The following text by Dr. Stephen Thomas is an authorized reprint of *Vocational Evaluation and Assessment: Philosophy and Practice* presented as published in 1997.

Vocational Evaluation and Assessment: Philosophy and Practice CHAPTER

SIX

Instruments of Evaluation

Vocational Evaluation Tools

The generic term tools refers to the comprehensive collection of instruments, techniques, and strategies available to vocational evaluators in the routine performance of their job duties.
Instruments, the focus of this chapter, are those standardized tests, work samples, and evaluation systems used to collect objective, norm-referenced data (e.g., time and error scores) on skill and ability. The techniques covered in the next chapter refer to the criterion-referenced approaches including functional assessment, situational assessment, continuity-based assessment, curriculum-based assessment, ecological/environmental assessment, behavioral observation, and interviewing used to subjectively assess behavior, performance, and attitude toward work. The strategies covered in the chapter on vocational evaluation processes are the accommodations, modifications, and supports for learning and performance that are applied during the use of instruments and techniques to ensure an accurate assessment of potential.

Vocational evaluation also relies on a wide variety of work-related tools (e.g., mechanical tools, office tools, electronics tools) to assess an individual's current and future potential. Unlike counseling, vocational evaluation is an equipment-oriented process. Evaluators in comprehensive units rely on similar kinds of tools and equipment used by workers on their jobs to assess a consumer's work-related needs and abilities. It is this reliance on simulated and real work tools and equipment that makes vocational evaluation uniquely different from other assessment disciplines that rely primarily on file review, interviewing, psychometric testing, and career counseling. Although vocational evaluation techniques will be briefly reviewed in this chapter to illustrate their interrelationship with evaluation instruments, they will be covered in greater detail in the following chapter.

Although evaluation instruments provide the opportunity to apply techniques and strategies (e.g., behavioral observation, modification) both instruments and techniques can stand alone as assessment methods. There are times when a technique can be used to collect information that is also available through the use of instruments. For example, the assessment of learning style can be accomplished by observing how individuals best understand what to do when being administering different tests, work samples, and situational assessments that are not designed to evaluate learning style. However, there are a variety of standardized instruments that were specifically designed to identify the preferred learning style. If such an instrument is used to determine the learning style, then the outcome can be validated through the observation of applied learning on other tests, work samples, and situational assessments. When instruments

and techniques are used together, more subtle and detailed information can be collected that will provide greater insight into behaviors, interests, abilities, and needs of the consumer.

It is this highly individualized mix of instruments, techniques, and strategies within the evaluation process that make vocational evaluation a unique and creative venture for the participant and practitioner alike. The key to a successful vocational evaluation is knowing how to efficiently plan when and what instruments and techniques will be administered to meet the individual needs of different consumers and referral sources. This chapter will provide a basic overview of the widely used instruments and techniques of evaluation and assessment. The application and interpretation of these instruments, techniques, and strategies will be described in later chapters.

Recognized Instruments and Techniques

The instruments and techniques of vocational evaluation have been adapted from a variety of other professions and fields that also engage in various forms of assessment (Neff, 1985; Pruitt, 1986). Psychology, for example, contributed standardized tests, the first work samples, and the testing laboratory. Pruitt (1986, p. 6) feels the most important concept psychology has given vocational evaluation is that "information derived from evaluative methods or instruments may be used to understand current behavior and to make predictions about future adjustment." Industry and industrial psychology created job analysis, behavioral rating scales, simulated tasks, and job tryouts for work classification, and employee screening and selection.
The military has given evaluation the group testing approach and contributed to the further refinement of work samples through the development of instruments, such as flight simulators. The simulator has been applied to other fields to evaluate and train ship pilots and captains, and assess applicants for bank teller positions through computer simulations. Lastly, the rehabilitation facility in the United States can be credited with the organization and refinement of these different assessment approaches into the process known as vocational, or work, evaluation. Neff (1985, p. 180) indicates that, due to a lack of available assessment procedures, rehabilitation facilities were forced to develop their own "assessment devices, which largely fall under the work-sample and/or situational assessment categories."

Neff (1985) identified four instruments and techniques commonly used in the assessment of work potential. They include the:

- Mental testing approach,
- Job analysis approach,
- Work-sample approach, and
- Situational approach.

These are consistent with (Nadolsky, 1973, p. 51) five evaluation methods, which include: "(a) the psychological testing approach, (b) the work sample approach, (c) the situational approach, (d) the job tryout approach, and (e) the job analysis approach." With the

exception of job analysis, the literature has recognized the same fundamental tools of evaluators (Sax & Pell, 1985; Tenth Institute on Rehabilitation Services, 1972). Similarly, Lesnik (1983) identified six generic "techniques" of vocational evaluation under the umbrella of occupational exploration. These techniques, which are listed in the general order used, and lead to the goal of real work, include:

- Interviewing,
- Psychological testing,
- Work samples situational assessment,
- Job site evaluation, and
- Job tryout.

The Vocational Evaluation and Work Adjustment Association (1975) classified the tools of vocational evaluation into three categories, situations as tools, resource tools, and applied tools, with a listing of the appropriate instruments and techniques under each one.

Situations as Tools

1. On-the-Job Evaluation, consisting of:
    - Job site situation,
    - Production work situation,
    - Trial training evaluation, and
    - Simulated job stations.
2. Work Samples, consisting of:
    - Actual job samples,
    - Simulated job samples,
    - Single trait samples, and
    - Cluster trait samples.
3. Psychometrics

Resource Tools

1. Occupational information
2. Client information
3. Job analysis
4. Audio-visual materials

Applied Tools

1. Interviewing procedures
2. Observational procedures
3. Reporting procedures

Sitlington, Neubert, Begun, Lombard, and Leconte (1996) identified methods for gathering information through transition assessment, which include:

- Analysis of background information interviews
- Psychometric tests
- Work samples

- Curriculum-based assessments
- Behavioral observations
- Situational assessments
  - In vocational settings
  - In community settings (e.g., home, recreation sites, banks, and stores)
- Assessing potential environments
  - Analysis of community environments
  - Job analysis
  - Analysis of postsecondary education environments

A national study by Hayward, Wine, Thorne, Stoddard, and Wilhite (1992) reported the percentage of vocational evaluations conducted for Vocational Rehabilitation that used the following instruments and techniques (Hayward & Thomas, 1993, p. 337).

Table 1
Most Common Vocational Instruments and Techniques Used in Vocational Evaluation

Vocational Instruments and Techniques	**Percentage of Use**
Specific tests and work samples	92.9
Clinical interview	50.4
Situational assessment	30.5
Functional assessment	22.3
Other	2.9
On-the-job evaluation	4.8

Thomas (1986, pp. 150–151) found relatively similar distributions of use to the Hayward et al. (1992) study, in a national survey of 106 full-time vocational evaluators in public, private, and school-to-work settings, who were members of VEWAA. Psychometric tests and work samples were listed separately, and there was a much higher reported use of interviewing, situational assessment, and job site (on-the-job) evaluation.

Table 2
Most Common Vocational Instruments and Techniques Used Among VEWAA Members in Vocational Evaluation

Vocational Instruments and Techniques	Percentage of Use
Psychometric/standardized testing	100.0
Work samples and systems	96.0
Interviewing	95.0
Situational assessment	67.0
Job site evaluation	30.0
Other	16.0

In the early years of vocational evaluation, work samples were the instruments of choice. Over time, as evaluation became shorter, there was greater reliance on quicker and cheaper psychometric tests. Today, however, newer work sample and evaluation systems have been significantly shortened, increasing their frequency of use. Situational assessments, which rely on behavioral observation, are performed in-house or in the community and take considerable time to set up and administer, as do on-the-job evaluations (OJE) that rely on consumer placement in community-based jobs. As a result, situational assessment and OJE are used less frequently. The "Other" category at 12.9 % in the Hayward et al. (1992) study and 16 % in the Thomas (1986) study represent the range of creative activities employed by evaluators in assessing potential.

More current instruments and techniques used in evaluation and not mentioned above include computers for assessment, occupational information, job search, and report writing; functional capacity assessment; training analysis; and, checklists and rating scales used in areas, such as functional assessment and ecological (environmental) assessment. Variations of recognized instruments and techniques are also identified in the *CARF Standards Manual* (1996), and the *CCWAVES Standards and Procedures Manual* (1996). Although the administration and interpretation of various standardized instruments will be presented later in the book, brief definitions and descriptions will be provided in this chapter to give the reader a basic familiarity with selected evaluation instruments.

Considerations in Choosing Appropriate Instruments

There is a well-defined hierarchy of vocational evaluation instruments (Cutler & Ramm, 1992; Thomas, 1991). It is based on the relationship of the instrument to real and simulated work—a fundamental consideration in vocational evaluation. As illustrated below, the most commonly used assessment instrument is the psychometric test. It is the quickest and most cost- effective means of obtaining information. Because psychometric tests are abstract in design, often require reading, and frequently have time limits, they look the least like work of all the evaluation instruments. Therefore, they may create

testing anxiety in people who do not perform well on standardized tests, and the results may not adequately represent current ability or future potential. Since they rely on a comparison of the individual being tested to a group of individuals in a norm table, they are often referred to as "norm-referenced" instruments or procedures.

Work samples (which are simulations of work) are initially more expensive to buy than psychometric tests and take longer to administer. As a result, they are not given as frequently but are more appropriate for the assessment of work-related performance, behavior, and manifest interest than psychometric tests. Situational assessments and community-based assessments (on- the-job evaluations) take the longest time to administer because of their focus on work-related behavior and are the least used evaluation techniques. Community-based assessments, in particular, have the highest relationship to work and rely on a "criterion-referenced" interpretation (i.e., how the individual performed each job task), rather than a norm-referenced interpretation. Situational assessments and community-based assessments are particularly useful for lower functioning individuals who might benefit from supported employment placement but, by no means, are limited in their application to lower functioning populations. Since work samples share many of the same characteristics with tests and situational assessments, they provide an opportunity for both norm-referenced and criterion-referenced assessment and interpretation.

Continuum of Vocational Evaluation Instruments

The tools an evaluator uses are a personal choice; what works for one may not work as well for another. As a result, many evaluators are somewhat eclectic—that is, they prefer to use a variety of instruments and techniques that best fit their assessment style and philosophy, as well as evaluation setting. A unit must maintain a widest possible variety of instruments to relate to: changes in populations; referral source needs; differences in consumer interests and abilities; and, the variety of available community resources, training and education programs, and employment/career opportunities. Following are a number of important questions evaluators must answer when choosing a repertoire of instruments and techniques for the unit (Brown, McDaniel, Couch, & McClanahan, 1994; McDaniel & McClanahan, 1993; McFarlane, Bellinger, Paulsen, Wesolek, & Modahl, 1988; Thomas, 1991).

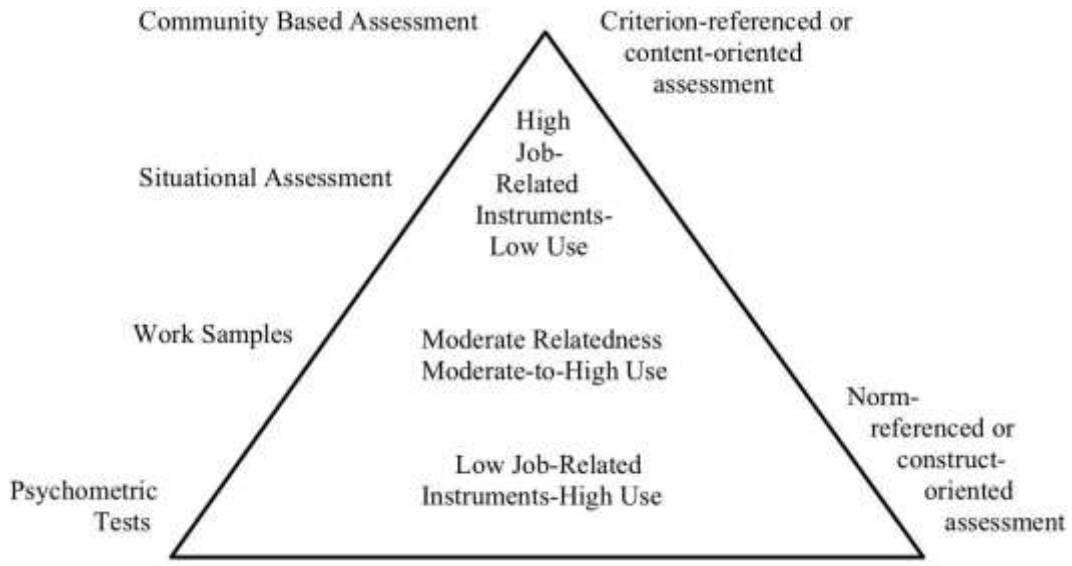

Figure 1: Continuum of Vocational Evaluation Instruments

1. What is the composition of the population being served? Although an evaluator will want to anticipate occasional variations in the type of individual served, instruments should be targeted to the typical referral. Tests and work samples should not be too easy or too difficult for participants to take. Otherwise, motivation in the evaluation will be affected. Instruments for readers and non-readers that assess for the same information (e.g., written and picture interest inventories) should be available.

2. What are the composition of the labor market, course/curriculum offerings at local schools and colleges, and community resources? There is little value in using instruments that do not represent available jobs (or job families), education, and training. In addition, evaluators need to know what community resources (e.g., remedial programs, adjustment services) and community supports (e.g., supported employment, supported living, accessible public transportation) are available to consumers to enhance learning, living, and working.

3. What are the goals and objectives of the evaluation/assessment unit? Similarly, what are the needs of consumers/students and referral sources? For example, if the goal is to assess curriculum placement for secondary special needs students, then more attention must be given to class placement rather than job placement issues. In this case, the repertoire of instruments must represent curriculum and community training opportunities first. The needs of consumers and referral sources must be consistent with the goals and objectives of the unit, and the two must be considered simultaneously when change is contemplated.

4. What is the size of the unit? In small fixed or mobile units, only instruments that are easily stored and setup when needed can be purchased. Bulky evaluation systems and work samples can be permanently set up in larger fixed and mobile units.

5. What is the length of the evaluation? Length (e.g., several hours, several days, or several weeks) dictates the number and types of instruments and techniques that can be administered. Therefore, the time necessary to give, score, and interpret lengthy work samples, entire evaluation batteries, or situational assessments may take longer than is provided for evaluation. In this case, the evaluator must rely on brief work samples and psychometric tests. The needs of the consumer and level of functioning will influence the length of the service and the types of instruments and techniques employed.

6. What is the consumer-to-evaluator ratio (i.e., will there be a group or individual administration)? Evaluators need to be fully aware of the demands on their time made by the different instruments they use. If the ratio is one to one, then evaluators can use instruments that require their undivided attention; where they must be present throughout the entire administration. If the ratio of two to one or higher, then instruments that take up less of the evaluator's time for instruction and assistance must be chosen. Although instruments designed for use with a high ratio or with groups can usually be administered individually, but instruments requiring an individual administration cannot be used with two or more evaluees at the same time. In general, instruments that allow the evaluator to give the instructions and walk away to observe behavior at a distance are often preferred.

7. What is the cost to buy, administer, score, and maintain an instrument or battery? Although some work samples and evaluation systems have a high front-end purchase cost, their durability and low maintenance may make them cheaper, in the long run, than psychometric test materials and packets that are expended with each administration. It is best to conduct long-term price comparisons based on at least two years' worth of administrations to determine which approaches are most cost-effective. Computer scored tests may be costly if charged by the person, as compared to unlimited access software. Because of the equipment orientation, vocational evaluation units will require sufficient funds to routinely purchase materials and supplies, and maintain or replace instruments. Evaluators must target how many participants will be served each year and set their budget based on the overall costs.

Other considerations for instrument choice may surface depending on the unit. Over time, personal experience will provide a better understanding of instrument needs. Following is a review of three selected instruments (i.e., psychometric tests, work

samples, evaluation systems) commonly used in vocational evaluation and assessment.

Psychometric Tests

Psychometric tests are standardized instruments (paper-and-pencil and performance- based) used primarily for counseling and planning; in this case, vocational/career counseling and planning. Psychometric tests are different than the more high level "psychological tests" (e.g., intelligence tests, personality tests, projective tests) used by licensed psychologists for clinical diagnosis. The American Psychological Association originally developed a test classification system that is used today by companies that market standardized tests to qualified users. Companies such as American Guidance Service, Consulting Psychologists Press, Psychological Assessment Resources, and The Psychological Corporation give the "user qualification level" for each test listed in their catalogs. There are three user qualification levels: A, B, and C.

User Qualification Levels.

Level A. This ranges from no qualifications for test use, other than employment with an appropriate company or organization ordering the tests, to the completion of a course in measurement, guidance, or related area. Supervised experience in test administration and interpretation is also acceptable. Tests in this category include dexterity tests traditionally used for employment screening in industry, and some self-administered and self-scored interest and aptitude tests. Generally, the range of tests available for purchase at this level is quite limited.

Level B. Depending on the company, users must have graduate training in measurement, guidance, or psychological assessment, or a Bachelor's or a Master's degree in psychology, counseling, education, or closely related field. Membership in specified professional associations or licensure/certification in appropriate areas will also qualify a user. This level contains the largest number of tests that are most frequently used by appropriately trained and qualified vocational evaluators. They include interest and work values tests, achievement and aptitude tests, and some intelligence tests used for quick screening.

Level C. Requirements in this category range from a graduate degree (a doctoral degree is preferred) in psychology, education, or closely related field; with coursework, training, and/or supervised practical experience in the administration and interpretation of clinical assessment instruments (i.e., psychological tests). Appropriate professional association membership or licensure is also acceptable for purchase.

Psychology licensure requirements for testing vary from state to state, and the ability to buy a test may not necessarily qualify someone to use it. Evaluators must check their own state regulations to determine the minimum qualifications needed to use specific kinds of psychometric and psychological tests. States generally do not limit

appropriately trained and/or certified evaluators (CVE or CRC) from administering and interpreting Level A and B psychometric tests. However, restrictions may be placed on the use, especially the interpretation, of Level C psychological tests.

Ethical Considerations in Testing

It has been this author's experience that standardized tests are frequently misused. Much of this misuse comes from a lack of knowledge of tests and measurements theory, including norms and norm groups, reliability, validity, Standard Error of Measurement, standardization in administration and scoring, and interpretation strategies. Knowledge of tests and measurements theory is just as important to the work of vocational evaluators as counseling theory is to counselors. Anyone engaged in any form of standardized testing should have, at a minimum, a course in testing that emphasizes measurement principles, ethics in testing, and a review of the different types of standardized tests. In particular, it is the violation of well-recognized ethical standards, which are generally accepted across related professional disciplines, that can create harm to the consumer, misinformation for the referral source, and potential legal trouble for the individual in charge of testing. Ignorance of ethical standards is not an acceptable defense for uninformed evaluators who are facing a hearing, grievance, or litigation resulting from inappropriate test use.

Following is a brief list of the major ethical guidelines that test users must firmly adhere to in all aspects of purchasing, storing, administering, scoring, and interpreting standardized tests. These ethical "themes" were taken from the codes of ethics of the American Counseling Association, the American Psychological Association, the Commission on Rehabilitation Counselor Certification, the Commission on Work Adjustment and Vocational Evaluation Specialists, and the Joint Committee on Testing Practices (1988).

1. Vocational evaluators must recognize the limits in competence and qualifications they have on using certain tests. They should also understand the purposes and limits of the tests they are using and know how they will benefit or potentially harm the consumer.

2. Evaluators must be sensitive to the impact that disability, socio-economic status, education, age, gender, race, and culture have on choosing, administering, and interpreting standardized tests. Many evaluation units will have a variety of tests that assess the same areas (e.g., mechanical reasoning) for readers and non-readers, and for individuals who approach learning and processing of information differently. The goal is to eliminate any adverse impact in testing and level the "playing field" (i.e., give all examinees the same unbiased opportunity to demonstrate their best performance). Adverse impact (as with differential prediction) is where one group performs better on a test than another but with no appreciable difference in the performance between the two groups on the outcome (e.g., job or classroom performance). Cross-cultural issues must be

considered by the evaluator. Be attentive to how a test is designed to handle variations in motivation, working speed, language facility, experiential background, and any bias in response to its content by individuals taking it (Alston & McCowan, 1994; Colyer & Smith, 1993; Joint Committee on Testing Practices, 1988; Prediger, 1993; Smart & Smart, 1993; Suzuki, Meller, & Ponterotto, 1996). Understand that the individual being evaluated is a "cultural entity" (Feist-Price, Harley, & Alston, 1996). Review the manual to determine if there are representative samples of minorities, women, individuals with disabilities, and individuals from a wide age range in the normative sample, and in the reliability and validity studies as well. Determine if studies of equity are reported in the manual or literature on the test in question. Review test content to determine if the wording is free of stereotypes and cultural bias. During test orientation and administration provide appropriate accommodations, when necessary, to minimize language, processing, and time barriers, and describe the accommodations made when reporting test results. Test developers and publishers are becoming more sensitive to the need to minimize bias in testing related to age, gender, race, culture, and disability. In the future, existing tests will be revised, and new tests developed that can be accurately used across an inclusive range of groups and environments.

3. To protect the confidentiality of tests, evaluators must maintain all unused test materials in a secure place. All used tests must be maintained in consumers' files and also stored in a secure place. It is the responsibility of the professional using the test to safeguard the materials. In particular, tests should not be given or mailed to consumers to take at home unless the test is designed for that purpose.

4. Participants in evaluation and assessment must ensure that informed consent is obtained before tests can be administered. The evaluee must be notified of and agree to three things: (a) the purpose of the testing program; (b) the kinds of information being sought; and, (c) what will be done with the information obtained. Some referral sources, such as Vocational Rehabilitation state agencies, workers compensation rehabilitation companies, and welfare-to-work programs, will have blanket consent forms signed before evaluation is provided. However, this does not circumvent the evaluator's responsibility to cover these three facts with the participant during the orientation phase. School systems often require that a separate consent form be completed for every service including vocational assessment. Vocational evaluators should not release evaluation reports to individuals who were not identified as recipients of the report; nor should they share any test results with unauthorized individuals. Since the report is considered the property of the referral source and the consumer, anyone else requesting a copy should be directed to contact the referral source. However, if a subpoena is issued for report or test information, the evaluator should turn over only that information that is requested in the subpoena, excluding actual copies of the test. Test score forms and profiles can be attached to reports, but as indicated in the

previous ethical guideline (see 3.), not the actual test itself. This will compromise the test's confidentiality if the report is subpoenaed, in which case it will end up in public court records, or if a copy of the report and attachments are given to the consumer or consumer's family, it will be released into the public domain. Since tests are protected by copyright, their unauthorized public distribution is further restricted, and if attorneys insist on receiving a copy, refer them to the publishing/marketing company.

5. Strict adherence should be paid to administering, scoring, and interpreting the test as specified in the manual. Variation in the instructions and interpretive guidelines set down in the manual will negatively affect the accuracy and utility of the results. Reasonable accommodations are allowed in administration and test performance; however, there is no set rule or formula that can predict how the modification will affect the validity of the instrument. In this case, criterion-referenced procedures will take precedence over norm- referenced procedures when scoring and interpreting the instrument. Modifications in standardized testing are appropriate when it is found that the test is unsuitable for use as is, and when other tests that measure the same trait (that would eliminate the need for modification) are not available. Modification is appropriate as long as a description of why and how the test was modified is included in an oral and written review of the test results. Although some professionals argue against the modification of standardized tests, this author would sooner be in violation of tests and measurement principles than to be in violation of the ADA. A detailed description of specific modification procedures will be presented later in the book.
6. Since most standardized tests today employ a norm-referenced approach to scoring and interpretation, use of appropriate norm groups is essential. When interpreting performance, take into account any major differences between the norm groups and the individual taking the test (Joint Committee on Testing Practices, 1988; Prediger, 1993). In reference to "substantial limitations to the activities of living and working," the ADA indicates that comparisons should be made to the general population (for living activities) and to the working population (for work activities; Thomas, Hiltenbrand, & Tibbs, 1997). Therefore, general population norms, applicant norms, and job trainee or worker norms should be used when available. If an individual wants to go to school, then the use of appropriate school norms would be recommended. "*The Civil Rights Act of 1991* (P.L. 102–166)" specifically addresses the issue of norms through *Section 106 Prohibition Against Discriminatory Use of Test Scores*. An amendment to the section states that:

> It shall be an unlawful employment practice for the respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment

related tests on the basis of race, color, religion, sex, or national origin.

7. The testing environment should be quiet, comfortable, and conducive to optimizing performance. Every attempt should be made before and throughout the evaluation to minimize testing anxiety and ensure the best possible performance of the participant.

8. Provide an accurate, understandable interpretation of the results and relate them to the purposes of the test and overall evaluation (e.g., employment and training, goals and needs). This interpretation may be offered to consumers and their families, to the referral source, and to other involved professionals through written and oral communication. Scores should be considered as approximations since no percentile score is an absolute representation of performance. The evaluator may also want to use more than one norm group for a broader comparison of performance to other environments and populations. Scores from one test, or work sample, will have less utility than scores and observations from a variety of different evaluation instruments and techniques (e.g., work samples, situational assessments, job or classroom tryouts). Comparison of scores to other evaluation information from sources, such as file review, interviews, staffing, behavioral observations, job analysis, and occupational information, will increase accuracy in decision-making. This broad interpretive approach, which uses multiple sources of information, will help the evaluator account for a significant number of variables that affect the outcome.

9. Update test versions when they become available. If a publisher continues to support an older version of a test, it can be used until the forms are no longer stocked. Outdated tests may result in outdated outcomes.

The codes of ethics of professional associations, and licensure and certification bodies that represent testing disciplines, provide specific standards for the appropriate choice, administration, scoring, interpretation, and safe keeping of standardized tests. Federal regulations available through the Equal Employment Opportunity Commission (EEOC) also provide guidelines for employment screening and testing. The following two ADA regulations, administered by the EEOC, caution professionals on how tests should be chosen and used with individuals with disabilities.

It is unlawful for a covered entity [employer] to use qualification standards, employment tests or other selection criteria that screen out or tend to screen out an individual with a disability or a class of individuals with disabilities, on the basis of disability, unless the standard, test or other selection criteria, as used by the covered entity, is shown to be job- related for the position in question and is consistent with business necessity ("Americans with Disabilities Act of 1990," 1991a).

It is unlawful for a covered entity to fail to select and administer tests concerning employment in the most effective manner to ensure that, when a test is administered to a job applicant or employee who has a disability that impairs sensory, manual or speaking skills, the test results accurately reflect the skills, aptitude, or whatever other factor of the applicant or employee that the test purports to measure, rather than reflecting the impaired sensory, manual, or speaking skills of such employee or applicant (except where such skills are the factors that the test purports to measure; "Americans with Disabilities Act of 1990," 1991a)

These awkwardly worded regulations relate more to testing performed in employment settings than to rehabilitation or transition settings. However, the regulations have general application to vocational evaluation as well since tests are often used to determine employment potential and placement. In short, the two regulations stress the need to carefully choose and use tests and work samples that evaluate individuals' abilities rather than their disabilities. Keep in mind that psychometric tests should not be used in evaluation and assessment as diagnostic instruments but to provide direction for vocational/career counseling and planning—with the ultimate goal of achieving satisfying and meaningful employment for the consumer.

Review of Standardized Tests

Eight different categories of Level A and B standardized tests commonly used in vocational evaluation and assessment will be reviewed. These categories include achievement, aptitude, basic skills, dexterity, intelligence, interest, learning style, and temperament and work values tests.

Achievement Tests. VEWAA's *Glossary of Terminology* (Dowd, 1993, p. 1) defines an achievement test as one "that measures the extent to which a person has 'achieved' something, acquired certain information, or mastered certain skills—usually as a result of planned instruction or training." Whereas professional certification and licensure examinations are considered to be achievement tests, the focus for evaluators is on assessing the more fundamental and traditional skills of reading, spelling, and mathematics. A few achievement tests may also include vocabulary or information subtests. Most comprehensive achievement tests are available at different levels (e.g., by school grade, age, or functional level), have time limits, and take several hours to administer. Results are reported using a combination of percentile scores, stanine scores, standard scores (which can be used to compare results to IQ scores), and/or grade level scores (e.g., 3rd Grade, 7th Grade, post high school). Although grade level scores are routinely requested by counselors and reported by evaluators, they cannot be used to draw direct comparisons to grade levels in local school systems.

Formats for different reading subtests often consist of word recognition, vocabulary, or comprehension questions, with some comprehensive tests using a combination of subtests.
Mathematical questions consist of math problems and/or word problems and

require the examinee to write down the answer or choose the correct one from a list of four or five possible answers (forced-choice format). Spelling subtests (which are frequently omitted by evaluators with limited time) can be dictated to the examinee, or the correct word chosen from a list of four or five similarly spelled words. As a rule, items on the tests are arranged in an increasing order of difficulty. Examples of achievement tests include:

- Adult Basic Learning Examination (ABLE)
- Comprehensive Adult Student Assessment System (CASAS)
- Key Math—Revised
- Peabody Individual Achievement Test–Revised (PIAT-R)
- Tests of Adult Basic Education (TABE)
- Wide Range Achievement Test 3 (WRAT-3)
- Woodcock Reading Mastery Tests–Revised (WRMT-R)

Achievement tests are often given first to determine whether reading or non-reading instruments should be used. Most psychometric tests require reading at the 7th-grade level or higher, with low-reading versions around the sixth grade or less. The consumer populations of most referral sources (e.g., Vocational Rehabilitation, school-to-work and welfare-to-work transition programs) read at around the 6th-grade level. Therefore, tests must be chosen and used with caution to ensure that low reading level does not unknowingly affect performance.
Functional illiteracy in the United States is defined as reading, spelling, and math at or below the 4.9th-grade level. The "local" section of the average newspaper is written at the 6th-grade level (ranging from 4th for want ads, to 9th or 10th for the international section). Reading ability is a critical factor for success in the United States today.

Aptitude Tests. Aptitude is "a combination of abilities and other characteristics, whether native or acquired, that are indicative of an individual's ability to learn or develop proficiency in some particular area if appropriate education or training is provided" (Cronbach, 1990, p. 701; Dowd, 1993, p. 2) defines an aptitude test as "a measure intended to predict success in a job, educational program, or other practical activity." The purpose of aptitude testing in vocational evaluation is to determine an individual's potential to succeed in a particular course or job where there has been no prior exposure or experience.

A fine line between an aptitude and achievement test and a certain amount of achievement is always needed (e.g., reading) to do well on an aptitude test. The U.S. Department of Labor (1991b) incorporates the following 11 aptitudes into its occupational classification system: intelligence or general learning ability (G), verbal ability (V), numerical ability (N), spatial ability (S), form perception (P), clerical perception (Q), motor coordination (K), finger dexterity (F), manual dexterity (M), eye-hand-foot coordination (E), and color discrimination (C). Other aptitudes appearing in the testing literature include mechanical reasoning, abstract reasoning, sales aptitude, and musical aptitude, to name a few.

Some evaluators attempt to use a limited number of aptitude tests, such as verbal, spatial, and manual ability, to develop a general aptitude composite. While these three scores have utility, uncovering subtle differences only available when all subtests are administered, can improve the comparison of scores to the complex aspects of work (Cronbach, 1990). For example, a vocabulary subtest may be a good measure of verbal aptitude, but it may not fully represent performance on other aptitude subtests, such as verbal reasoning, spelling, and language usage. Ultimately, scores can be categorically grouped (e.g., cognitive, spatial, motor) for interpretive purposes, as long as score differences within each group are not significant.

Aptitude tests are available individually to measure a single aptitude (e.g., mechanical reasoning, clerical ability, spatial relations), or as multiple aptitude (or multi-aptitude) test batteries. Multiple aptitude test batteries are composed of a collection of eight-to-twelve subtests that cover a broad range of aptitudes, similar to the Department of Labor's 11 aptitudes (i.e., general learning ability, verbal, numerical, spatial, form perception, clerical perception, motor coordination, finger dexterity, manual dexterity, eye/hand/foot coordination, color discrimination) All subtests within a battery are universally similar in their layout, administration, scoring methods, norm groups, and interpretation strategies. This "universality" allows for the comparison of subtest scores—a procedure that is difficult to do with individual aptitude tests that do not share the same developmental philosophies or norm groups. Profiles are available for interpreting and comparing results on multiple aptitude test batteries. Percentile scores in the low thirties and higher is often indicative of average and better performance when compared to the chosen norm group.

Examples of individual and multiple aptitude tests are as follows:

- Individual Aptitude Tests
    - Bennett Mechanical Comprehension Test (BMCT)
    - Computer Operator Aptitude Battery (COAB)
    - Computer Programmer Aptitude Battery (CPAB)
    - Minnesota Clerical Test (MCT)
    - Minnesota Spatial Relations Test–Revised (MSRT)
    - Office Skills Test (OST)
    - Revised Minnesota Paper Form Board Test (MPFB)
    - SRA Clerical Aptitudes
    - SRA Test of Mechanical Concepts
- Multiple Aptitude Batteries
    - Career Ability Placement Survey (CAPS)
    - Differential Aptitude Tests (DAT; several editions are available)
    - Employee Aptitude Survey (EAS)
    - General Aptitude Test Battery (GATB)
    - Occupational Aptitude Survey and Interest Schedule (OASIS)

For the sake of administrative convenience, most aptitude tests have time limits. These time limits often affect the performance of individuals who do not respond well to the pressure of time, do not read or process quickly, and have difficulty marking the answer sheet rapidly. Work samples are a more work-related, "hands-on" method of aptitude assessment that can more easily minimize (or accommodate) the negative effects of time, processing, and manipulation on performance.

Basic Skills Instruments. Basic skills comprise those fundamental competencies related to independent living and working. They include activities, such as telling time, money handling, measuring (e.g., weight, volume, linear), sign recognition and survival words, consumer skills, job search skills, and knowledge of job keeping behavior. These are particularly important traits to assess in the prevocational phase with individuals whose basic skills are in question, as a result of a lack of community exposure, limited or no education, or processing disabilities, such as mental retardation or traumatic brain injury. Basic skills instruments are generally administered orally with the aid of pictures and other "hands-on" activities. In the strictest sense, they are not considered psychometric tests but standardized tests. Examples of basic skills instruments include:

- Life Centered Career Education Assessment System (Competency Rating Scale and Knowledge Battery–LCCE)
- Social and Prevocational Information Battery (SPIB–moderate and low level forms available)
- Street Survival Skills Questionnaire (SSSQ)
- Tests for Everyday Living (TEL)

Instructional materials and remedial suggestions are also furnished with these instruments that can be used to recommend or provide accommodations or improvements in identified skill deficits. With basic skills instruments, it is particularly important to supplement norm-referenced interpretation with criterion-referenced interpretation. For example, both methods of interpretation can be incorporated into statements, such as: "On the Ruler Reading subtest, Ms. Salazar's score at the 25th percentile, when compared to general population norms, indicated that she could only measure and draw lines down to a quarter of an inch." In a more dynamic, prognostic assessment, the evaluator would take the time to teach the participant how to read a ruler and administer the subtest again to see if learning took place, noting the method of instruction. If particular fine measuring skills are needed by the consumer, then goal-specific remediation or accommodation can be recommended.

Dexterity Tests. Dexterity is the "adroitness or skill in using fingers, hands, arms, and shoulders, sometimes in combination with other body parts. It is usually measured by observing performances on various work activities, such as work samples, or by administering standardized performance tests" (Dowd, 1993, p. 8). A dexterity test is a timed performance-based measure of various types of finger and manual manipulation, and eye-hand coordination. It is considered to be an aptitude test but is being reviewed separately because of its unique and extensive use in

vocational evaluation.

Dexterity tests (also referred to as performance tests) either assess hand use or tool use at both fine and manual levels. The Crawford Small Parts Dexterity Test measures fine finger and hand dexterity with tools (small screwdriver and tweezers); and the Hand Tool Dexterity Test (by Bennett) measures manual dexterity of hand, arm, and shoulder using larger tools (screwdriver, pliers, and wrench). Since individuals who have never used tools before do not perform as well on tool-oriented dexterity tests as people with experience, non-tool-oriented dexterity tests are often preferred for an unbiased assessment of general dexterity. Tool-oriented dexterity tests should be reserved for individuals with experience using tools or who seek training or employment in jobs requiring the use of related tools. The Purdue Pegboard is an example of a non-tool fine finger dexterity test and the Minnesota Rate of Manipulation Tests (MRMT) is an example of a non-tool manual dexterity test.

It is important to remember that dexterity tests are a measure of timed dexterity—how quickly someone can perform an activity requiring dexterous ability. In addition, previous experience in a job or activity involving dexterous skills will tend to increase performance on related dexterity tests. Likewise, practice effect (improvement resulting from repeated administration of a test in close time intervals) will have an effect on dexterity test scores. For this reason, dexterous ability can best be observed using tests, work samples, and situational or community-based assessments that provide sufficient opportunities for finger and/or manual involvement. In conjunction with performance scores, the evaluator can determine through observation if the person is fast and accurate, slow but accurate, or slow with difficulty in grasping, moving, aiming, and/or placing an object. Observation of frustration, attention to detail, motivation, retention of a sequence of activities, organization, and problem-solving can also be observed during the administration of dexterity tests. Many evaluators also include range- of-motion, strength, and motor coordination tests and activities under this category, especially those used to assess the functional abilities and limitations of persons with physical and motor impairments.

Intelligence Tests. Intelligence "is the global capacity of the individual to act purposefully, think rationally, and deal effectively with the environment" (Power, 1991; Wechsler, 1981, in Power 1991, p. 87). Power (1991, p. 87) further states, "Intelligent behavior is as much a function of drive and incentive as the more traditionally conceived components of intellectual ability, such as abstract and logical thinking, reasoning, judging, and retaining knowledge." Intelligence tests are also considered to be aptitude tests that measure general learning ability (or general mental ability), but they are usually classified separately from other aptitude tests. The more well-known Level C intelligence tests, such as the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the Stanford-Binet Intelligence Scale (SB) are restricted in use to licensed psychologists and psychological associates for clinical diagnosis. These and other Level C intelligence tests are designed to measure verbal (left brain) and performance (right brain) ability through a series of different cognitive and motor subtests. This level of

intelligence testing is only used in rehabilitation or transition when there is a need to diagnose the possible existence of mental retardation, a learning disability, or other cognitive/motor impairment.

Level B intelligence tests, which are available to vocational evaluators for individual or group administration, are geared to a brief screening of either verbal or performance ability. Evaluators, counselors, educators, and psychologists frequently compare IQ scores obtained from Level B and C intelligence tests to the standard scores from achievement tests to determine the possible existence of a learning disability (i.e., one or several achievement scores that are one or more standard deviations lower than an average IQ score). Although intelligence tests are a poor predictor of general employability, they are a better predictor of the level of traditional placement in training, education, and employment. The following examples of Level B intelligence tests have been classified as either verbal or performance measures.

Verbal Measures:

- Otis-Lennon School Ability Test-6th Edition (OLSAT–formerly the Otis-Lennon Mental Ability Test)
- Peabody Picture vocabulary Test-3rd Edition (PPVT-III)
- Shipley Institute of Living Scale (SILS)
- Slosson Intelligence Test-Revised (SIT)
- Wonderlic Personnel Test

Performance Measures:

- Culture Fair Intelligence Test
- Raven's Progressive Matrices (Standard and Advanced Progressive Matrices for Adults)
- Revised Beta Examination–Second Edition (Beta-II)
- Tests of Nonverbal Intelligence-3 (TONI-3)

In situations where an individual's verbal skills are not strong, performance measures can be used and may indicate the consumer's ability to develop verbal skills (e.g., someone with limited English-speaking proficiency, or someone who has had limited education or quality learning experiences). Verbal measures are particularly useful when direct placement into academic courses, formal education, or training is being considered.

Interest Tests/Inventories. Terms like self-awareness, motivation, drive, need, and level of interest describe the constructs of vocational interest inventories (Power, 1991). Interests are often a reflection of our values, attitudes, personality, and to some degree, our aptitudes. Interest Tests are one of the most widely used instruments in vocational evaluation and career counseling. They are often given at the very beginning to set a vocational tone for the evaluation and to identify any personal goals that can be used in the development of the evaluation plan. Because interest tests are self-report inventories of personal likes and dislikes, their classification as a psychometric test has been questioned. Interest inventories are

available in written and picture (non-reader) versions. Each test item may consist of two or three choices of work activities (written or pictorial) that allow the examinee to choose the most preferred, and sometimes the least preferred, activity in the set. Another format provides only one work activity (written or pictorial) at a time and examinees rate their level interest on a Likert-type scale (e.g., from very disinterested to very interested).

Results of an individual's high and low "tested interest" areas should be compared to "expressed interest" (statements made during the interview or contained in the file), and "manifest interest" (what was observed during evaluation, or performance on a related job or school subject; Power, 1991; Pruitt, 1986; Siefker, 1996; Super, 1949, in Power, 1991). When all three are consistent (expressed, tested, manifest), a career or vocational decision has been internalized. When one or all three are inconsistent, or inventory profiles are relatively flat (i.e., no significant difference in interest category levels throughout the inventory), then career exploration and counseling would be warranted. High- and low-interest areas should also be compared when exploring employment options. For example, if someone scores high in management and low in computation, then jobs in human services management would be preferred by that person over jobs in fiscal management. There are some Workers' Compensation and Social Security evaluations that do not focus on interest but on what exists in the local economy that an injured worker would be able to do, regardless of personal preference. Some of these evaluations may also focus on wage loss and lost earning capacity unrelated to interest in the job.

The following examples of interest inventories are divided into written and picture interest inventories. Written interest inventories:

- Campbell Interest and Skill Survey (CISS)
- Career Assessment Inventory (CAI)
- Career Decision-Making System (CDM-R)
- Career Occupational Preference System Interest Inventory (COPS–available in a variety of versions from intermediate to professional)
- Kuder Occupational Interest Survey (KOIS)
- Ohio Vocational Interest Survey–2nd Edition (OVIS II)
- Self-Directed Search (SDS; available in standard and low reading versions; also covers self-report of abilities)
- Strong Interest Inventory (SII)
- Vocational Research Interest Inventory (VRII)

Picture interest inventories:

- Career Occupational Preference System–Picture Interest Inventory of Careers (COPS-PIC)
- Geist Picture Interest Inventory–Revised (GPII-R)
- Reading-Free Vocational Interest Inventory–Revised (R-FVII)
- Wide Range Interest-Opinion Test (WRIOT)

Versions of interest tests are available for students in a range of grades from junior

high to college, for adults both vocational and professional, and for individuals who are mentally retarded. Tests can be self-administered and scored by the examinee, hand scored by the evaluator, or computer administered and scored. Profiles generally cluster specific interest scales under broader occupational themes. Some interest profiles will contain attitudes and/or values scales as well.

Interest inventories are not as useful for individuals who have limited knowledge of or experience with the world of work, and who may not understand the jobs or activities contained in the test. In addition, a lack of definitive validity studies on interest tests has brought their accuracy into question.

Learning Style Tests. VEWAA defines learning style as "the way in which an individual learns new material. Learning style is usually defined in terms of the sensory modalities (e.g., visual, auditory, tactile, and kinesthetic) by which the person learns the fastest" (Dowd, 1993, p. 17). Learning assessment is the "determination of the potential to learn by identifying what teaching or behavioral change techniques are most effective" (Dowd, 1993, p. 17). The ability to understand, remember, and recall a set sequence of information essential to success in the classroom, on the job, or in the community, is a key element in targeting preferred (or primary) learning styles. The assessment of cognitive and learning styles has become a very important part of the vocational evaluation process. CARF requires that it be available for use when necessary; and schools require a learning style assessment with all students with learning disabilities, so that classroom accommodations can be provided.

Each personal way of dealing with information and experience, which forms the basis of learning style, can be related to the conditions, content, modes, and expectations of learning as well as to the stimuli and elements of the learning environment (Blakemore, McCray, & Coker, 1984). Dunn, Dunn, and Price (1979) identified five major factors (or stimuli) that affect learning: environmental (e.g., sound, lighting, temperature), emotional (e.g., motivation, persistence, structure), sociological (e.g., working alone or in a team with peers or authority figures), physical (e.g., time of day, mobility, presentation format), and psychological (e.g., analytical/global, reflective/impulsive, cognitive style). Many learning style inventories, such as the CITE Learning Styles Inventory, identify (a) how a student gathers information (auditorily, visually, with language, numerically, or kinesthetically), (b) the student's preferred working conditions (alone or with others), and (c) his/her expressive preferences (verbal or written)" (Blakemore et al., 1984, p. 49). Other test formats use an assessment of brain dominance (left brain versus right brain learning), or an identification of values and temperaments (sensing/intuiting, thinking/feeling) that influence how individuals learn.

Learning style instruments come in two basic forms: self-report tests and performance- based tests. **Self-report tests** can be obtained in either paper-pencil or computer formats, and are quick and easy to administer and score. They require readers to rate their preference for statements that describe conditions, situations, and study/learning approaches with which they are most comfortable (e.g., I study

best alone; I learn more from listening; I like to study with background noise). For low readers or individuals who have had limited or unsuccessful learning experiences, self-report instruments will not be particularly accurate.
Performance-based tests require the evaluee to engage in a series of activities that involve looking at or hearing a series or sets of letters, colors, and/or geometric patterns and recalling the information. The sets become progressively longer, and recall of a series can occur immediately after the presentation and again at the end of the test. The evaluator can assess short- and long-term memories and the level of sequencing (i.e., how may items be remembered in their correct order).

Following are examples of learning style instruments listed by self-report and performance-based formats. Self-report tests:

- CITE Learning Styles Inventory
- Learning Style Inventory (Dunn, et al.)
- Productivity Environmental Preference Survey
- TLC Learning Style Inventory (Hanson & Silver)
- Vocational Learning Styles Media Kit
- Your Style of Learning and Thinking

Performance-based tests:

- Pathfinder (formerly the Trainee Performance Sample, assesses at the trainable level)
- Learning Efficiency Test–II (LET-II)
- Perceptual Memory Task (PMT, assesses at the educable level and above)
- Personnel Tests for Industry–Oral Directions Test (PTI-ODT)

A more informal process of learning style assessment will be discussed later in the book. If a formal assessment and identification of preferred learning style are requested, or a learning problem is suspected, then an evaluator may choose to use a standardized learning style test early in the evaluation process. These results can be verified through informal observations of how well individuals follow instructions on other tests, work samples, and situational assessments. On the other hand, if during the evaluation an informal assessment uncovers a possible learning problem, standardized learning style instruments can be used to identify strengths and limitations in learning style. Appropriate accommodations in administration and instructional style can be explored during the remaining evaluation, and noted in the final staffing and report. As Leconte & Rothenbacher (1987, p. 164) put it:

> "As in recommended practices for interest assessment, it is important to look beyond formal test instrument results and use observations and other informal techniques to substantiate findings. In other words, evaluators are encouraged to synthesize the results of tested, expressed, and manifested learning styles into a unique individual profile."

Temperaments and Work Values Tests/Inventories. Temperaments are "the adaptability requirements made on the worker by specific types of jobs.

Temperaments became one of the components of job analysis because it was found that different job situations called for different personality traits on the part of the worker" (Dowd, 1993, p. 27). The U.S. Department of Labor (1991b) included ten factors under the heading of Temperament, such as: working alone, expressing personal feelings, dealing with people, performing repetitive work, performing under stress, performing a variety of duties.

Work values are defined as "an intrinsic value placed on a construct, internal or external, of the worker, such as creativity, independence, altruism, attitude toward and pride in work, and so on. Identified strengths in values may help in vocational exploration and/or job placement" (Dowd, 1993 p. 33). It has been argued that there is little difference between temperaments and work values, because they are both used to supplement interest information. When temperaments and work values are consistent with tested, expressed, and manifest interests, greater reliance can be placed on the vocational decision made by the consumer. However, when there are little, if any, expressed or tested interests, results from temperaments and work values inventories can be used as a starting point for career exploration. For example, the individual who states "I don't know exactly what I want to do, but I want to work by myself," may lead the evaluator to explore jobs or environments where contact with others is minimized (e.g., night security guard, accountant, computer programmer, on-line office at home).

These self-report instruments classify the tested range of temperaments or work values from their highest to lowest ranking, or on a profile with dichotomous values/temperaments on either end of the scale (e.g., introvert to extrovert). Examples of work values and temperaments inventories include:

- Career Orientation Placement and Evaluation Survey (COPES)
- Minnesota Importance Questionnaire (MIQ)
- Myers-Briggs Type Indicator (MBTI)
- The Salience Inventory (SI)
- The Values Scale (VS)
- Temperament and Values Inventory (TVI)
- Work Temperament Inventory (WTI)
- Work Values Inventory (WVI)

Other Tests. There is a broad range of standardized tests that cannot be classified in one of the previous eight categories but are useful to vocational evaluators. They include instruments, the Dvorine Color Vision Test (sometimes classified under aptitude) and the PDI Employment Inventory, and various standardized behavior rating scales, such as the Becker Work Adjustment Profile (BWAP), Prevocational Assessment and Curriculum Guide (PACG), Vocational Assessment and Curriculum Guide (VACG), and the AAMR Adaptive Behavior Scales— Residential and Community. Also, various emotional state, work personality, and counseling tests/inventories, such as the Eight State Questionnaire (8SQ), FIRO-B Awareness Scale, Gordon Personal Profile-Inventory, Hogan Personality Inventory–2nd Edition,

Manson Evaluation— Revised (ME), Million Index of Personality Styles, Motivation Analysis Test (MAT), Occupational Stress Inventory, and the Work Personality Profile can be used when such assessments appear relevant to rehabilitation, transition, and employment.

There are far too many tests to list in this section that can provide useful information to vocational evaluators and consumers. The Rehabilitation Resource publication Tests and Test Use in Vocational Evaluation and Assessment (Siefker, 1996) reviews a variety of tests often used in the field. Many other books are available that give an overview of tests and measurement theory (Anastasi & Urbani, 1997; Cronbach, 1990; Drummond, 1996; Lyman, 1991), and that review tests commonly used in assessment, counseling, and human services (Kapes, Mastie, & Whitfield, 1994; Keyser & Sweetland, 1984–1994; Kramer & Conoley, 1992; Maddox, 1997). Refer to the Appendix section for a selected list of test publishers/marketers. A copy of their most current product catalogs can be requested at no charge, which gives descriptions of available tests and prices. Specimen sets are often available for review at a lower cost than complete test packages.

Not everyone can profit from psychometric testing (Power, 1991; Thomas, 1991). There are times when tests can underestimate potential and screen certain individuals or groups out of appropriate opportunities. Some of the circumstances that adversely affect testing include memory or processing problems, motor difficulties, low-performance speed, difficulty with the English language, cultural difference, and test anxiety. For example, individuals who are clinically depressed do not process information quickly and should be given power tests (untimed tests) rather than speeded tests whenever possible. Obtaining and reporting both a timed and untimed score on a timed test would also yield meaningful information. Evaluators must determine what barriers will prevent psychometric tests from accurately assessing an individual's current potential, and make appropriate accommodations. When this is not feasible, they must choose other instruments (e.g., work samples) or techniques (e.g., situational assessment and/or OJE) that will provide a more valid assessment. As Owings (1992, p. 176) describes it:

> The dichotomy is valid test scores versus valid assessments of individuals. They are not the same. Despite previous admonitions, tests can be successfully modified to obtain better information about the client—not necessarily better test scores. There is an enormous difference in the vocational usefulness of accurate information versus accurate test scores. If the test is inappropriate for the client, correct use of it will produce valid scores but not necessarily information that will be useful in predicting job success.

Work Samples and Systems

What are Work Samples? As the name implies, a work sample is simply a "close simulation," a "mock-up," or a "sample" of work (Neff, 1985). More specifically it is:

> A well-defined work activity involving tasks, materials, and tools that are identical or similar to those in an actual job or cluster of jobs. Work samples are used to assess a person's vocational aptitude(s), work characteristics, and/or vocational interests. There are several specific types of work samples: Cluster Trait, Job Sample, Simulated, and Single Trait (Dowd, 1993).

Hugo Munsterberg has been credited with developing the first work sample in the early 1900's (Nadolsky, 1973; Pruitt, 1986). It was a model of a streetcar used to evaluate applicants for operator positions with the Boston Railroad Company (Bregman, 1969)**.** Considered to be one of the first attempts at personnel selection for a particular job, Munsterberg also attempted to compare scores of applicants to their performance as operators (Rosenberg, 1973).

As mentioned earlier, work samples are initially more expensive to purchase, and generally take longer to give than psychometric tests. But with these disadvantages come advantages. Since work samples take longer than many psychometric tests, they provide an opportunity to observe task-related behaviors, involve the evaluee in hands-on career exploration and decision-making, and try out various accommodations and modifications to determine what might improve learning and performance (Kaiser & Modahl, 1991; Power, 1991; Pruitt, 1986; Thomas, 1991). Work samples can be used as situational tools to assess stamina, evaluate improvements in learning and performance over repeated trials, and engage in work adjustment to modify unacceptable work behaviors. Because they look more like work than a test, Nadolsky (1973, p. 3) found that culturally disadvantaged "clients who received vocational evaluation services viewed work samples as being less threatening than psychological tests and responded in a positive manner to the work sampling procedures. In general, through the use of work samples, both the client and the counselor received information about the client's work behavior and vocational potential that was highly relevant and previously unavailable to them."

Types of Work Samples. A work sample is based on a job analysis, or other occupational information, and is a closer approximation of work than a psychometric test. The Vocational Evaluation and Work Adjustment Association (1975, p. 55) identified four types of work samples, which include:

- an actual job itself moved into the evaluation unit,
- a simulation of an actual operation,
- a trait sample, which assesses a single factor, such as finger dexterity, and
- a cluster trait sample, which measures a group of traits.

The "actual job" or job sample, and the "simulation" or simulated work sample have high face validity (i.e., they look similar to work activities). These are often referred to as content- based or criterion-referenced instruments. On the other hand, the "trait sample" or Single-Trait Work Sample, and the "cluster trait sample" or Cluster Trait Work Sample, are more abstract and do not readily resemble a real or

simulated work activity (i.e., they look more like a test). These are referred to as construct-based or norm-referenced instruments.

Training assessment samples are similar to work samples and are used to assess the potential for training in an area where formal preparation is required. Someone cannot be employed as a Registered Nurse, for example, without completing training and becoming registered. Since it is not feasible to develop a nursing work sample, a two-phase approach would be warranted—a cognitive and a performance evaluation. The first phase (the cognitive evaluation) would require an assessment of the mental and academic abilities needed for nursing (e.g., verbal and mathematical achievement or aptitude) and an ability and interest in using the common language of the chosen professional field; in this case, medical terminology. Having an evaluee read the first chapter of a medical terminology text and take a written test to assess retention and application would help both consumer and evaluator explore interest and potential.

The second phase, the performance evaluation, would require that a series (or cluster) of tests, work samples (e.g., a vital signs work sample), and/or situational assessments (e.g., reading and completing medical charts, making a bed) be used to assess the performance aspects of a nursing job. If an individual does not currently have the potential (or motivation) to master medical terminology and succeed in a rigorous educational program but demonstrates interest and potential during the performance phase, then an entry-level job or on-the-job training as a nurse aide might be considered. In relation to long-range career development, it could be recommended that the consumer pursue training, possibly as a licensed practical nurse at a local community college or training hospital, following a year or two of successful employment as a nurse aide. This allows additional time to become familiar with medical terminology and procedure and decide if there is sufficient interest and motivation to seek further career training. If interest and potential surface during the cognitive phase but not the performance phase, other medically related jobs could be explored with the consumer.

A training assessment sample can be created by standardizing the medical terminology activity and choosing an appropriate cluster of related cognitive and performance instruments and techniques. Training assessment samples can be developed to cover terminology in electronics, computer programming, accounting, engineering, psychology, or other technical and professional fields. More applied activities, such as using terminology in case studies or problem-solving exercises (e.g., a lab experiment, reading a technical graph or schematic), should also be incorporated into training assessment.

Basic skills samples are commonly used to assess functional skills essential to independent living, training, and working. These include, but are not limited to, telling and using time (e.g., clocks, calendars, bus schedules, appointment schedules), money handling (e.g., making change, budgeting, writing checks), recognizing signs and survival words, using maps, reading dials and gauges (e.g., stoves, washing machines, automobiles), measuring (e.g., linear, volume, weights),

and using the telephone and telephone book. Many evaluators design their own assessment devices or purchase standardized basic skills instruments, such as the Street Survival Skills Questionnaire (SSSQ), and the Social and Prevocational Information Battery (SPIB), described in the previous section on psychometric tests. If locally developed basic skills samples are not normed, then they are classified as situational assessment activities.

Work Sample Development and Standardization. A work sample can represent an entire job, or one or several tasks of a job (or course). A card filing work sample may be designed to assess the skills needed for a file clerk position, or it can assess one of the skills (i.e., card filing) required of a secretary or clerk-typist. It may not be feasible to include all tasks of an elaborate job into a single work sample; therefore, a combination (or cluster) of instruments and techniques will be needed to assess all essential job duties. There are also compounding factors that limit the broad application of a work sample to identical jobs in the same or different work environments. All jobs with the same title: (a) do not possess the same job duties, (b) do not place corresponding value to the same duties, and/or (c) do not use the same technology on similar duties in different environments.

For example, there are three general performance criteria for the job of a cashier (or cash register operator): speed, accuracy, and use of technology. A high volume of customers often requires cashiers in a grocery store to work faster than cashiers in a small specialty store where volume is not as great, and where other tasks (e.g., waiting on customers, stocking shelves) are equally important. Cashiers in all environments are required to accurately operate the cash register and make change; however, the medium of exchange differs. In grocery stores, just about any medium of exchange (just short of bartering) is used, including cash, checks, credit and debit cards, coupons, and food stamps. On the other hand, some small stores and restaurants will only accept cash and credit cards, but no checks.

Technology also varies greatly. Cash registers in some fast food restaurants only require the operator to press keys that correspond to the item ordered (e.g., large drink key, small French fries key, cheeseburger key), and the tax is automatically totaled with the sale. Scanners and bar code readers on many new cash registers have made entering correct prices easier. Some cash registers in department store chains are similar to computer terminals and require a variety of codes to be entered (e.g., sales clerk number, item inventory number) before a sale can be made. Cash register technology will continue to improve to the point where the job of a cashier may become obsolete. If one were to develop norms for and validate a cashier work sample, what job, criteria, and technology would be used? This dilemma in standardization not only affects vocational evaluation but training as well, and the use of work samples to generalize performance to a variety of different work environments calls for skill and caution on the part of the evaluator.

In order to evaluate the ability to succeed in a particular job or course, the evaluator must analyze the job tasks or classroom activities and select instruments and techniques that relate to the tasks/activities in question. Again using the example of

a cashier, an evaluator must first know what the job entails. This can be accomplished through the review of a local job analysis (e.g., grocery store cashier) or a more general job description available in occupational information documents, such as the *Dictionary of Occupational Titles* (DOT; U.S. Department of Labor, 1991a). From there, instruments and techniques can be chosen to assess the "objective" functions (e.g., machine operation, manual dexterity, money handling) and the "subjective" functions (e.g., communication and interaction skills, standing, reaching, and lifting).

The best way to evaluate for potential as a cashier is through a supervised community- based assessment (on-the-job evaluation). However, when such opportunities are not feasible, cash register operation and change-making work samples can be used. They provide relatively high content orientation and face value (i.e., close relationship to the job) for improved career exploration, decision-making, and modification/accommodation purposes. As a final consideration, a variety of construct-oriented instruments can be clustered together, such as a clerical aptitude test that measures speed and accuracy in matching letters and numbers, a manual dexterity test to assess the ability to manipulate a keyboard, and a math test to evaluate change-making. The latter method lacks realism and provides less content match, requiring more subjective judgments on the part of the evaluator and evaluee. Therefore, mixing content and construct instruments (e.g., an adding machine operation work sample, a change-making basic skills instrument, and a clerical matching aptitude test) would offer the best of both approaches. Developing a cluster of instruments and techniques around a job or course will ensure that all essential tasks and performance areas are covered. A single work sample or test cannot assess all of the aptitudes, physical demands, temperaments, behaviors, communication needs, social requirements, and environmental conditions of a particular job (Power, 1991). Clustering of appropriate instruments and techniques that can address these varied issues is essential.

In situations where all of the essential job tasks or course activities cannot be assessed, the evaluator will need to focus attention on key essential tasks. Three essential tasks must be considered: (a) the most time-consuming task, (b) the most difficult task, and (c) the most important task ("Americans with Disabilities Act of 1990," 1991b; Connolly, 1975; U.S. Department of Labor, 1991b). If these three areas can be addressed in the evaluation, there is a high likelihood of accurately assessing potential in individuals who have the ability to generalize skills or learning style. There may be more than one essential task that shares the same characteristic (e.g., they are both considered the most time consuming). Likewise, there may be a task that has more than one characteristic (e.g., it is the most important and the most difficult).

Similar to psychometric tests, work samples are standardized. Work samples should have industrial norms or standards for comparison to employee or applicant populations (Botterbusch, 1981; CARF, 1996; McCray, 1980; McFarlane et al., 1988; Power, 1991; Stout Vocational Rehabilitation Institute, 1977). Although disability norms in and of themselves provide little opportunity for comparison to working

populations, employed client norms can be highly useful (Berven & Maki, 1982). Industrial standards in the form of predetermined time standards are available on some work samples and systems (Hume, 1973; Shinnick, Black, & Decker, 1983; Vactor & Hubach, 1979). Two predetermined time methods used to standardize work samples are Methods-Time-Measurement (MTM, MTM2, MTM3) and MODAPTS (MODular Arranged Predetermined Time Standards).

These and many other predetermined time techniques were originally developed by industrial engineers to determine the most cost-effective way to assemble and package a product on a work line. Ergonomic principles and body mechanics are used to determine how long it takes the "average" worker to perform a series of movements on a specific industrial or office job. Once a series of movements is identified for a task, their predetermined, or standard, times are added together to determine how long it takes to perform each work task and the total job. This information is then used to set production quotas and determine labor costs. Percentile scores on tests and predetermined time percentages are different. When the norm table of a standardized test is used, the 99th to 100th percentile represents optimum performance on the test items. With a predetermined time standard, 100% Standard or 100% IN (Industrial Normal) refers to the performance needed by a competitive worker to meet the expected production quota. Predetermined percentages can range well above 100% (e.g., 150%).

When standardized work samples are used as designed, they can yield pertinent information on current functioning. However when used prognostically, they can evaluate improvement in performance that is only available through a dynamic assessment. If the standardized approach does not initially result in a positive outcome, evaluators should determine what affected performance and make appropriate modifications/accommodations to overcome the problem in subsequent administrations. The evaluator must never forget that it does not matter as much how an individual "scored" on a work sample, but rather what the person got right and wrong and how performance could be improved if the sample was administered again. Criterion-referenced rather than norm-referenced assessment is the "key" to a creative, flexible, and successful work sample-based evaluation. If modifications or accommodations serve to improve performance, then recommendations for similar changes on the job or in the classroom can be made. If performance does not improve on a work sample as a result of the prognostic approach, then other work samples or instruments should be selected and used.

Commercial Work Sample and Evaluation Systems. Vocational evaluators can develop and standardize their work samples (Botterbusch, 1981; McCray, 1980; Stout Vocational Rehabilitation Institute, 1977), or purchase commercially available work samples and evaluation systems (Brown et al., 1994; McFarlane et al., 1988). The first work sample system developed specifically for use in vocational/work evaluation were the TOWER and JEVS work sample systems (Pruitt, 1986; Rosenberg, 1973). The TOWER (Testing, Orientation, and Work Evaluation in Rehabilitation) system began development in the 1930s at the Institute for the Crippled and Disabled (now the International Center for the Disabled). In 1958, the

Philadelphia Jewish Employment and Vocational Service began work on the JEVS work sample system. Due to the limited training available for evaluators, the TOWER and JEVS systems provided training for purchasers of their system. This training focused on all aspects of the evaluation process, including interviewing, administration, behavioral observation, scoring, interpreting results, and report writing. TOWER and JEVS were selling more than a well-organized collection of work samples, but a process of evaluation as well.

Another early system, the THOMASAT, was developed by the Highland View Hospital in Cleveland to evaluate the cognitive-motor functioning of individuals for jobs performed in a sheltered workshop (Rosenberg, 1973). The TOWER, JEVS, and THOMASAT incorporated a variety of work samples or activities to evaluate a wide range of tasks and job functions. Although these three systems are no longer being marketed, they were originally developed to evaluate and predict job placement and success of adults with disabilities and disadvantaged youth. One of the earliest single work samples still available, the Pennsylvania Bi-Manual Work Sample, was developed by the McDonald Training Center in Florida (Pruitt, 1986). Today there are approximately 18 work sample and evaluation systems commercially available (Brown et al., 1994).

Commercially available **work sample systems** and **evaluation systems** are composed of a group of individually designed and standardized work samples, tests or activities that share the same developmental philosophy and norm groups. They also share similar methods of administration, scoring, and interpretation. This "universality" allows for the comparison of the results of all instruments within the system or battery. This type of comparison is more difficult with independent tests and work samples developed with different philosophies, norm groups, and approaches to scoring and interpretation. Universality carries over to both work sample systems and evaluation systems; however, there is a difference between the two.

Work sample systems (also referred to as work sample batteries) are composed of standardized instruments that resemble work or work-related activities. Therefore, they have high face value and content orientation, which readily lend themselves to both norm-referenced and criterion-referenced interpretations. Activities, such as card filing, message taking, proofing/editing, adding machine operation, data entry, sorting, assembly, tool usage, electrical wiring, and sewing machine operation, are often found in many work sample systems. Since they look like work, evaluees relate to them more as a work activity rather than a test. Thus testing anxiety is reduced and the consumer is more motivated to participate (Pruitt, 1986). On work samples lasting more than 20 minutes, work-related behaviors can be observed, and job-related modifications attempted. Examples of some commonly used work sample systems (or work sample batteries) include Micro-TOWER, Skills Assessment Module (SAM), System for Assessment and Group Evaluation (SAGE), Talent Assessment Program (TAP), VALPAR Component Assessment Systems, Vocational Evaluation Systems (VES), Vocational Information and Evaluation Work Samples (VIEWS), and the Vocational Interest Temperament Aptitude System (VITAS). These

and related systems/batteries represent the oldest and most traditional approach to vocational evaluation.

Some commercial work sample systems must be purchased in their entirety, and it is recommended that all instruments in the system be administered to provide the most comprehensive interpretation possible. With other systems, the evaluator can buy one or several work samples and use them independently or in combination with other evaluation instruments. Most all of these systems either require training for purchase (depending on the skill of the evaluator) or offer it as an option. Similar to the original TOWER system, the SAVE (Systematic Approach to Vocational Evaluation) system is sold as an evaluation manual that contains all forms and information for building, administering, and scoring the work samples listed in the manual. The evaluator purchases the materials and supplies locally that are needed to build all or selected work samples from the manual.

Work samples within a system may take anywhere from ten to 45 minutes to administer. Total battery administration may last from a half-day to nearly a week, depending on the length and number of work samples in the system (batteries can have anywhere from ten to 28 individual work samples). For 1:1 ratios, a participant can take all or parts of a system depending on individual needs and the types of referral question(s). With higher ratios, they can be administered two different ways. The first is a group administration where everyone in the group takes the same instrument at the same time. This requires that the evaluator has as many systems as there are people in the group (e.g., three systems for a 3:1 ratio). One administration can be given to the entire group for each work sample, followed by a group discussion of their results and interests in the sample just taken.

In the second method, two or more individuals are placed on different work samples in a battery at the same time. Over the course of the evaluation, all participants may eventually take the same work samples but at different times. Most systems do not have a set order in which work samples must be administered. This is up to the discretion of the evaluator and the availability of the instrument.

Evaluation systems are composed of a series of standardized tests or activities that are more abstract than work samples. These construct-oriented instruments generally use a norm- referenced approach and result in percentile scores or occupational codes (e.g., aptitude codes). Terms, such as abstract reasoning, verbal ability, numerical ability, visual tracking, finger dexterity, manual dexterity, hand strength, eye-hand-foot coordination, spatial relations, and form perception are often used to describe the instruments contained in most evaluation systems. Examples of some evaluation systems include APTICOM, Career Evaluation Systems (CES), Career SCOPE, Hester Vocational Evaluation System (MVE), Key Educational Vocational Assessment System (KEVAS), McCarron-Dial Evaluation System (MDS), and Vocational Transit. Although the Computerized Assessment (COMPASS) is classified here as an evaluation system, it uses a criterion-referenced rather than a norm-referenced approach to scoring. Some of the systems look very much like a computerized aptitude test battery and can be used successfully with moderate to

higher functioning individuals.

These newer-generation evaluation systems are usually sold as a package (individual instruments are not sold separately). Most of them take a day or less to administer, and several only take a few hours. Training is either required or optional, depending on availability and evaluator need. Although many work sample systems use a computer for scoring and report writing, nearly all of the evaluation systems require a computer for administration, scoring, and report writing. Some evaluation systems may not have high face value to work or be as easy to modify as work samples (except Vocational Transit); however, they are generally quicker to administer and score. Some evaluators use the shorter evaluation systems as a tool for deciding if a work sample evaluation would be beneficial, and what instruments should be administered.

Not all batteries, or instruments within batteries, can be easily classified as a work sample system or evaluation system. Some individual instruments and batteries fall somewhere in the middle ground of the continuum. Discretion must be used in choosing appropriate instruments to ensure that they do not intentionally screen individuals out, but at the same time, are not so easy that they insult the intelligence of the participant and underestimate potential. In addition, these systems have good standardization, and many report a variety of norm groups, reliability, and validity studies.

Review of Commercial Systems

The publication *Vocational Evaluation Systems and Software: A Consumer's Guide* (Brown et al., 1994) provides a description of nearly all of the commercial work sample and evaluation systems on the market today. It is available from The Rehabilitation Resource listed in the *Resources* section of this book. Although the Brown et al. (1994) publication does not list the more recent CareerScope and the Wide Range Employability Sample Test (WREST) that was reviewed is no longer being marketed, this publication is an excellent starting point for narrowing down evaluation and work sample systems to be considered. The publication also contains a section on how to assess and choose a system. Some of these considerations include: (a) purpose of the system, (b) populations for which the system was developed, (c) administration method (to groups or individuals), (d) cost to purchase and maintain, (e) space needed to house the system, (f) time needed to administer the system, (g) scoring and interpretive strategies, and (h) availability of training and support. It also reviews 12 commercially available job search software systems frequently used by vocational evaluators. Following is a brief review of commercial evaluation and work sample systems (including the CareerScope) abstracted from the Brown et al. (1994) publication. These descriptions are by no means comprehensive, and the developers should be contacted directly for additional information on each system.

APTICOM. A computer-driven, hardware-oriented evaluation system consisting of ten aptitude tests, one interest inventory with 12 interest areas, and several language

and math skills tests. The entire battery can be completed in under 2 hours.

Vocational Research
Institute 1528 Walnut
Street, Suite 1502
Philadelphia, PA 19102-3619
800-874-5387 or 215-875-7387

Career Evaluation Systems (CES). Three separate computer systems that score batteries of standardized tests include: (a) CareerView for average or above average persons with no physical impairments seeking career guidance (190 minutes), (b) VocScan for individuals with physical disabilities (250 minutes) and low reading levels (200 minutes), and (c) JobSupport for individuals who are mentally retarded (200 minutes).

Career Evaluation
Systems, Inc. 6050 West
Touhy
Chicago, IL 60648
312-774-1212

CareerScope. A software-based alternative to the APTICOM that uses a standard computer for administration and scoring of an aptitude and interest inventory. Both aptitude and interest components can be completed in under 2 hours.

Vocational Research
Institute 1528 Walnut
Street, Suite 1502
Philadelphia, PA 19102-3619
800-874-5387 or 215-875-7387

Computerized Assessment (COMPASS). A battery of 12 computer-based subtests, three work samples, and two surveys that yield 17-factor scores related to 11 aptitudes, as well as to reasoning, math, and language. The system can be administered in about 70 minutes.

VALPAR International Corporation
P.O. Box 5767
Tucson, AZ 85703-5767
800-528-7070 or 602-293-1510

Hester Vocational Evaluation System (MVE). The system is composed of eight apparatus-type and nine standardized paper-and-pencil tests that result in 19 ability factors and 17 personal characteristics. The administration time is approximately 3.5 hours.

Hester Evaluation
Systems, Inc. 2410
Southwest Granthurst
Topeka, KS 66611-1274
800-832-3825 or 913-357-0362

Key Educational Vocational Assessment System (KEVAS). A computer-assisted system supplemented with performance-based hardware and standardized paper-and-pencil tests. Twenty-two areas of functioning are measured under three categories: psychophysical functioning, work-related competencies, and social and motivational functioning. A total of 3.5 hours is required for administration.

Key
Evaluation,
Inc. 673
Broad Street
Shrewbury, NJ 07702
201-747-0048 or 800-25-KEVAS (outside NJ)

McCarron-Dial Evaluation System (MDS). A series performance-based, standardized tests (including a paper-and-pencil test, and a behavior rating scale and inventory) designed to assess five factors within the three basic dimensions of verbal-spatial-cognitive, sensorimotor, and emotional coping. The basic battery takes around three hours while the comprehensive battery requires up to five days, including 10 hours for behavioral observations.

McCarron-Dial Systems, Inc.
P.O. Box
45628
Dallas,
TX 75245
214-247-
5945

Microcomputer Evaluation of Career Areas (MECA). Composed of 15 microcomputer, work-oriented career exploration and assessment kits (e.g., automotive, business and office, health care, manufacturing). Each kit takes approximately 30 minutes to administer through the computer and simulated work activity.

Conover Company
P.O. Box
155
Omro,
WI
54963
800-
933-1933

Micro-TOWER. Consists of 13 self-contained, group-administered work samples under the five aptitude clusters of verbal, motor, numerical, spatial, and clerical perception. The administration time for all work samples is between 14.5 and 25 hours, including time for breaks and group discussions.

Micro-TOWER
ICD Rehabilitation & Research
Center 340 East 24th Street

New York, NY 10010

Skills Assessment Module (SAM). Assesses 25 affective, cognitive, and psychomotor abilities using three paper-pencil tests and 12 hands-on work samples. The battery can be administered in 2.5 to 3.5 hours.

Piney Mountain Press, Inc.
P.O. Box 333
Cleveland,
GA 30528
800-255-
3127

System for Assessment and Group Evaluation (SAGE). The battery contains 17 test instruments and work samples consisting of five components: Vocational Assessment Battery of 11 aptitudes, Cognitive-Conceptual Abilities Test of general educational development, Vocational Interest Inventory, Assessment of Work Attitudes, and Temperament Factor Assessment. The total administration time is 4 to 5 hours.

Train-Ease
Corporation
PESCO
21 Paulding Street
Pleasantville, NY 10570
800-431-2016

Systematic Approach to Vocational Evaluation (SAVE). Package A assesses 16 worker trait groups (for the mentally retarded and academically deprived) and Package B expands the assessed worker trait groups to 46, for broader use. A manual is sold with no equipment but with information on building 47 work samples. The entire battery takes 15 to 20 hours to administer.

SAVE
Enterprises
16 Downing
Street
Rome, GA 30161
706-295-6407

Talent Assessment Program (TAP). Composed of 10 performance-based tests and activities grouped into three categories: Visualization and Retention; Discrimination; and, Dexterity. The administration time is 2.5 hours or less.

Talent Assessment, Inc.
P.O. Box 5087
Jacksonville,
FL 32247 904-
260-4102

VALPAR Component Assessment Systems. Contains 19 separate work samples

and activities covering areas, such as small tools use, clerical comprehension, problem-solving, assembly, sorting, range of motion, drafting, and physical capacity. The administration time is 15 to 90 minutes each, depending on the work sample.

VALPAR International Corporation
P.O. Box 5767
Tucson, AZ 85703-5767
800-528-7070 or 602-293-1510

Vocational Evaluation Systems (VES). It contains 28 separate, audio-visually administered, work sample carrels. Examples include bench assembly, drafting, electrical wiring, sales processing, cooking/baking, engine service, cosmetology, and office services.

Approximately 2.5 hours are required for each work sample.

New Concepts Corporation
2341 South Friebus Avenue, Suite #5
Tucson, AZ 85713
800-828-7876 or 602-323-6645

Vocational Information and Evaluation Work Samples (VIEWS). It consists of 16 work samples grouped into the four lowest worker skill groups: Materials Sorting, Clerical Matching and Counting, and Assembling; Machine Feeding; Routine Tending; and, Fabricating. The battery is designed for the mentally retarded and takes between 15 and 20 hours to administer.

Vocational Research
Institute 1528 Walnut
Street, Suite 1502
Philadelphia, PA 19102-3619
800-874-5387 or 215-875-7387

Vocational Interest Temperament Aptitude System (VITAS). It contains 21 work samples related to work groups from the DOL's Guide to Occupational Exploration. The battery is designed for the educationally and/or culturally disadvantaged and takes approximately 15 hours to administer.

Vocational Research
Institute 1528 Walnut
Street, Suite 1502
Philadelphia, PA 19102-3619
800-874-5387 or 215-875-7387

Vocational Transit. This computer-based evaluation system consists of four electronic test modules that assess the lowest level of General Educational Development, and the four aptitudes of motor coordination, manual dexterity, finger dexterity, and form perception. It is designed to evaluate low functioning individuals in around 90 minutes.

Vocational Research
Institute 1528 Walnut
Street, Suite 1502

Philadelphia, PA 19102-3619
800-874-5387 or 215-875-7387

Most of these work sample and evaluation systems are both norm-referenced and criterion-referenced. The majority is related to the Department of Labor's occupational coding and classification systems created for the use with the *Dictionary of Occupational Titles (4th Edition;* (U.S. Department of Labor, 1991a), the *Revised Handbook for Analyzing Jobs* (U.S. Department of Labor, 1991b), and related publications. These include, but are not limited to, the Department of Labor's data/people/things codes, 11 aptitudes, General Educational Development (reasoning, math, and language), Physical Demands, and Temperaments. Almost all of these systems offer computer-generated profiles and/or reports, and many of the software programs are capable of printing a list of job titles with DOT codes that relate to the results of that particular battery.

Conclusion

Vocational evaluators are fortunate to have such a unique variety of tools (instruments, techniques, and strategies) at their disposal. Given the limits of the evaluation environment, deciding what instruments to purchase and use requires sensitivity and sound judgment; especially when considering the most accurate yet efficient way to meet the needs of the consumer and referral source. Evaluators must also be aware of their limits in using certain standardized tests, and routinely apply ethical guidelines when choosing, storing, administering, scoring, and interpreting any standardized instrument. The ability to use work samples and evaluation systems in lieu of or in addition to psychometric tests strengthens an evaluator's ability to creatively incorporate techniques into the evaluation experience–and to offer the most comprehensive and valid evaluation possible.

References

Alston, R., & McCowan, C. (1994). Aptitude assessment and African American clients: The interplay between culture and psychometrics in rehabilitation. *Journal of Rehabilitation, 60*(1), 41–46.

Americans with Disabilities Act of 1990 FR 35737 (1991a). Americans with Disabilities Act of 1990, Pub. L. No. 101-336, 104 Stat. (1991b). Anastasi, A., & Urbani, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Berven, N., & Maki, D. (1982). The concept of employed client norms: An application to philadelphia JEVS work samples. *Vocational Evaluation and Work Adjustment Bulletin, 15*(1), 22–28.

Blakemore, T., McCray, P., & Coker, C. (1984). A guide to learning style assessment. Menomonie, WI: University of Wisconsin–Stout, Stout Vocational Rehabilitation Institute, Research and Training Center.

Botterbusch, K. F. (1981). *Work sample norms, reliability and validity*. Menomonie, WI: University of Wisconsin–Stout, Vocational Rehabilitation Institute, Materials Development Center.

Bregman, M. (1969). *Some components in vocational evaluation*. Paper presented at the National Rehabilitation Association Annual Conference, New York, NY.

Brown, C., McDaniel, R., Couch, R., & McClanahan, M. (1994). *Vocational evaluation systems and software: A consumer's guide*. Menomonie, WI: University of Wisconsin–Stout, Stout Vocational Rehabilitation Institute, Materials Development Center.

CARF. (1996). Standards manual and interpretive guidelines for employment and community support services. Tucson, AZ: Commission on Accreditation of Rehabilitation Facilities.

CCWAVES. (1996). Standards and procedures manual for certification in vocational evaluation. Washington, D.C.: Commission on Certification of Work Adjustment and Vocational Evaluation Specialists.

Civil Rights Act, Pub. L. No. 102–166 § 106 (1991).

Colyer, N., & Smith, F. (1993). The role of language and culture in vocational assessment: Factors to consider when reviewing assessment tools. In R. Fry & W. Garner (Eds.), *The issues papers. The Sixth National Forum on Issues in Vocational Assessment* (pp. 57–64). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Connolly, W. (1975). *A Practical guide to equal employment opportunity: Law, principles, and practices* (Vol. I + II). New York, NY: Law Journal Press.

Costello, J., & Corthell, D. (1991). Assessment for community-based employment. In R. Fry (Ed.), *The issues papers. The Fifth National Forum on Issues in Vocational Assessment* (pp. 79–84). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Cronbach, L. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: HarperCollins Publishers.

Cutler, F., & Ramm, A. (1992). Introduction to the basics of vocational evaluation. In J. Siefker (Ed.), *Vocational evaluation in private sector rehabilitation* (pp. 31–66). Menomonie, WI: University of Wisconsin–Stout, Vocational Rehabilitation Institute, Materials Development Center.

Dowd, L. (Ed.). (1993). *Glossary of terminology for vocational assessment, evaluation and work adjustment*. Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Drummond, R. (1996). *Appraisal procedures for counselors and helping professionals* (3rd ed.). Englewood Cliffs, NJ: Merrill.

Dunn, R., Dunn, K., & Price, G. (1979). *The learning style manual.* Lawrence, KA: Price Systems, Inc.

Feist-Price, S., Harley, D., & Alston, R. (1996). A cross-cultural perspective for vocational evaluation and assessment. *Vocational Evaluation and Work Adjustment Bulletin, 29*(2), 48–54.

Gustin, Q., & Petterssen, R. (1978). From vocational evaluation to job training: What can go wrong – a case study. *Vocational Evaluation and Work Adjustment Bulletin, 11*(3), 23– 28.

Hayward, B. J., & Thomas, S. W. (1993). Analysis of a national study on vocational assessment procedures with vocational rehabilitation clients. In R. Fry & W. Garner (Eds.), *The issues papers. The Sixth National Forum on Issues in Vocational Assessment* (pp. 333– 338). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Hayward, B. J., Wine, J., Thorne, J., Stoddard, S., & Wilhite, J. (1992). *Evaluation of vocational assessment procedures and the IWRP process used by state VR agencies*. Research Triangle Park, NC: Research Triangle Institute.

Hume, B. (1973). The MODAPTS approach to vocational evaluation. In R. E. Hardy & J. G. Cull (Eds.), *Vocational evaluation for rehabilitation services* (pp.

167–176). Springfield, IL: Charles C. Thomas.

Joint Committee on Testing Practices. (1988). Code of fair testing practices in education. Washington, D.C.: American Psychological Association.

Kaiser, J., & Modahl, T. (1991). Work sample usage: A different perspective. In R. Fry (Ed.), *The issues papers. The Fifth National Forum on Issues in Vocational Assessment* (pp. 69–71). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Kapes, J., Mastie, M., & Whitfield, E. (1994). *A counselor's guide to career assessment instruments* (3rd ed.). Alexandria, VA: National Career Development Association.

Keyser, D., & Sweetland R. (Eds.). (1984–1994). *Test critiques* (Vol. I-X). Austin, TX: Pro-Ed. Kramer, J., & Conoley J. (Eds.). (1992). *The eleventh mental measurement yearbook*. Lincoln, NE: The Buras Institute of Mental Measurements.

Leconte, P., & Rothenbacher, C. (1987). Cognitive and learning styles in vocational assessment: Importance and practice. In R. Fry (Ed.), *The issues papers. Third National Forum on Issues in Vocational Assessment* (pp. 159–167). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Lesnik, M. (1983). Techniques of vocational evaluation. In R. Lassiter, M. Lassiter, R. Hardy, J. Underwood, & J. Cull (Eds.), *Vocational evaluation, work adjustment, and independent living for severely disabled people* (pp. 30–39). Springfield, IL: Charles C. Thomas.

Lyman, H. (1991). *Test scores and what they mean* (5th ed.). Boston, MA: Allyn and Bacon. Maddox, T. (1997). *Tests: A comprehensive reference for assessments in psychology, education, and business* (4th ed.). Austin, TX: Pro-Ed.

McCray, P. (1980). *Suggested guidelines for evaluating work samples*. Menomonie, WI: University of Wisconsin–Stout, Vocational Rehabilitation Institute, Materials Development Center.

McDaniel, R., & McClanahan, M. (1993). Selecting vocational evaluation tools. In R. Fry & W. Garner (Eds.), *The issues papers. The Sixth National Forum on Issues in Vocational Assessment* (pp. 39–44). Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

McFarlane, F. R., Bellinger, F., Paulsen, A., Wesolek, J., & Modahl, T. (1988). *Consumer's guide to evaluate work samples*. Athens, GA: E & F Vocational Services.

Nadolsky, J. (1973). *Vocational evaluation of the culturally disadvantaged: A comparative investigation of the JEVS system and a model-based system*. Auburn, AL: Auburn University, Department of Vocational and Adult Education.

Neff, W. S. (1985). *Work and human behavior*. New Brunswick, NJ: AldineTransaction.

Owings, S. (1992). Using vocational evaluation in determining employability, wage loss, lost earning capacity, and other aspects of expert witness work. In J. Siefker (Ed.), *Vocational evaluation in private sector rehabilitation* (pp. 169–192). Menomonie, WI: University of Wisconsin–Stout, Vocational Rehabilitation Institute, Materials Development Center.

Power, P. (1991). *A guide to vocational assessment* (2nd ed.). Austin, TX: Pro-Ed. Prediger, D. (1993). Multicultural assessment standards: A compilation for counselors. Alexandria, VA: Association for Assessment in Counseling and Education.

Pruitt, W. A. (1986). *Vocational evaluation* (2nd ed.). Menomonie, WI: Walt Pruitt Associates. Rosenberg, B. (1973). The work sample approach to vocational evaluation. In R. E. Hardy & J. G. Cull (Eds.), *Vocational evaluation for rehabilitation services* (pp. 139–166). Springfield, IL: Charles C. Thomas.

Sankovsky, R. (1969). *State of the art in vocational evaluation: Report of a national survey*. Pittsburgh, PA: Research and Training Center in Vocational Rehabilitation.

Sax, A., & Pell, K. (1985). A primer on tools of evaluation. *Vocational Evaluation and Work Adjustment Bulletin, 18*, 57–60.

Shinnick, M., Black, J., & Decker, R. (1983). Industrial engineering procedures and vocational evaluation. *Vocational Evaluation and Work Adjustment Bulletin, 1*, 24–29.

Siefker, J. (1996). *Tests and test use in vocational evaluation and assessment*. Menimonie, WI: University of Wisconsin–Stout, The Rehabilitation Resource.

Sitlington, P., Neubert, D., Begun, W., Lombard, R., & Leconte, P. (1996). *Access for success: Handbook on transition assessment*. Reston, VA: Council for Exceptional Children, Division on Career Development and Transition.

Smart, J., & Smart, D. (1993). Vocational evaluation of Hispanics with disabilities: Issues and implications. *Vocational Evaluation and Work Adjustment Bulletin, 26*(3), 111–122.

Stout Vocational Rehabilitation Institute. (1977). *Work sample manual format*. Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Super, D. (1949). *Appraising vocational fitness*. New York, NY: Harper & Row.

Supported Employment Technical Assistance Project. (1989).Traumatic brain injury: Supplemental manual *(employment training specialist series)*. Lansing, MI: Michigan State University.

Suzuki, L., Meller, P., & Ponterotto, J. (Eds.). (1996). *Handbook of multicultural assessment: Clinical, psychological and educational applications*. San Francisco, CA: Jossey-Bass.

Tenth Institute on Rehabilitation Services. (1972). *Vocational evaluation and work adjustment services in vocational rehabilitation*. Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Thomas, S. (1986). *Report writing in assessment and evaluation*. Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Thomas, S. (1991). *Vocational evaluation and traumatic brain injury: A procedural manual*. Menomonie, WI: University of Wisconsin–Stout, Materials Development Center.

Thomas, S., Hiltenbrand, D., & Tibbs, S. (1997). A new look at outcomes: Validating the vocational evaluation and work adjustment processes in the contemporary setting. In R. Fry (Ed.), *The issues papers. The Eighth National Forum on Issues in Vocational Assessment* (pp. 117–126). Menomonie, Wi: University of Wisconsin–Stout, The Rehabilitation Resource.

U.S. Department of Labor, Employment and Training Administration. (1991a). *Dictionary of occupational titles* (4th, revised ed.). Indianapolis, IN: JIST Works, Inc.

U.S. Department of Labor, Employment and Training Administration. (1991b). *The revised handbook for analyzing jobs*. Indianapolis, IN: JIST Works, Inc.

Vactor, H., & Hubach, J. (1979). The industrial engineering approach to work evaluation. *Vocational Evaluation and Work Adjustment Bulletin, 4*, 14–17.

Vocational Evaluation and Work Adjustment Association. (1975). Vocational evaluation project final report. *Vocational Evaluation and Work Adjustment Bulletin, 8, Special Edition*.

Wechsler, D. (1981). *WAIS-R manual: Wechsler adult inteligence scale—Revised*.

San Antonio, TX: Psychological Corporation.