TEST REVIEW:
Workplace Assessment Tools

Maurice Taylor
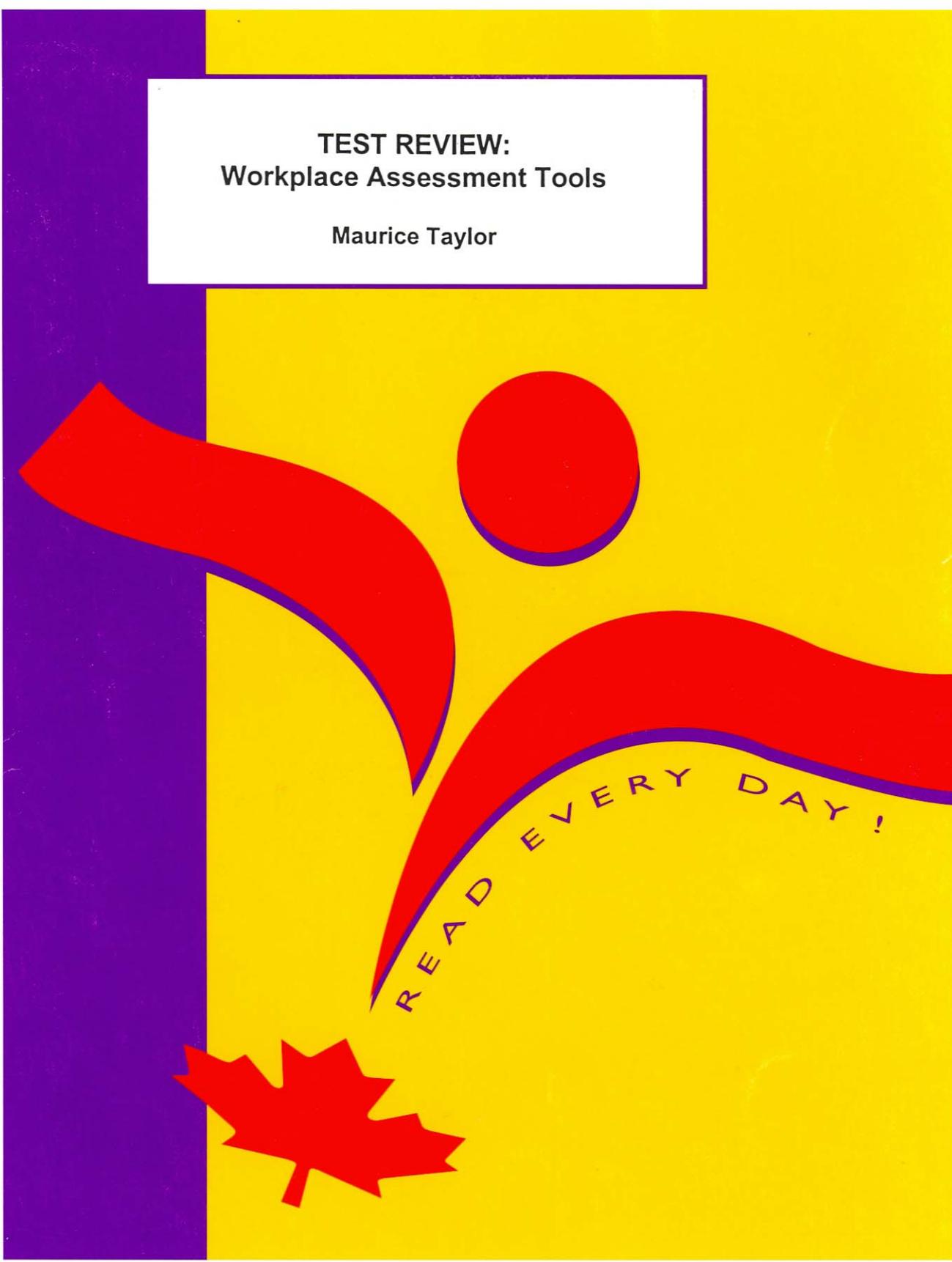
READ EVERY DAY !

Canada

Table of Contents

INTRODUCTION

Learner assessment in workplace literacy programs is still a much debated issue in the field of adult education. It continues to mean different things to different people. On the one hand, the debate can focus on the essential components of a standardized test, and on the other hand it can center on ways to obtain relevant information in order to make decisions. In this discussion paper, one part of the assessment question is addressed—that of testing. By means of a technical test review, seven assessment tools that have been used in workplace programs are discussed. Each instrument is described in terms of the test content, the description of scores and the **test development process.**

As a way of anchoring this review, the topic is introduced with a discourse on some of the dimensions of the testing question. These sections will serve as the backdrop for helping practitioners make decisions around testing when it is a necessary component of a workplace education program. The thrust of this paper is not to suggest that some tests are better than others but rather to argue that workplace educators need to find the test that is most appropriate to their learner and program requirements.

PART 1
DIMENSIONS OF THE TESTING QUESTION

Assessment in Learning

In workplace education programs, assessment can be viewed like a map. It helps people arrive at a destination. The map or learning plan guides the trainee to a destination which are the goals and outcomes of the program. People often use maps to find a desirable route to their destination—some want the fastest route while others enjoy a more scenic route with points of interest along the way. The same holds true for assessment. It should help learners move toward their goals and tell them how much time will be needed to reach their goals. In many of the current workplace education programs assessment is being used to place learners into levels, to prescribe specific learning plans, to monitor progress and to certify attainment and goal completion. Many practitioners and scholars have advocated that good assessment, curriculum and instruction are all part of a learning plan. This plan is made up of various strategies and techniques for instructing, assessing and reinforcing learning. Just as no single teaching strategy is effective for all learners in all situations, no single assessment instrument or process is adequate in providing quality information for all purposes.

Generally, there are three major approaches to assessment in the broader area of adult literacy and these can shed some light in the domain of workplace literacy. Each approach carries certain assumptions and limitations. In the traditional approach, functional literacy is equated with the attainment of a particular grade level score on tests of achievement. Through the use of such tests, it is possible to estimate percentages of various population groups performing at or above specified reading grade levels. Persons performing at or above a specified level are considered to have adequate skills to perform successfully on materials or tasks judged to be of comparable grade level difficulty.

The second approach is called competency based. In this approach tasks are developed from non-school types of materials and administered to national samples. By analyzing task responses, it is possible to estimate the proportion of the adult population that can be expected to perform the tasks successfully and to determine the extent to which various background characteristics such as educational level, race, ethnicity, gender and income level are related to the estimates of successful performance.

In the profile approach the conception of literacy is not simply viewed as a set of isolated skills associated with reading and writing, but more importantly it is seen as the application of those skills for specific purposes and specific contexts. Because of the diversity of literacy demands that people encounter at home, at work and in their communities, literacy is seen as a set of context utilized practices rather than as a single standard. Three distinct aspects of literacy are included in this approach—prose, document and quantitative.

As the reader moves into the technical test review section, each one of these three approaches to assessment is represented.

_____

What Are the Essential Skills of Today?

Another side of the assessment and testing issue is related to the types of skills required in the workplace. One factor in choosing an instrument is to consider whether or not the skills assessed on the test are the same kinds of skills being used in the job or in future jobs. This line of inquiry has been going on for some time now. Most recently the goal of the Essential Skills Research Project being conducted by Human Resources Development Canada was to find out what workers do in their jobs. Through approximately 3,000 interviews with workers in a broad range of industries across Canada, the study examined the essential skills actually required in lower skill entry level occupations. The objective of the study was to produce for each of these occupations, a profile describing the essential skills required in the Canadian workplace. These profiles

contain descriptions of how the essential skills are used in an occupation as reported by the actual job incumbents and include the more traditional basic skills, thinking skills, teamwork and skills required for continuous learning. Having a general understanding of these different categories of skills will enhance decisions around what types of instruments to use. The reader should note here that the seven tests later reviewed do not purport to measure all of the skill categories described in the following paragraphs.

For the traditional basic skills such things as the reading of textual materials, the use of documents, writing, numerary and oral communication are included. The following are examples of some of those skill categories. For textual reading some of the typical and more complex tasks are rated on a scale ranging from reading relatively short texts to locating a single piece of information or following simple written instructions to the more complex task of interpreting dense and complex texts.

In the skill category—use of documents—there are typical and complex tasks which involve reading. These are rated on a scale of difficulty from reading signs, labels and other table like texts which require only identification of key words to making high level text-based inferences and using specialized knowledge. For the skill of writing there are typical and more complex tasks also rated on a scale with level 1 skills including such things as preparing written materials in preset formats requiring very brief text to the more complex task of preparing written materials on a diverse range of topics to persuade or present an analysis, comparison, evaluation or critique.

For numeracy, tasks are categorized by application setting: money math, scheduling or budgeting math, measurement in calculation math or data analysis math. Numeracy is further described by specific math skills such as whole numbers, fractions, proportions, measurement conversions and geometry. The skill of oral communication is described by type and purpose such as listening, speaking, interacting with co-workers, interacting with those who supervise, interacting with customers, participating in group discussions, presenting information to large groups and leading formal discussions.

The second major category of skills is called thinking skills with one of the first sets of sub-skills related to problem solving. Problem solving skills are also rated and include such factors as complexity of the problem, complexity of identifying the problem, complexity of identifying the solution steps and complexity of assessing the situation. Related to problem solving is decision making. Typical and most complex tasks are again rated on a point scale and include such factors as whether there is a set of procedures or decision tree to follow, the consequence of error and the extent to which judgement is

required to make an appropriate decision. The next set of sub-skills under thinking skills are planning and organizing one's own job tasks. The planning and organizing of job tasks is rated on a scale and includes such factors as the extent of variety in work activities to the extent to which the worker's own work plan must be integrated with the work plans of others. The next category is use of memory. The content is described by type, for example remembering information for brief periods of time, remembering unique events and remembering job specific information for long-term use.

The third major category of skills is working with others. Content is described by the work context in which the job is performed, for example working alone, independently, with a partner or a helper or in a team, and by participation in leadership or supervisory activities such as orientating new employees, assigning tasks to other workers, selecting contractors or suppliers or participating in formal group discussions about work processes. Computer skills is the next category. The use of computers is rated on a five-point scale ranging from "uses computers but no knowledge of software required" to "assesses needs and designs computer systems".

In summary, essential skills are those skills that are found in virtually all occupational areas, though their specific form or level of complexity varies. These same skills provide the foundation for learning other more specific skills and enable an individual to perform the tasks required in the job.

_____

The Nature of Testing

A common concern for many practitioners trying to make sound decisions about assessment is that the technical terminology of testing is cumbersome. Although reported in other sources, this part of the discussion paper reviews some of the terms related to testing. For example standardization, norms, absolute standard and competency-based testing are all terms used in the technical review section of the report. A standardized test is a test that is administered under standard conditions to obtain a sample of learner behavior that can be used to make inferences about the learner's ability. A standardized test differs from an informal test in that the latter does not follow a fixed set of conditions. For example in a standardized achievement test, the same reading materials are read by different learners following the same procedures, answering the same types of questions and observing the same time limits.

There are two types of standardized tests: norm referenced and criterion referenced. Many standardized tests have been adopted to permit a learner score to be interpreted in relation to or in reference to the scores of other people who have taken the test. This type of test is called norm referenced. In such a case an individual's standardized test score is interpreted by comparing it to how well the referenced group normally performs on the test.

The second type of standardized test is criterion referenced. This testing format is closely related to the development of self-paced, individualized program instruction. In programs following this approach, a domain of knowledge or skills is carefully defined. An absolute standard or criterion of performance is set and everyone's score is established

in relation to that standard. Learning objectives that can be assessed are specified and the units of instruction frequently called modules are developed to teach the various subsets of knowledge and skill identified by the learning objectives.

Another type of test is a competency-based test. Typically, developers first identify the important objectives or competencies to be learned. Next, test items are developed to determine whether learners already possess the competencies or if instruction is needed to develop certain competencies. Then, various commercially available materials with a variety of learning exercises are identified that teach each of the competencies so that teachers can select the materials their learners need to master. This approach is a form of teaching to the test, even though the exact contents of the assessment instrument may not appear in the curriculum to avoid directly teaching to the specific test items. In the competency-based programs, the competencies are identified first and the curriculum is designed to help the learner achieve these specific competencies.

PART 2
Tests of Adult Basic Education (TABE)
Work-Related Foundation Skills

CTB MacMillen-MacGraw Hill,
20 Ranch Road,
Monterey, California,
93940-5703
Tel.:1800-538-9547

Test Content

The Work-Related Foundation Skills (WRFS) is a series of norm referenced tests designed to measure achievement of basic skills in reading, mathematics and language commonly found in a majority of adult basic education curricula. Using appropriate content and language to represent the workplace context, WRFS assesses achievement skills and individual needs required to function in the workplace and in society.

WRFS has four forms: health, trade-technical, business-office and general. The test items, graphics and passages are presented in a context of these workplace environments. All four forms of the WRFS measure comparable skills. Because the four forms are comparable in content coverage, they can be used effectively in a pre-test / post-test situation.

Each of the four forms of WRFS contain the following subtests: reading, mathematics computation, applied mathematics and language. It take about 30 minutes to complete each subtest. The reading subtest focuses on the learner's understanding of the reading process as well as on the ability to construct meaning from the text. In the context of passages, graphics and reference sources, this subtest measures the examinee's ability to recognize words in work milieus, recall information, and recognize main ideas. The reading items focus on functional reading through the use of diagrams, charts, forms and schedules.

The mathematics computation subtest measures the basic computational skills necessary for any successful mathematical program. It measures the examinee's understanding of the operations of addition, subtraction, multiplication and division as they relate to whole numbers, decimals and fractions. Also measured is the student's understanding of money, integers and percents. The applied mathematics section of the test stresses concept formation, reasoning and problem solving in context. It measures the examinee's ability to apply mathematical concepts related to Enumeration, computation in context, number theory, measurement, data interpretation, geometry and prealgebra. Attention is given to the reasoning skills required for practical problem solving.

The language subtest measures the learner's skills in language usage, sentence formation, paragraph development and writing conventions such as capitalization and punctuation. The subtest emphasizes skills essential to the achievement of full literacy and stresses the link between language and reading and writing. It also places importance in using language properly in the workplace. A variety of authentic work related context are presented as stimuli.

_____

Description of Scores

The WRFS provides norm referenced scores in the form of scale scores, grade equivalence, and percentile ranks. Additionally, it provides criterion reference information for objectives embedded in the tests. The WRFS subtest and objective structures were determined by compiling information from various state adult education curricula and by studying the outcomes and competency frameworks of adult programs throughout the United States. The WRTS standardization sample was collected in 1994 and invoked a nation wide sample of approximately 1700 adult learners. Each study participant took the TABE WRTS and Form 7 and 8.

Test Development

      The creation of the TABE WRFS followed a typical test development process which included the following seven steps: assessment design, writing, item review, tryout, item analysis, test selection and norms development.

Reliability. The Kuder-Richardson Formula 20 coefficient based on the number of correct scores of the WRFS range from .82 - .91. The higher the KR 20 coefficient, the greater the internal consistency of the subtest level. Internal consistency indicates that each subtest level as a whole measures a single trait, such as reading ability rather than a combination of two or more abilities.

_____

Validity. Content related validation in achievement tests is evident by a correspondence between test content and instructional content. To ensure such correspondence, the test developers conducted a comprehensive curriculum review and met with educational experts to determine common educational goals and the knowledge and skills emphasized in today's curricula. This information was used to guide all phases of the design and development of this work related assessment tool.

Construct Validity. Construct validity which refers to the meaning of test scores and the kind of inferences they support, is a central concept underlining the work related assessment test validation process. Evidence for construct validity is comprehensive and integrates evidence from both content and criteria related validity. For example, to demonstrate comprehensiveness, the work related assessment tool must contain items that represent essential instructional objectives. Additionally, patterns of inter-correlations among the work related assessments and between them and other instruments in the TABE family of assessments, demonstrate convergent and divergent validity. That is, tests which are designed to measure similar skills correlate more highly than tests designed to measure distinctly different skills. For example, the WRFS language subtest correlates highly with the reading subtest, as expected, while its correlation with the WRFS mathematics subtests are lower.

Tests of Applied Literacy Skills

Educational Testing Service,
Language Learning and Assessment,
Rosedale Road,
Mail stop 05-9,
Princeton, New Jersey 08541,
Telephone: 1-800-551-1230

Test Content

The Tests of Applied Literacy Skills (TALS) measures three distinct and important aspects of literacy—prose, document and quantitative. Each of these three subtests is available in two forms. Prose literacy tasks involve the knowledge and skills needed to understand and use information from texts that include editorials, news stories, poems and the like. The two forms of the Prose Literacy Test contain a total of 48 tasks. The materials in the test are mostly expository, which means that they describe or define. The texts used in the test are reprinted in their entirety and replicate the layout and typography of the original sources. Three aspects of prose literacy are represented in this subtest: locating, integrating and generating information. Tasks from each of these three areas and the three subtests extend over a range of difficulty. A twenty minute time limit is specified for each of the two sections of each test form.

Document literacy tasks involve the knowledge and skills required to locate and use information contained in job applications, payroll forms, transportation schedules, maps, tables, indexes and so forth. The two forms of the Document Literacy Test contain a total of 52 tasks. Many of the tasks are a necessary part of meeting the requirements of a job and managing a household. These skills involve strategies needed to locate information in various complex arrays such as graphs and charts and to transfer this information from one document to another. Three types of questions or directives are used in the document scale: locating, cycling and integrating.

Quantitative literacy tasks involve the knowledge and skills needed to apply arithmetic operations, either alone or sequentially to numbers that are embedded in print materials such as balancing a checkbook, figuring out a trip, completing an order form or determining the amount of interest from a loan investment. There are 46 tasks in both forms of the Quantitative Literacy Test. The respondent must perform addition, subtraction, multiplication and division using numbers that are buried in printed materials encountered in everyday practical situations. The difficulty of the tasks are a function of the particular arithmetic operations called for, the extent the number is embedded in the material and the extent to which an inference must be made to identify the type of operation to perform.

Description of Scores

The first step in interpreting the test results is to calculate the raw scores. The raw scores are the total number of right answers from each section of a particular form and are obtained by scoring the questions on a right/wrong basis using the guides in the test manual. The second step is to convert raw scores to proficiency scores using the tables in the manuals for each form of the three literacy tests. To obtain an individual's proficiency score the test administrator selects the appropriate table of the form of the test and locates the number correct for each of the two sections of the test. The number in the table that is defined by the intersection of the appropriate section rows is the proficiency score. This score provides an estimate of an individual's demonstrated skill within each of the literacy domains.

With a proficiency score in hand, the test administrator uses the interpretive framework in the manual which characterizes selected tasks at various levels of difficulty along each of the literacy scales. The scales are designed to range from 0-500, however the majority of tasks fall between 200 and 400 on each scale. To understand what each test measures, the tasks are interpreted in light of the interactions between the stimuli and the questions or directive that can account for placement of tasks along each of the scales.

The interpretive framework for the Prose Literacy Test focuses on locating information in the text, integrating information from text and generating information from text. Similarly, the interpretive framework for the Document Literacy Test includes locating information in documents, cycling through information in documents, and integrating information in documents. Conditions and levels of difficulty for the Quantitative Literacy Test center on addition, subtraction, multiplication and division tasks and combination tasks.

At the student level, the TALS can serve a variety of needs. Each of the tests provide an estimate of an individual's demonstrated proficiency in applying a set of literacy skills to selected materials frequently encountered in work, home and community contexts. The proficiency score levels provide a useful indication of the range of tasks that an individual can be expected to perform with a high level of consistency. As well, the test scores can indicate the most appropriate materials to be used in a program. In addition, the test scores provide learners with planning information and can assist in estimating individual learner growth and progress.

Test Development

The underlying principle driving the development of the simulation tasks was that demonstrated performance on any given task reflects the interaction among the structure of the stimulus material, the content represented and the nature of what the individual is asked to do with the material. Using Item Response Theory (IRT) in developing the TALS allowed for the estimation of the difficulty level of a particular exercise relative to the difficulties of all exercises in the pool, as well as the estimation of an individual's proficiency level in the assessed area. Four procedures were followed to establish the conversion tables that transform the raw score to the IRT estimated scale score: a) the establishment of IRT scales and estimation of national distribution from the 1985 Young Adult Literacy Survey b) the establishment of separate IRT scales with the TALS items c) the linking of the TALS scales to the Young Adult Literacy Survey scales via linking samples and d) the production of score conversion tables for the TALS.

Scale and Test Properties

Response probabilities. A particular benefit to the tests as a result of the application of an IRT model was the availability of item information that allows for the description of various proficiency levels over the range of each scale. This means it is possible to identify tasks that an individual scoring at a particular level can be expected with high probability to perform successfully, while a person scoring at the next lower score has a much lower probability of success.

Inter-scale relationships. Estimated correlations among the three literacy skills were .59 (Prose and Document), .60 (Prose and Quantitative) and .62 (Document and Quantitative).

Differential group performance. A deliberate attempt was made to select stimuli across a broad range of adult interests and contexts so that highly specialized background knowledge would not provide an advantage to any one group over another. As well, questions were developed to tap skills relevant to the types and levels of processing associated with the three literacy domains as reflected and practiced in everyday situations.

Basic English Skills Test

Centre for Applied Linguistics,
1118 22nd Street, N.W.,
Washington, DC 20037,
Tel.: 202-429-9292

Test Content

The Basic English Skills Test (BEST) was developed to measure listening comprehension, speaking, reading and writing skills at a basic and intermediate level. It provides information on the attainment of basic functional language skills and consists of two sections: an oral interview section and a literacy skills section.

The oral interview section is made up of 49 items divided into personal identification, home environment, time, directions, shopping and money, signs, doctors office, accidents, employment and data forms. A reading passage and a writing task are also included. These two tasks serve as a screening device to identify examinees for whom the literacy skills section may be appropriate. All questions in this section of the test are a series of simulated real life listening, comprehension and speaking tasks. This part of the test is administered individually in a face to face interview and takes about fifteen minutes to complete.

The literacy skills section contains about fifty items categorized under personal background, calendar, food and clothing labels, rent cheque, envelopes, telephone directory, train schedule, signs, a reading passage and a writing question. This part of the test can be administered either individually or on a group basis and takes about an hour.

Description of Scores

       Each item in the oral interview section of the BEST is designed to test one of four skills areas of listening comprehension, communication, fluency and reading and writing. Each of these four areas is also represented by a symbol on the test booklet. For each item of each skill area a range of scores is provided along with a summary of the oral interview scoring rules. This range includes values from 0-3 with a brief description and relevant examples of what is meant by the numerical weight. After scores are totalled using a scoring sheet, the test administrator derives an oral interview total score. Examinees who obtain a score of five or above can be given the literacy skills test.

       A literacy skills scoring sheet is used for the second part of the test. Scoring is based on a scale of 1 or O points for the eleven parts of the test. The test manual provides scoring instructions and relevant examples. Total scores for all questions are written on the scoring sheet. Using a competency based approach seven student performance levels are related to the different ranges of scores for both sections of the test.

Information from both sections of the test can be used for several purposes. Placement testing or entrance testing evaluates the extent and nature of the students language proficiency on entering an English language training course. Placing students according to BEST test scores can result in more or less homogeneous classes. The test can also be used to measure the progress of students in developing functional English language proficiency. It can determine progress made in a particular course of instruction provided the course is designed to teach survival and pre-employment language skills. A third use of the test can be for diagnostic purposes. The BEST is designed to reflect as closely as possible the actual language use situation with which students need to cope in independent daily living. An item by item review of a student performance on the test can provide important information on the student's ability or lack of ability to perform particular language use tasks. A final purpose of the test is for screening such as assessing the readiness for a particular course of study or training program.

Test Development

Content specifications for the test focused on topics identified as crucial to survival level competency in English. The total number of students who participated in the US field testing of forms B C and D was 987 for the oral interview section and 632 for the literacy skills section. Selection of the items for inclusion in the operational form of the test was based primarily on the statistical results of an item analysis. Level of difficulty and discrimination (r-biserial coefficients) were examined for each test item. Results showed that very few field test items needed to be eliminated. Internal consistency reliability estimates for sub-scores of the three forms range from .77 - .83 and for total scores .96 - .97. These scores are quite high for a test composed mostly of free response items. Inter rater reliability or how different raters score the same examinee indicated that there was a high degree of consistency across raters.

The BEST demonstrated high face validity since its content is quite similar to real life language use tasks it aims to test. Some evidence is provided on construct validity when test scores were correlated with the pre-assigned group rating of the examinees' overall language proficiency. In terms of inter-subscale correlations, each subscale of the oral interview section of the test shows substantial positive correlation with the total section score. As well there is a high correlation between reading and writing scores in all three forms.

Canadian Adult Achievement Test

The Psychological Corporation,
Harcort Brace and Company,
555 Homer Avenue,
Toronto, ON M8Z4 X6,
(1 -800-387-7278)

Test Content

The Canadian Adult Achievement Test (CAAT) is designed to measure an adult's level of educational achievement. This battery includes subtest in Vocabulary, Reading Comprehension, Spelling, Number Operation, Problem Solving, Mechanical Reasoning, Language, Science, and Study Skills. There are four levels of CAAT: level A is for adults who have had one to three years of formal education; level B is for adults who have had four to six years of formal education; level C is for adults who have had at least seven to ten years of formal education and may or may not have finished high school and level D is for adults who have had eleven to twelve or more years of formal education and have probably finished high school. Not all levels include all subtests.

The CAAT assesses and describes learners' generic skills in such areas as phonics, reading school related text, identifying misspelled words, drawing inferences from different types of reading material, addition and subtraction or consumer related problem solving abilities. To give the reader a sense of the content of these various subtests, a few are described below.

The vocabulary subtest is designed to assess the knowledge and understanding of words that are frequently encountered by adults in their work or daily activities. Thirty-four multiple choice items include words sampled from applied or general vocabulary, from vocabulary of the physical and natural sciences and from vocabulary of the social sciences. The problem solving subtest consists of twenty-one problems and measures an examinee's ability to determine an outcome, to record and retrieve information, and to use geometric concepts to measure. Using whole numbers, decimals, percents and fractions, the learner is requested to solve problems an adult is likely to encounter. The language subtest is taken from the Stanford Test of Academic Skills and is designed to measure a learner's functional knowledge and effective use of the English language. This subtest is made up of fifty-four items and is organized into four parts: reference skills, language sensitivity, conventions and paragraph arrangements.

Description of Scores

Two main kinds of scores can be obtained from the CAAT series: content referenced and norm referenced scores. An important score under the first category is the objective raw score. Average (mean) raw scores for all CAAT objectives at each level have been determined for the sample taking each level of the test as part of the CAAT research program. The performance of individual learners or groups of learners on each CAAT's objective can be evaluated in relation to those average scores for the purpose of identifying relative strengths and weaknesses.

Norm referenced scores are used to compare a learner's performance to that of an appropriate reference group. Grade equivalence and scale scores are types of norm referenced scores that can be used for this purpose when CAAT has been administered. Grade equivalence relate learners' scores to the typical performance of a number of groups of adults in specified grade tested in a given month of the school year. The CAAT grade equivalence ranges from K.0 to 12.9 with grade equivalence higher than 12.9 designated as post secondary. Despite the limitations, grade equivalence scores can provide an indication as to the approximate level at which adult instruction can be initiated.

CAAT scaled scores express performance on subtests and totals that extend across levels. For example, since the Reading Comprehension subtest appears in all three levels of CAAT, scaled scores for this subtest are comparable from level A to level B. to level C to level D. The continued nature of the scale and its equal interval property makes scale scores particularly suitable for studying change over an instructional period, regardless of the level of CAAT that was taken before and after instruction.

_____

Test Development

Because CAAT is designed to assess the educational achievement of adults, it was considered necessary that adults, not school children be used in the standardization program. However, despite the fact that much work is being done at the present time concerning the measurement of adult achievement, there are still no suitable criteria for defining the population of adults across Canada for whom the CAAT would be appropriate. In the absence of such criteria, research was conducted with a number of adult groups whose characteristics would define the users of this type of instrument. The sample was drawn from lists of adult educational and vocational programs in communities and schools and from mailing lists of provincial and regional and penal institutions. This research yielded information about the appropriateness of the items and objectives, the difficulty and sensitivity of the item for the adult group and the effectiveness of the options for all multiple choice questions.

Reliability. The reliability of a test is defined as the extent to which the test yields

consistent results. The Kuder-Richardson Formula 20 was used to determine the

reliability of CAAT's scores. Reliability estimates range from .84 - .95 for level A,B, and

C.

Validity. The content validity of the test can be evaluated through the examination of the

test content to determine whether it measures the fundamental skills that are necessary for

an adult to function in today's society. Since the objectives were written to reflect the

content of adult education programs throughout Canada, CAATis valid for measuring

these common objectives. However, individual programs must determine the validity of

CAAT for measuring their own objectives. Evidence also indicated that inter-correlations

of the CAAT are positively correlated with one another.

Measurement, Reading and Arithmetic

Ramsay Corporation,
Boyce Station Offices,
1050 Boyce Road,
Pittsburg, PA 15241-3907,
Tel.: 412-257-0732

Test Content

This criterion referenced test is made up of three basic skills in the areas of measurement, reading and arithmetic. The measurement of wholes, halves, quarters, eighths, and sixteenths comprise the first portion of the test intended for use where trainees must be able to measure accurately. Twenty items in a multiple format make up this section. The reading portion of the test consists of five paragraphs which are progressively more difficult in readability level. Each of the test passages is followed by eight multiple choice questions. The readability estimate range is from 8.0 for the instructions to 12.0 for the final passage. The average readability of the test is 10.6.

The job related passages are designed to assess ability to read job related information found in an industrial environment in the following areas: hydraulic fluids, bench tools, hydraulics, lubrication and solenoids. The arithmetic portion of the test was developed to evaluate the ability to perform basic computations of addition, subtraction, multiplication and division of whole numbers, decimals and fractions. It also contains simple formulas and charts. Forty multiple choice items make up this part of the test. Each part of the test has a time limit. Measurement is timed at 15 minutes; reading takes 40 minutes to complete and the arithmetic subtest in 30 minutes.

Description of Scores

For this criterion referenced test, an absolute standard or criterion of performance is set and everyone's score is established in relation to that standard. Thus, 90 percent and above correct might be necessary to receive a grade of A, 80 to 89 percent correct for a B and so forth. Percentile equivalents for the three portions of test have been conducted on a sample of 1,441 metal workers.

Test Development

Reliability studies have been conducted on other tests from the Ramsey Corporation such as the Technician Electrical Test, the Technician Mechanical Test, the Building Maintenance Test, the Electronics Test and the Electrical Maintenance Trainee Test but not this newly developed one. In terms of the basic skills test validation procedures, seven steps have been taken. First, the client provides a company job description which is then placed into the Ramsay Corporation format of job summary, work performed and consequences of errors. Ten subject experts add to the description to ensure that it is up to date and reflects the job activities. Then the ten subject experts rate the job with respect to importance of job activities and percent of time spent on job activities. The inter rater agreement is then evaluated. The third step is the evaluation of the readability levels, mathematical skills and other requirements of the job and the training program. The testing company then evaluates the job analysis information and determines the knowledge, skills and ability and the methods for measurement. The fifth step entails editing, composing and printing the test. They are brought to a group of at least four subject experts for review and matching with knowledge, skills and ability. The Angoff procedure to set cutting scores and review the final test are used. This means that

5 to 10 subject experts review the test item by item, select a correct answer, are told the keyed answer and are asked what percent of qualified candidates would pass each item. Difficulty and point biserial discrimination indices have been calculated. The last step includes the writing of a content validation report in the format suggested by the Uniform Guideline on Employee Selection Procedures. This also means that the company reviews a draft and the Primal report detailing the activities undertaken is bound and provided to the client.

Wide Range Achievement Test 3

Test Content

The Wide Range Achievement Test 3 (WRAT3) is designed to measure the codes which are needed to learn the basic skills of reading, spelling and arithmetic. This recent addition of the test can be used with adults ranging in ages from 17 to 75. Each of the two alternate test forms provide three subtests. The first subtest is reading which consists of a letter reading section and a word reading section. In the letter reading section the respondent is asked to name 15 letters of the alphabet and in the word reading section he is asked to pronounce 42 words. Both parts of the reading subtest are to be administered individually. For adults, the word reading section is administered first. It is a word recognition test and not a test of speech or diction. If a more comprehensive reading test is required both test forms can be administered. One point is given for each correct letter and word. Raw scores are transferred to a section on the test form.

For the second subtest spelling, respondents write letters of the alphabet and words to dictation. There are 55 items on both forms of the spelling test. These consist of 15 items on the letter writing section and 40 items on the word spelling section. For adults, the word spelling section is administered first. A pronunciation guide is provided for the test administrator with the list of spelling words on plastic cards. One point is given for each of the correctly written letters and spelled words. Total spelling raw scores are then transferred to a section on the test form.

The arithmetic subtest consists of counting, reading number symbols, solving oral problems and performing written computations. It is also broken down into an oral arithmetic section with 15 items and a written arithmetic section with 40 items. Adults begin with the written arithmetic section. Each item is counted as one point and the total arithmetic raw score is transferred to a section on the test form. Each form of the WRAT3 takes 15 to 30 minutes to complete.

Description of Scores

There are several types of scores used to report the WRAT3 results. Four kinds of scores will be highlighted here. The raw scores are the number counts of the correct responses. They cannot give a precise measure of how much better one performance is over another. The absolute scores of an individual or group of individuals give measurement of each variable across the whole continuum of that variable without regard to grade or age. These scores are most suitable for pre- and post-testing. Standard scores are the types of scores used for comparisons within individuals and between them. Not only is it possible to tell that there is a difference between the two scores, it is appropriate to describe the amount of difference between such scores. Examples of the ratings of standard scores are: high average (110-119), average (90-109), low average (80-89). Grade scores are slightly higher level scores than raw scores. For the WRAT3, grade scores are only rough guides to grade achievement because there is as much variance within a grade as there is between grades. These scores should be used with the information from standard scores or absolute scores in describing the test results.

Test Development

The WRAT3 is an age-normed referenced test of the basic academic coding skills of word recognition, spelling and arithmetic computation. Restandardization of this edition of the test began in the early 90s as a result of the Rasch analysis of item difficulties. It was found that there were many duplications of items in bold forms of the WRAT-R. The WRAT3 was given to a norm sample in the United States. An adult sample was selected by using a national stratified sample design controlled for age, regional residence, gender, ethnicity and socio-economic level. The sample size for all age groups(5-75) was 4,834.

Reliability. When a skill is measured there should be a reasonable expectation that another measurement of the same skill will give similar results. The extent that an instrument can make consistent measurement of a trait is described by the reliability of the test. Three of the many measures of reliability are reported here. The median test coefficient alphas range from .85 - .95 over the different test forms. This is considered desirable. Correlation coefficients for the raw scores of the three subtests are also very high—.98, .98 and .98 respectively. Using the test-retest method corrected stability coefficients range from.91 to.98.

Content validity. The Rasch statistics of item separation was used to indicate how well the items selected were representative of the three domain areas: written decoding, written encoding and arithmetic computation. For each of the subtests the highest item separation score possible 1.00 was found which provides strong evidence that there is content validity on each of the three variables.

Construct validity. Evidence regarding the developmental nature of the WRAT3 shows a steady increase in mean scores across the norm age groups until the 45-54 year old age groups. From here the scores begin to fall. IJsing the data from the total norm sample, there is a significant correlation between each WRAT3 test raw score and the age and month of the test taker. Evidence also indicated that inter-correlations of the WRAT3 are positively correlated with one another. The range of correlation for the reading-spelling comparison is .81 -.91; reading-arithmetic .54 - .78; and spelling-arithmetic .58 - .82. The intercorrelations show a moderate to high positive relationship among the variables.

Metric Skills Assessment

Lakeshore Technical College,
Gary Heeler,
1290 North Avenue,
Cleveland WI 53015,
Tel.: 414-458-4183

Test Content

The Metric Skills Assessment is a competency-based test designed to measure a predeveloped series of metric learning experiences. The test which consists of 35 items is divided into two sections. The first section covers areas of metric origin, basic units of measure, definitions, and numerical conversions. The second section is a series of questions asking the test taker to record measurements provided by different job-related instruments such as the micrometer and the vernier caliper. The test can be administered individually or to groups of adult learners and takes about 20 minutes. The test also can be used as a pretest/post-test training tool.

Description of Scores

The Metric Skills Assessment is a type of criterion-referenced tool which means that learner gain is interpreted in terms of how many units of instruction are mastered at the prescribed criterion level. The test can be administered as a training program pretest and depending on the criterion used by the instructor can direct the learner into the appropriate part of the curriculum. As a post-test, the predetermined criterion is usually 90% correct or 80% correct.

Test Development

       The test developer first identified what important competencies should be learned through a panel of mathematical experts. Then various commercially available curriculum materials with a variety of learning exercises were identified that teach each of the competencies. In other words, the competencies were first identified and a curriculum was designed to help the learner achieve those competencies. The curriculum is divided into six key areas. In the first module the everyday units of the metric system are introduced with a review of decimal operations. Traditional and metric measurement and the use of metric scales and tapes to measure items are also taught. The focus of the second module is conversion between metric and English units of measure while the third area centers on metric thread pitch and hardware identification. In the fourth module, precision measuring equipment use and identification is taught while the fifth area focuses on the use of verniers. A review of the metric system of measurement and skills reassessment concludes the program.