FREQUENTLY ASSOCIATED WITH NORM REFERENCED TESTS

	DEFINITION	MAJOR ADVANTAGES	MAJOR DISADVANTAGES
PERCENTILE RANK	The percentile rank establishes a student's standing relative to a norm group in terms of the percentage of students who scored at or below his or her raw score. For example, a student who scored at the 98th percentile achieved a raw score which was higher than the raw scores of 98 percent of the norm group who took the same test under the same conditions.	1. Percentiles show the relative standing of individuals compared to a normative group. 2. They are familiar to most public school personnel, though probably not the general public. 3. Percentiles are relatively easily explained.	1. Percentiles are frequently confused with the percent of the total number of test items answered correctly. 2. Since the percentile scale does not have equal units of measurement, percentiles should not be used in the computation of group statistics.
GRADE EQUIVALENT SCORE	The grade equivalent score indicates the performance of a student on a particular test relative to the median performance of students at a given grade level and month; e.g., a fifth grader who receives a grade equivalent score of 8.2 on a reading test achieved the same raw score performance as the typical eighth grader in the second month of eighth grade would be expected to achieve on the same fifth grade test.	1. It appears easy to communicate the standing of an individual student relative to a grade level (most people believe they understand what is meant by grade equivalent scores).	1. Grade equivalents are easily misunderstood and misinterpreted. 2. Achievement expressed in grade equivalent score units cannot be meaningfully compared with each other in several instances. a. Grade equivalent scores cannot be meaningfully compared for the same student (or group of students) over time. b. Grade equivalent scores cannot be meaningfully compared for the same student (or group of students) across subject matter areas. c. Grade equivalent scores cannot be meaningfully compared for the same student (or group of students) across different tests. 3. Many grade equivalent scores are statistical projections (interpolations or extrapolations). In the later grades it is not uncommon to find grade equivalent scores of two or three grade levels above or below the student's actual grade level, but these scores are of doubtful accuracy. 4. The grade equivalent scale is not composed of equal sized units. Having equal sized units implies that the underlying difference between any two scores is the same throughout the scale.

SCORES FREQUENTLY ASSOCIATED WITH NORM REFERENCED TESTS

	DEFINITION	MAJOR ADVANTAGES	MAJOR DISADVANTAGES
STANDARD SCORE	Standard scores are derived from raw scores, but express the results of a test on the same numerical scale regardless of grade level, subject area or test employed.	1. Since the mean and standard deviation of the standard score scales are pre-specified, a student's standard score immediately communicates two important facts about his or her performance on that test: a. Whether the student's score is above or below the mean. b. How far above or below the mean, in standard deviation units, his or her performance is. 2. The constant numerical scale of standard scores facilitates comparisons: a. Across students taking the same test. b. Across subject matter areas for the same student. 3. Standard scores are derived in a way that maintains the equal interval property in their units which is absent in percentile and grade equivalent scores. Therefore, summary statistics may be meaningfully interpreted when calculated on standard scores.	1. The most useful interpretation of standard scores requires some knowledge of statistics (i.e., mean and standard deviation) and hence may not be appropriate for audiences who have not been exposed to these concepts (e.g., parents, the news media). 2. Given the variety of standard scores available, there may be potential confusion in expressing the same test performance with so many different numerical values. 3. The conversion of raw scores to standard scores may either maintain the shape of the distribution observed, or may transform the distribution to another, more interpretively convenient shape (e.g., the normal distribution); and the procedures employed in specifying the conversion process may not be immediately obvious.
NORMAL CURVE EQUIVALENTS	A standard score system having 99 equal intervals. The average corresponds to the 50th centile; the 1st & 99th NCEs correspond to the 1st & 99th centiles. Range: generally 1-99 but can be higher and lower.	1. Same as standard score systems. 2. Permit aggregation of data from a wide variety of tests.	1. They are relatively new. 2. They depend upon standard scores or percentiles. 3. Not all test publishers use them.

SCORES FREQUENTLY ASSOCIATED WITH NORM REFERENCED TESTS

	DEFINITION	MAJOR ADVANTAGES	MAJOR DISADVANTAGES
EXPANDED SCALE SCORE	Expanded scale scores are a type of standard score whose scale is designed to extend across grade levels and whose mean increases progressively as the grade level increases.	1. Expanded scores facilitate longitudinal comparisons of an individual across grade levels.	1. Different test publishers use different terms to refer to their expanded scale scores (e.g., growth scale values, achievement development scale scores, standard score, scale score) and this may be confusing when considering results from different tests.
		2. Expanded scale scores provide the vehicle for expressing a performance obtained at one grade level to the norm group of another. This is useful when the appropriate level of a test to be administered to a student is judged to be other than that of his or her grade level (i.e., functional level testing).	2. Different tests use different ranges, and standard deviations in deriving their expanded scale scores. Thu , results from different tests expressed in expanded scale score units cannot be readily compared.
		3. Since they were designed as equal interval, their scores may be mathematically manipulated (e.g., averaged).	3. The statistical properties of expanded scale scores are often not as uniform as theoretically desired.
STANINE	Stanines are a standard score scale consisting of nine values with a mean of five and a standard deviation of two. If the distribution of scores is normal, each stanine includes a known proportion of the scores in the distribution.	1. As in all standard scores, stanines have the same meaning across different tests, different grade levels and different content areas. 2. Stanines consist of only nine possible scores and thus may be easier to communicate to audiences not familiar with measurement terminology. Verbal labels may be given to each stanine value to facilitate interpretation.	1. Since some of the stanines encompass a wide range of score , their use in reporting can be insensitive to differences between students' performance that are more apparent from the use of other test scores.

SCORES FREQUENTLY ASSOCIATED WITH OBJECTIVE REFERENCED TESTS

	DEFINITION	MAJOR ADVANTAGES	MAJOR DISADVANTAGES
RAW SCORE	The number of items on a test or subtest answered correctly by the student.	1. Virtually no statistical or measurement expertise is needed to calculate raw scores. 2. Raw scores are the necessary first step in expressing test performance in any of a number of other ways (e.g. standard scores, percentiles.)	1. By themselves, raw scores offer no indication as to how a student who has mastered the skills represented on the test "should" perform (i.e., criterion referenced) or how other students at the same grade level have performed (i.e., norm referenced.)
ITEMS CORRECT	The proportion of the total number of items answered correctly by the student.	1. Very little statistical or measurement expertise is required to understand this expression of test performance. 2. If the content area is sufficiently represented by the items on the test, the percent correct provides an expression of the proportion of the subject matter mastered by the student.	1. No notion of test difficulty or expected performance is contained in this score. Unless accompanied by a standard for mastery or information as to how a student's peers have performed in the test, misinterpretations may arise.
OBJECTIVE MASTERY SCORE	When a standard for mastery has been applied to a set of items for a specific objective, a student's performance in terms of that objective is expressed as having mastery or non-mastery of the objective.	1. The objective mastery score compares the student's performance on that objective to a judged standard of what he or she should know of the skills required to master it. This score can be very useful in diagnosing a student's specific strengths and weaknesses. 2. When the subject matter requires a successive accumulation of skills (e.g., elementary math), objective mastery scores may be extremely useful in monitoring the progress of students in specific skill areas.	1. Objective mastery scores are difficult to compare across different tests. Items designed to measure the same objective may differ in difficulty or have different standards for mastery on different tests. 2. If a purpose in testing is to differentiate among students, objective mastery scores do not present a very useful index. Different raw scores above or below the mastery level are viewed as the same--either mastery or non-mastery.

APPENDIX C

PUBLISHED TESTS OF BASIC READING SKILLS

Most reading tests are part of a multisubject battery. The first list in this appendix provides information on such tests. The second list gives information on tests which measure only reading skills. These lists are modified versions of the ones in Anderson, B. L., R. J. Stiggins, and S. B. Hiscox, Guidelines for Selecting Basic Skills and Life Skills Tests, Portland, OR: Northwest Regional Educational Laboratory, 1980.

Although the lists are intended to include all appropriate tests published since 1970, some may have been inadvertently overlooked. Inclusion on the list does not imply endorsement.

Appendix C (continued)

MULTISUBJECT ACHIEVEMENT BATTERIES

Tests and Subscores	Grade Level(s)*	Publication Date	Publisher**
Adult Basic Learning Examination (ABLE)	Adult	1974	PSYCH CORP
Reading			
Arithmetic			
Spelling			
Vocabulary			
Adult Performance Level Survey (APLS)	9-Adult	1976	ACT
Reading			
Computation			
Writing			
Identifying Facts and Terms			
Problem Solving			
Alaska Instructional Diagnostic System (AIDS)	1-8	1977	SRRC
Reading			
Mathematics			
American School Achievement Tests, Revised Edition (ASAT)	1-9	1975	BMC
Reading			
Arithmetic			
Language			
Spelling			
Social Studies			
Science			
California Achievement Tests Forms C & D (CAT)	K-12	1977	CTB
Reading			
Mathematics			
Language			
Spelling			
Reference Skills			

*Not all subtests are available at all grade levels.

**Names, addresses and phone numbers are given in Appendix H.

Appendix C (continued)

MULTISUBJECT ACHIEVEMENT BATTERIES

Tests and Subscores	Grade Level(s)	Publication Date	Publisher
California Assessment Program Survey of Basic Skills Reading Mathematics Written Expression Spelling	12	1974	CSDOE
Comprehensive Assessment Program The Achievement Series Reading Mathematics Language Arts	PreK-12	1980	SFC
Comprehensive Tests of Basic Skills Expanded Edition Form S & T (CTBS) Reading Mathematics Language Arts Reference Skills Science Social Studies	K-12	1976	CTB
Criterion Test of Basic Skills Reading Arithmetic	Kinder-garten-8	1976	ATP
Diagnostic Skills Battery Reading Mathematics Language Arts	1-8	1976	STS
Iowa Tests of Basic Skills Multi-level Edition Forms 7 & 8 Reading Comprehension Mathematics Skills Language Skills Work-Study Skills Vocabulary	3-9	1978	RPC
Iowa Tests of Educational Development: SRA Assessment Survey Reading Mathematics Language Arts Social Studies Science	9-12	1974	SRA

Appendix C (continued)

MULTISUBJECT ACHIEVEMENT BATTERIES

Tests and Subscores	Grade Level(s)	Publication Date	Publisher
Metropolitan Achievement Tests (METRO '78) Reading Comprehension Mathematics Language Social Studies Science	K-12	1978	PSYCH CORP
National Educational Development Tests Mathematics Usage English Usage Social Studies Reading Nature Sciences Reading Word Usage	7-10	1974	SRA
Scholastic Testing Service Educational Development Series Scholastic Tests Reading Mathematics English Social Studies Science Solving Everyday Problems USA in the World Nonverbal Ability Verbal Ability School Interests School Plans Career Plans	2-12	1976	STS
Science Research Associates Achievement Series (ACH) Forms 1 & 2 Reading Mathematics Language Arts Social Studies Science Reference Materials Applied Skills	K-12	1978	SRA
Science Research Associates High School Placement Reading Arithmetic or Modern Math Language Arts Social Studies Science	9	1973	SRA

Appendix C (continued)

MULTISUBJECT ACHIEVEMENT BATTERIES

Tests and Subscores	Grade Level(s)	Publication Date	Publisher
Science Research Associates Norm Referenced/ Criterion Referenced Testing Program Reading Mathematics	4-10	1977	SRA
Sequential Tests of Educational Progress: Series III Levels E-J Reading Mathematics Computation Mathematics Basic Concepts Writing Skills Social Studies Science Study Skills/Listening Goal Orientation Index	1-12	1979	AWPC
SOI Learning Abilities Test Reading Arithmetic	1-11	1975	SOI
Stanford Achievement Test 1973 Edition (SAT) Reading Mathematics Social Studies Science Listening Comprehension Spelling	1-9	1973	PSYCH CORP
Stanford Test of Academic Skills (TASK) Reading Mathematics English	8-Adult	1975	PSYCH CORP
Tests of Adult Basic Education (TABE) Reading Mathematics Language Arts	Adult	1976	CTB

Appendix C (continued)

MULTISUBJECT ACHIEVEMENT BATTERIES

Tests and Subscores	Grade Level(s)	Publication Date	Publisher
Tests of Achievement and Proficiency Form T (TAP)	9-12	1978	RS
Reading Comprehension			
Mathematics			
Written Expression			
Social Studies			
Science			
Using Sources of Information			
Applied Proficiency Skills			
United States Employment Service Basic Occupational Literacy Test (USES BOLT)	Adult	1974	USDL
Reading Vocabulary			
Reading Comprehension			
Arithmetic Computation			
Arithmetic Reasoning			
Wide Range Achievement Test Revised Edition (WRAT)	K-Adult	1978	JA
Reading			
Arithmetic			
Spelling			

Appendix C (continued)

READING TESTS

Tests	Grade Level(s)	Publication Date	Publisher
Analysis of Skills: Reading	1-8	1976	STS
Analytical Reading Inventory	2-9	1977	CEMPC
Clarke Reading Self-Assessment Inventory	11-Adult	1978	ATP
Criterion Referenced: Reading Tactics	7-12	1977	SFC
Cutrona Reading Inventory (Oral)	K-Adult	1975	CEI
Developmental Reading: Diagnostic/ Prescriptive Tests: Fundamental Stage	3-Adult	1975	PSAA
Diagnostic Reading Scales	1-12	1975	CTB
Diagnostic Reading Test: Pupil Progress Series	1-8	1970	STS
Diagnostic Screening Test: Reading	1-12	1979	SC
Fountain Valley Teacher Support System in Secondary Reading	7-12	1976	RLZA
Gates-McGinitie Reading Tests, Second Edition	1-12	1978	RPC
Gates-McGinitie Reading Tests, First Edition	1-12	1970	RPC
Individualized Criterion Referenced Testing: Reading (ICRTR)	K-8	1976	EDC
Literacy Assessment Battery	Adult	1976	ERIC
Mastery: An Evaluation Tool: (SOBAR)	K-9	1976	SRA
McCarthy Individualized Diagnostic Reading Inventory--Revised	2-Adult	--	EPS
McGrath Diagnostic Reading Test	1-13	1976	MPC
Minimum Reading Competency Test	1-13	1976	KHSD
Nelson-Denny Reading Test Forms C & D	9-12	1976	RPC
Nelson Reading Skills	3-9	1977	RPC
Objectives-Referenced Bank of Items and Tests: Reading and Communication Skills (ORBIT: RCS)	K-Adult	1975	CTB
Oral Word Recognition Test	1-13	1973	MPC
Performance Assessment in Reading (PAIR)	7-9	1978	CTB
Power Reading Survey Test	1-12	1975	BRA
Prescriptive Reading Performance Test	K-12	1978	WPS
Reading Skills Diagnostic Test	2-8	1971	BP
Reading Skills Competency Test	K-7	1979	CARE
SRA Reading Index	9-Adult	1974	SRA
Stanford Diagnostic Reading Test	1-13	1976	PSYCH CORP
Test of Reading Comprehension (TORC)	1-8	1978	PRO-ED
Woodcock Reading Mastery Tests (WRMT)	K-12	1973	AGS

APPENDIX D

PROCEDURES FOR SCORING WRITING SAMPLES

While many objective tests can be machine scored, writing tests that rely on writing samples must be hand scored by persons trained to use designated criteria and performance standards. Several different methods have been devised for scoring writing samples, depending on the assessment purpose. Appropriateness depends upon what information is needed, how it will be used, and what resources are available. Some scoring methods are more complicated and costly than others.

Three scoring methods will be discussed here: holistic, analytic and primary trait scoring. The methods are similar in that all require careful training of scorers (usually language arts teachers) in a group setting for about half a day. Once scorers are trained, scoring is usually conducted in a carefully structured setting to allow reading of papers by two readers (and a resolver, if necessary) and exercising of quality control procedures. Under such conditions high reliability can be achieved.

Holistic Scoring

Holistic scoring involves reading a paper for an overall or "whole" impression. No specific trait, such as organization, syntax or originality, is individually addressed. A reader makes a judgment in much the same way that he or she decides whether a novel or an essay in Newsweek magazine is superior, mediocre, or slipshod.

Raters are asked to make no marks upon the paper. The objective is to score quickly, trusting first impressions.

Raters use "range finders" as guides. These are actual student papers, chosen for their representativeness, from the total group of papers to be scored. There is a range finder for each score level: e.g., 4, 3, 2 and 1. Virtually all trained readers agree on the score each paper should receive, so these papers serve as effective models to assist raters in assigning scores.

Range finder papers are intended to represent the approximate midpoint of each range. For example, a 4 range finder should not be considered the best possible paper from the sample. Rather, it should be considered a "middle 4"; that is, some 4's in the sample will be a little better, some not quite so good.

Range finders are intended as guides. Clearly, not every contingency can be anticipated. Readers are encouraged to (1) trust their own judgment, and (2) confer with a head reader to resolve any unusual problems.

One holistic score is determined to represent the overall quality of the writing. Although the specific strengths and weaknesses of a given paper are not delineated, final reports to parents and others can include descriptions of the typical characteristics of papers receiving each score.

Analytical Scoring

Analytical scoring is a trait-by-trait analysis of a paper's merit. Individual traits considered important to any piece of writing in any context are selected for analysis. For example, papers of students asked to write a letter expressing an opinion may be scored on ideas, organization and wording.

The traits are scored one at a time. The scorer's impression of one trait should not influence the scoring of any other trait. Readers are presented with guidelines for scoring each trait. The guide presents an elaboration of each point on the rating scale for each trait. Scores are reported separately for each trait. Range finders such as those used in holistic scoring are sometimes used along with the scoring guides.

Primary Trait Scoring

Primary trait scoring is similar to analytical scoring in that it focuses on one or more specific characteristics of a given piece of writing. But while analytical scoring attempts to isolate general important characteristics, primary trait analysis is situation specific. That is, the most important--or primary--trait(s) in a letter to the editor will not likely be the same as that (those) in a set of directions for assembling a toy.

The primary trait system is based on the premise that all writing is done in terms of an audience, and that successful writing will have the desired effect upon that audience, whether newspaper reader or employer. In scoring, papers are judged on the likelihood of their producing the desired response.

Because they are situation-specific, primary traits differ from item to item, depending on the nature of the assignment. Suppose a student were asked to give directions for taking the bus from home to school. The primary trait might then be sequential organization; any clear, unambiguous set of directions would necessarily be well organized. The chart on the next page summarizes the key features of the scoring procedures discussed above.

Comparison of Alternatives

	Holistic	Analytical	Primary Trait
GOAL	Overall Impression	Trait-by-trait Analysis	Situation Specific Analysis
PREPARATION	Allow 1/2 day to find range finders; 1/2 day to train readers	1 day to I.D. traits; 1 day per trait to dev. criteria; 1 day to refine scoring; 1/2 day to train readers	Same as analytical
SCORING TIME	1 to 2 min. per paper depending on length	1 to 2 min. per trait per paper	1 to 2 min. per trait per paper
CONTEXT	Contexts where rank order is useful	Contexts where skill analysis is useful	Contexts where skill analysis is useful

The following references provide further information on direct measures of writing.

Spandel, V., & R. Stiggins. Direct Measures of Writing Skill: Issues and Applications. Portland, OR: Northwest Regional Educational Laboratory, 1980. ($4.25 prepaid)

This 64-page monograph presents a status of writing assessment, an overview of direct writing assessment procedures and information on how to adapt writing assessment to specific purposes.

Bridgeford, N., & R. Stiggins. A Consumer's Guide to Writing Assessment. Portland, OR: Northwest Regional Educational Laboratory, 1981.

The Guide compares direct and indirect assessment methods, describes how to develop direct assessment measures and select indirect measures. Extensive lists are provided of organization and consultants across the country who can provide assistance in developing direct measures. Profiles of indirect measures are also provided.

APPENDIX E

PUBLISHED TESTS OF WRITING SKILLS

The test list is taken from Bridgeford, N., & R. Stiggins. A Consumer's Guide to Writing Assessment, Portland, OR: Northwest Regional Educational Laboratory, 1981. Although the list is intended to include all tests published since 1970 and designed for grade 8 or above, some tests may have been inadvertently overlooked. Inclusion on the list does not imply endorsement.

Appendix E (continued)

PUBLISHED TESTS OF WRITING SKILLS

Test	Grade Level(s)	Publication Date	Publisher
ACT Assessment English Usage Test (subtest of an achievement battery) Punctuation, grammar, style, diction, logic, organization	9-13	5/year	ACTP
Adult Performance Level* Adult Survey--Writing Subscore (score based on items embedded in a longer test) Ability to recognize appropriately written material on various forms and documents used in everyday life	Adults	1976	ACTP
Adult Performance Level High School Survey-Writing Subscore (score based on items embedded in a longer test) Ability to recognize appropriately written material on various forms and documents used in everyday life	9-12	1976	ACTP
American School Achievement Tests: Language, Spelling (subtests of an achievement battery) Correct usage, punctuation, capitalization, sentence recognition, grammar, spelling	7-9	1963	BMC
APL Content Area Measures (series of individual tests, each including a writing subscale) tests available in Occupational Knowledge, Community Resources, Consumer Economics, Health and Governmental Law Ability to recognize appropriately written materials on forms and documents used in everyday life	9-12 Adults	1977	ACTP

*Writing sample included

Appendix 3 (continued)

Test	Grade Level(s)	Publication Date	Publisher
Analysis of Skills (ASK) Language Arts Capitalization and punctuation, usage, sentence knowledge, composing process	2-8	1977	STS
Basic Skills Assessment Program* Writer's Skills Test Spelling/punctuation, capitalization, usage, logic, evaluation	7-12	1977	AWPC
California Achievement Tests/Language (Subtest on an achievement battery) Language: auding, mechanics, usage and structure, spelling Language Mechanics: capitalization and punctuation Language Expression: usage, sentence structure, and paragraph organization	1-12	1970	CTB
California Achievement Tests/Language (Subtest of an achievement battery) Mechanics, expression, spelling	K-12	C: 1977 D: 1978	CTB
College English Placement Test* Topic selection, organizing materials for presentation, editing, composition	College	1969	RPC
College-level Examination Program (CLEP)* General Examination in English Composition and Subject Examinations in College Composition and Freshman English English Composition: Logical and structural relationship within sentences; economy, precision and clarity of communication; logical and attention to purpose and audience College Composition: Sentence structure, paragraph and essay construction, style, logic, language history and reference skills Freshman English: Style, logic, syntax, usage, punctuation, paragraph construction, dictionary and research skills	College	Variable at CLEP test centers	CEEB

*Writing sample included

Appendix E (continued)

Test	Grade Level(s)	Publication Date	Publisher
College Outcome Measures Project,* Composite Examination Writing Subscale (Subtest of an achievement battery) Ability to address an audience, organize and develop an essay and use language and sentence structure	College	1980	ACTP
Comparative Guidance and Placement Program Sentences Test; also referred to as Written English Expression Test (Subtest of achievement battery) Grammar, usage, word choice, sentence structure, logical relationships within sentences, clarity of expression	College	1973	CEEB
Comprehensive Assessment Program High School Subject Tests/Writing and Mechanics Test, Language Test <u>Language Test</u>: Spelling, punctuation and capitalization, correctness of expression <u>Writing and Mechanics Test</u>: Paragraph development, usage, paragraph structure	9-12 Adult	1980	SFC
Comprehensive Assessment Program - Spelling, Capitalization and Punctuation, Grammar and Language Total (Subscores of an achievement battery) Spelling, capitalization, punctuation, grammar	2-8	1980	SFC
Comprehensive Tests of Basic Skills/ Language (Subtest of an achievement battery) Mechanics, expression, spelling	2-10	Q-1968 R-1969	CTB
Comprehensive Tests of Basic Skills/ Language (Subtest of an achievement battery) Mechanics, expression, spelling	1-12	S-1973 T-1975	CTB

*Writing sample included

Appendix E (continued)

Test	Grade Level(s)	Publication Date	Publisher
Content Evaluation Series Language Arts Test/Language Ability, Composition (Subtests of achievement battery) Language Ability: Sentence structure, word form and function, mechanics, diction Composition: Invention, arrangement, and style	7-9	1969	RPC
Description Tests of Language Skills/ Sentence/Structure Using complete sentences; using coordination and subordination appropriately	9-12 College	1977-81	ETS
Descriptive Tests of Language Skill/ Logical Relationships Categorizing ideas, using appropriate connectives, making analogies, recognizing principles of organization	9-12 College	1977-81	ETS
Descriptive Tests of Language Skills/Usage Ability to use pronouns, modifiers, diction and idioms, verbs	9-12 College	1977-81	ETS
Diagnostic Skills Battery/Language Arts (Subtest of an achievement battery) Capitalization and punctuation, usage, sentence knowledge and composing process	1-8	1977	STS
Educational Development Series/English (Subtest of achievement battery) Grammar, capitalization, punctuation, spelling	9-12	1972	STS
Educational Development Series/English (Subtest of an achievement battery) Capitalization, punctuation, usage	1-12	1977	STS
Essentials of English Tests Skills in spelling, grammatical usage, word usage, sentence structure, punctuation and capitalization	7-12 & 13	1961	AGS

Appendix E (continued)

Test	Grade Level(s)	Publication Date	Publisher
Hoyum-Sanders English Test (Four forms) Division one covers sentence recognition, capitalization, punctuation, contractions, possessives, spelling, correct usage, and alphabetization. The second and third divisions cover sentence recognition, capitalization, punctuation, correct usage, and reference materials such as guide words and index	2-8	1964	BEM
Iowa Tests of Basic Skills, Multilevel Battery/Language Skills (Subtest of achievement battery) Spelling, capitalization, punctuation, usage	3-9	1978	RPC
Iowa Tests of Basic Skills, Primary Battery/Language Skills (Subtest of an achievement battery) Spelling, capitalization, punctuation, usage	K-6	1979	RPC
Iowa Tests of Educational Development/ Language Usage (Subtest of an achievement battery) Punctuation, capitalization, manner of expression, word and sentence order, organization of ideas, spelling	9-12	1971	SRA
IOX Basic Skills Tests,* Secondary Level-Writing Subtest Using words correctly, checking mechanics, selecting correct sentences, expressing ideas in writing	9-12	1978	IOX

*Writing sample included

Test	Grade Level(s)	Publication Date	Publisher
Metropolitan Achievement Tests/Language* (Subtest of an achievement battery) Punctuation and capitalization, usage, grammar and syntax, spelling, study skills. Listening comprehension at grades 1.5 to 12.9	1.5 to 12.9	1978	PSYCH CORP
Metropolitan Language Instructional Tests Punctuation and capitalization, usage, grammar and syntax, spelling study skills. Listening comprehension at grades 1.5 to 4.9	1.5 to 9.9	1978	PSYCH CORP
Minnesota High School Achievement, Examinations/Language Arts Subtest (Separate test for each grade, three forms per grade) Content Areas: Gr. 7: Language study skills, spelling, word knowledge, kinds of sentences, usage, sentence structure, punctuation, capitalization; Gr. 8: Spelling, vocabulary, kinds of sentences, faulty expression, verb usage, use of words, types of sentences, capitalization and punctuation, usage, general information, literature; Gr. 9: Spelling, vocabulary, sentences, sentence structure, punctuation, usage, composition, library, literature (interpretation), literature (knowledge); Gr. 10: Sentence structure, word discrimination, spelling, punctuation, diction, reading and literature, general information; Gr. 11: Sentence structure, word discrimination, spelling, punctuation, organization, library skills, literary style, literary figures, quotations, literature; Gr. 12: Spelling, vocabulary, punctuation, word discrimination, word usage, sentence structure, library skills, literature	7-12	1976	AGS

*Writing sample in process

Appendix E (continued)

Test	Grade Level(s)	Publication Date	Publisher
Missouri College English Test Mechanics and Effectiveness of Written Expression: punctuation, capitalization, grammar, spelling, sentence style and structure, paragraph organization	College	1965	HBJ
National Educational Development Tests/ English Usage (Subtest of an achievement battery) Ability to use such basic elements of correct and effective writing as punctuation, capitalization, diction, sentence reconstruction, and paragraph organization	7-10	1971	SRA
Sequential Tests of Educational Progress Intermediate and Advanced; Forms D-J- Writing Skills (Subtest of achievement battery) Capitalization and Punctuation: ability to recognize errors in mechanics or usage Word Structure and Usage: ability to detect errors in use of parts of speech embedded in sentences Sentence and Paragraph Organization: language construction skill and ability to recognize appropriate organization	3.5 to 12.9	1979	AWPC
Sequential Tests of Educational Progress Series II/Mechanics of Writing, English Expression (Subtests of an achievement battery) Mechanics: Spelling, capitalization, punctuation Expression: Ability to evaluate the correctness and effectiveness of sentences	4-College	1971	ETS
Stanford Achievement Test/Spelling,** Language (Subtests of achievement battery) Spelling, capitalization, punctuation, usage, syntax, language sensitivity, dictionary and other reference skills	1-9	1973	PSYCH

*Writing sample available in 1982

Appendix E (continued)

Test	Grade Level(s)	Publication Date	Publisher
Test of Adolescent Language Ability to express thoughts in graphic form, ability to write, ability to understand and generate syntactic structures, abilityl to use language expressively	6-12	1980	PRO-ED
Test of Standard Written English Grammar, usage, sentence logic	9-12 College	6/year	ETS
Test of Written Language* Vocabulary, thematic maturity, ability to produce meaningful thought units, handwriting, spelling, word usage, style	2-8	1978	PRO-ED
Tests of Achievement and Proficiency (Subtest of achievement battery) Capitalization, punctuation, grammar, usage, organization, spelling	9-12	1978	RPC
Walton-Sanders English Test (Four Forms) Ability to recognize obvious errors in spelling, sentence structure, punctuation, the use of the past tense and past participle forms of verbs, the use of nominative and objective forms of pronouns, the use of English idioms, especially those involving a choice of prepositions, and other common faults	9-13	1964	BEM
WRITE: Senior High* Mechanics, punctuation, usage, vocabulary, spelling, organization and format	9-12	1979	CTB
Writing Proficiency Program* Sentence fragments, run-on sentences, subject-verb agreement, verb form, pronoun case, punctuation and mechanics, capitalization, spelling, paragraph and essay organization, paragraph coherence, topic sentences	9-13	1979	CTB

*Writing sample included

Test	Grade Level(s)	Publication Date	Publisher
Writing Proficiency Program/Intermediate* System Sentence structures, sentence mechanics, paragraph structures, sentence fragments, run-on sentences, adjectives and adverbs, personal pronouns, verb tense, verb agreement with subject, misplaced modifiers, conjunctions, commas, end marks and quotation marks, capitalization, topic/summary sentences, sequence of sentences, use of transitions	6-9	1981	CTB

*Writing sample included

APPENDIX F

TESTS OF EVERYDAY LITERACY ACTIVITIES

Test	Grade Level(s)	Publication Date	Publisher
Adult Performance Level Functional Literacy Test	9-Adult	1978	ACTP
Basic Skills Assessment	8-12	1977	ETS
Everyday Skills Test (EDST)	6-12	1975	CTB
IOX Basic Skills Test	9-12	1978	IOX
Life Skills: Tests of Functional Competencies in Reading and Math	9-12	--	RPC
Minimum Essentials Test (MET)	8-Adult	1980	SFC
NM Consumer Mathematics Test	9-12	1973	NMDOE
Reading/Everyday Activities in Life (REAL)	9-Adult	1972	CAL-P
SRA Coping Skills: A Survey Plus Activities	7-Adult	1979	SRA
SRA Survival Skills in Reading & Mathematics	6-Adult	1976	SRA
STS Educational Development Series: Scholastic Tests	2-12	1976	STS
Senior High Assessment of Reading Performance (SHARP)	10-12	1978	CTB
Stories About Real-Life Problems	5-8	--	NIU
Test of Consumer Competencies	8-12	1976	STS
Test of Everyday Writing Skills (TEWS)	9-12	1978	CTB

APPENDIX G

CLERICAL TESTS INVOLVING ON-THE-JOB LITERACY SKILLS

Tests and Subscores	Grade Level(s)	Publication Date	Publisher
General Clerical Test Clerical speed and accuracy Numerical ability Verbal facility	9-Adult	1972	PSYCH CORP
Short Employment Tests Verbal Numerical Clerical	Adult	1972	PSYCH CORP
Short Tests of Clerical Ability Arithmetic Business vocabulary Checking Coding Directions Filing Language	Adult	1973	SRA
Short Clerical Aptitudes Office vocabulary Office arithmetic Office checking	9-Adult	1973	SRA

APPENDIX H

PUBLISHERS' NAMES AND ADDRESSES

ACTP	American College Testing Program P. O. Box 168 Iowa City, IA 53340 (319) 356-3711
AGS	American Guidance Service Circle Pines, MN 50014 (612) 786-4343
ATP	Academic Therapy Publications 20 Commercial Blvd. Novato, CA 94947 (415) 883-3314
AWPC	Addison-Wesley Publishing Co., Inc. South Street Reading, MA 01867
BEM	Bureau of Educational Measurement Emporia State University Emporia, KS 66801
BFA	BFA Educational Media 2211 Michigan Avenue P. O. Box 1795 Santa Monica, CA 90406 (213) 829-2901
BMC	Bobbs-Merrill Co., Inc. 4300 West 62nd Street Indianapolis, IN 46268 (317) 298-5400
BP	Brador Publications, Inc. Livonia, NY 14487
CAL-P	CAL Press, Inc. 76 Madison Avenue New York, NY 10016 (212) 685-0892
CARE	The Center for Applied Research in Education, Inc. Route 59 West Nyack, NY 10994 (914) 358-8991

CEEB	College Entrance Examination Board Box 2815 Princeton, NJ 08541
CEI	Cutronics Educational Institute 128 W. 56th Street Bayonne, NJ 07002
CEMPC	Charles E. Merrill Publishing Company 1300 Alum Creek Drive Columbus, OH 43216 (614) 997-1221
CSDOE	California State Department of Education 721 Capitol Mall Sacramento, CA 95814 (916) 445-4688
CTB	California Testing Bureau/McGraw-Hill Del Monte Research Park Monterey, CA 93940 (408) 649-8400
EPS	Educators Publishing Service 75 Moulton Street Cambridge, MA 02138 (617) 547-6706
ERIC	ERIC Document Reproduction Service P. O. Box 190 Arlington, VA 22210
ETS	Educational Testing Service Princeton, NJ 08541
HBJ	Harcourt, Brace, Jovanovich, Publishers 757 Third Avenue New York, NY 10017
IOX	Instructional Objectives Exchange Box 24095 Los Angeles, CA 90024 (213) 474-4531
JA	Jastak Associates, Inc. 1526 Gilpin Avenue Wilmington, DE 19806 (302) 652-4990

KHSD	Kern High School District 2000 24th Street Bakersfield, CA 93301
MHBC	McGraw-Hill Book Company 8171 Redwood Highway Novato, CA 94947
MPC	McGrath Publishing Company P. O. Box 9001 Wilmington, NC 28402 (919) 763-3757
NIU	Northern Illinois University Alan M. Voelker Curriculum & Instruction De Kalb, IL 60115 (815) 753-1000
NMDOE	New Mexico State Department of Education Monitor Education Building State Capitol Santa Fe, NM 87501 (505) 827-2429
PRO-ED	PRO-ED 333 Perry Brooks Building Austin, TX 78701
PSAA	Paul S. Amidon & Associates, Inc. 1966 Benson Avenue St. Paul, MN 55116
PSYCH COR.	The Psychological Corporation Harcourt, Brace, Jovanovich, Publishers 757 Third Avenue New York, NY 10017 (212) 888-3500
RLZA	Richard L. Zweig, Associates, Inc. 20800 Beach Blvd. Huntington Beach, CA 92648 (714) 536-8877
RPC	Riverside Publishing Company 1919 South Highland Avenue Lombard, IL 60148 (312) 629-9700

SC	Stoelting Company 1350 S. Kostner Chicago, IL 60623 (312) 522-4500
SFC	Scott, Foresman and Company 1900 East Lake Avenue Glenview, IL 60025 (312) 729-3000
SIUP	Southern Illinois University Press P. O. Box 3697 Carbondale, IL 62901
SOI	SOI Institute 214 Main Street El Segundo, CA 90245 (213) 322-5995
SRA	Science Research Associates, Inc. 155 N. Wacker Drive Chicago, IL 60606 (800) 621-0664
SRRC	Southwest Regional Resource Center 127 S. Franklin Street Juneau, AK 99801 (907) 586-6806
STS	Scholastic Testing Service 480 Meyer Road Bensenville, IL 60106 (312) 766-7150
USDOL	United States Department of Labor Bureau of Labor Statistics 1515 Broadway New York, NY 10036 (212) 399-5405
WPS	Western Psychological Services 12031 Wilshire Blvd. Los Angeles, CA 90025