

VOCATIONAL ASSESSMENT & EVALUATION  
A PRIMER FOR REHABILITATION  
PROFESSIONALS:

WHAT TO EXPECT AND HOW TO MAXIMIZE YOUR REFERRAL DOLLARS

Michelle Hamilton  
University of Wisconsin-Stout  
CAVEWAS PRE-CONFERENCE  
June 5, 2005  
Shift Happens! Navigating the Changes  
2005 Annual Conference of the  
Canadian Association of Rehabilitation Professionals  
Montreal, Quebec

---

---

---

---

---

---

---

---

---

---

Terminology

SYNONYMOUS?

- Vocational Assessment (VA)
- Vocational Evaluation (VE)

---

---

---

---

---

---

---

---

---

---

Terminology

- VA and VE are NOT synonymous terms
- VA is the broad term used to describe various procedures and methods

---

---

---

---

---

---

---

---

---

---

---

## Assessment Defined

---

Systematic procedures to obtain information from a variety of sources to draw inferences about people

*(Standards for Educational and Psychological Testing, 1999)*

---

---

---

---

---

---

---

---

---

## Vocational Assessment Defined

---

- A general term for the process of identifying and appraising an individual's level of functioning in relation to vocational preparation and employment decision making

---

---

---

---

---

---

---

---

---

## Vocational Assessment

---

Purpose of vocational assessment in rehabilitation

- Plan a course of action
- Enhance client self-knowledge and vocational decision-making abilities
- Predict realistic employment outcomes that result in successful client vocational rehabilitation

---

---

---

---

---

---

---

---

---

## Vocational Assessment

---

### Role of the Rehabilitation Professional

- Collect enough information or 'data' about the client to *diagnose* and make predictive statements about his/her potential to obtain a successful rehabilitation outcome

---

---

---

---

---

---

---

---

---

## LEVELS OF VOCATIONAL ASSESSMENT

---



### Level 1 – Screening/Needs Assessment



### Level 2 - Clinical or Exploratory



### Level 3 – Comprehensive Vocational Evaluation

VEWAA Project 1974

---

---

---

---

---

---

---

---

---

## Vocational Assessment

### Level 1 – Screening/Needs Assessment

---

#### Initial Process

- Needs Assessment
- Determine what is necessary to develop a plan of action (e.g. vocational plan)

#### Level 1 includes:

- Interview with client
- Collection of routine background information (e.g. demographics)
- Reliance on subjective consumer statements
  - Vocational choice/interest
  - Self-estimates of competence
  - Reported work history
- Functional Assessment (PCQ; ABRS)
- Limited, if any, standardized testing (e.g. interest)

---

---

---

---

---

---

---

---

**Vocational Assessment**

**Level 2 – Clinical or Exploratory**

**Intermediate Process**

- In depth exploration or case study approach to the client and vocationally related circumstances. Builds on level 1 information through use of:
  - Additional interviews
  - Collect/analyze documentation (e.g. school records, medical records)
  - Career exploration
  - Vocational and/or adjustment counseling
  - Psychometric/standardized testing
  - Transferable skills analysis
  - Computerized job matching
  - Job analysis and/or environmental assessment
  - Assistive technology considerations

---

---

---

---

---

---

---

---

---

---

**Vocational Assessment**

**Level 3 Vocational Evaluation**

**‘Final’ process**

- Comprehensive individualized holistic process of assessment that utilizes specific instruments, procedures, and behavioral observation
  
- Designed to measure, observe, and document interests, values, temperaments, work-related behaviors, aptitudes, skills, and physical capacities, in order to predict viable employment and/or training outcomes
  
- Used when more in depth information about the client is necessary and not available from information in level 1 or 2

---

---

---

---

---

---

---

---

---

---

**Vocational Assessment**

**Level 3 Vocational Evaluation**

- Real or simulated work used as the focal point of the evaluation
- Systematic observation and recording of work behavior & performance
- Occurs over time and uses multiple methods and techniques to validate findings. Some combination of the following methods are used:
  - Work samples; job samples
  - Situational assessments; community based assessment
  - Standardized Testing
  - Interview
  - Transferable skills; job matching
  - Analysis of background information
  - Career Exploration/job shadow
  - Assistive Technology considerations
  - Prescriptive recommendations

---

---

---

---

---

---

---

---

---

---

---

## Distinction Between VA & VE

- VE is holistic: considers disability/medical, psychological, social, vocational, educational, cultural and economic issues
- Ideally conducted by a CVE
- VE occurs over time (days/week +)
  - evaluated over time with varying work-related demands and environments
- Systematic Behavior Observation and Recording
  - Work performance
  - Work behavior
- Adaptive and/or transferable skills are questionable or unknown
- Work is the specific focus of VE
  - Work samples
  - Situational assessments
  - Community-based assessments
  - Occupational resources
  - Career exploration

---

---

---

---

---

---

---

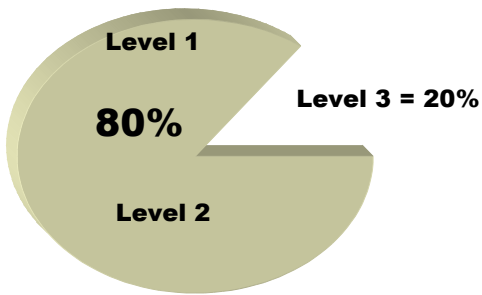
---

---

---

---

## Clients Served in Levels of Assessment (Estimate)



---

---

---

---

---

---

---

---

---

---

---

## Systems that Interface with VR

- Non Profit Organizations (CPA, MS, MD etc)
- CPPD
- Canadian Forces VRP
- LTD Insurance
- Auto Insurance
- Worker's Compensation
- Forensic
- Family law
- Post secondary institutions
- EAP
- HR/Risk Management

---

---

---

---

---

---

---

---

---

---

---

## Choosing the 'Right' Level of Vocational Assessment for Your Needs?

---

Three primary characteristics drive which level of vocational assessment is most appropriate:

- Purpose of referral
- Specific referral questions
- Severity of disability
  - And/or other barriers to employment

---

---

---

---

---

---

---

---

---

---

---

## Purpose of Referral

---

### Referral Source or Systemic Question

What technique gives the most accurate information about a client in the most economical and shortest period of time ?

### VE Practitioner Question

What techniques gives me the most accurate information about a client?

---

---

---

---

---

---

---

---

---

---

---

## Referral for Vocational Assessment

---

- Rehabilitation professionals always have a 'reason' they refer for a VA or VE
- What information do you have?
- What information is needed?
- Responsibility of referral source to articulate what they want or outcome expectation from the VA or VE
  - Purpose for referral
  - Referral questions/objectives

---

---

---

---

---

---

---

---

---

---

---

## Referral Questions

---

Referral questions are critical

- Provide focus and structure to the VA or VE process
- Give the 'evaluator' something to measure against (e.g. specific job or educational program)

Treat referral to VA or VE similar to referrals for any other type of assessment such as:

- Neuropsychological assessment
- Medical restrictions
- Functional capacity evaluation

See appendix "A" example referral questions

---

---

---

---

---

---

---

---

---

---

---

## Disability and/or Barriers to Employment

---

- Disability related functional limitations
- Impact of disability on career development and vocational decision-making
- Stability & prognosis of disability
- Personal Characteristics
  - Motivation
  - Coping skills
- Environmental concerns
  - access; transportation

---

---

---

---

---

---

---

---

---

---

---

## Client Levels of Need

---

Clients who use VR services have varying levels of need. Three most common levels are:

- > Information
- > Instructional
- > Advocacy

---

---

---

---

---

---

---

---

---

---

---

## Information Level Clients

---

- Brief involvement with VR
- Seek information
  - Interests; Transferable skills
  - Assistive Technology
  - Job seeking
- Generalize and apply information to own lives
- Rarely return for additional services, unless they sustain injuries, experience chronic health impairments that necessitate alternative skills or adaptive equipment

---

---

---

---

---

---

---

---

---

---

---

## Instructional Level Clients

---

Instructional level clients constitute the bulk of rehabilitation professional's caseload. Clients at this level of need typically benefit from:

- Vocational guidance & career decision-making
- Postsecondary education and/or specific training
- Access to independent living resources and/or skill training
- Job Seeking Services
  - secure & maintain employment
- Time limited VR service

---

---

---

---

---

---

---

---

---

---

---

## Advocacy Level Clients

---

- Require more extensive VR services
- Clients typically have multiple severe disabilities and/or multiple barriers to employment
- Little or no work experience or work history
  - Questionable transferable or adaptive skills
- Areas of need in independent living, vocational decision-making, goal setting, employment readiness
- Adaptive or Assistive technology
- Benefit from adjustment services and/or improved interpersonal functioning

---

---

---

---

---

---

---

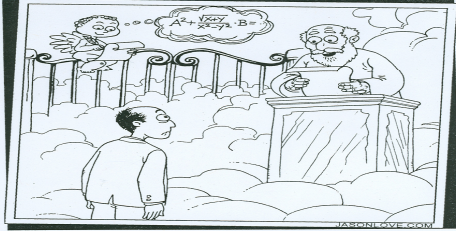
---

---

---

# The Psychology of Psychometrics

## Snapshots



"Well, all of your morals check out. Now we just need you to take this two-part exam..."

---

---

---

---

---

---

---

---

---

---

---

## Psychological Tests

### Test Advantages

- Economical
- Ease & speed of administration
- Most realistic for certain info (literacy, math etc)
- Some are legally mandated (e.g. WISC-R to diagnose DD)
  - Required by the referral source
  - contractually mandated
- Compare client performance to norm group
- Reduce overall assessment or evaluation time

---

---

---

---

---

---

---

---

---

---

---

## Problems in Test Use

3 reasons why tests fail to yield accurate data:

- Poor test selection
- Improper use of the test
  - Overuse
  - Indiscriminate use
  - Learning versus performance
- Test does not measure what practitioner or evaluator intends to assess

---

---

---

---

---

---

---

---

---

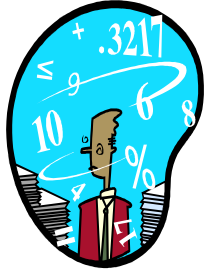
---

---

## Voc Assessment Results (assessment level dependent)

---

- Interest
- Values/Temperament
- Achievement
  - Reading, math, spelling
- Aptitude
- Cognitive/learning abilities
- Learning style
- Transferable Skills
- Adaptive Skills
- Work performance
- Work behaviors
- Functional assessment



---

---

---

---

---

---

---

---

---

---

---

## Test Selection

---

To effectively use psychometric tests need knowledge and competencies in 2 areas:

### Technical Knowledge

- Concepts of tests and measurements
  - Basic statistics
  - Reliability & validity of tests
  - Norm groups

### Practical Knowledge

- Application of test and measurement concepts to REALITY i.e. clients and/or the world of work
  - Limitations of tests
  - Factors that influence test performance
  - Clinical judgment
    - Type I & II errors

---

---

---

---

---

---

---

---

---

---

---

## The Foundation of all Testing: **Reliability**

---

Reliability = repeatability

- How consistently can an individual's test score be reproduced under varying conditions?
  
- Unreliable tests are WORTHLESS!

See appendix "B"

---

---

---

---

---

---

---

---

---

---

---

## Reliability: Standard Error of Measurement (SEM)

---

- Tests are not perfect!
- Variability in test performance
  - Observed score = true score + error score (SEM)
  - SEM represents the inherent error in all tests
- Simply put: the less error, the more reliable the test
- SEM predicts the variability of an individual's score if they took the test a large number of times
  - GATB "G" = SEM ± 6

---

---

---

---

---

---

---

---

---

## Reliability

---



Reliability is established and expressed as a correlation coefficient:

- Degree of relationship between two things
  - GATB "G" score 2 different administrations
- Degree to which test score fluctuates from one test time to another
- Coefficient is a numerical value between +/- 0.00 -1.00
- Closer to 1.00 (+/-) the stronger the relationship (more reliable)
  - 1.00 Perfect positive – as one scores increases other score increases
  - -1.00 Perfect negative – as one score increases other score decreases

---

---

---

---

---

---

---

---

---

## Interpreting Degree of Relationship

---

- Low = .00 - .29
- Moderate = .30 - .59
- High = .70 - 1.00

### **Ideally**

- We want tests with coefficient  $r = .90 +$

### **Reality**

- Many psychological and vocational tests have coefficients of  $r = .40 - .60$

---

---

---

---

---

---

---

---

---

## Validity

---

Validity simply means that a test measures what it says it will measure:

- Math computation test measures math computation

Validity is established by various methods (see appendix "B")

- Content (achievement tests, certification, licensure)
- Construct (
  - Predictive
  - Concurrent)
- Criterion (achievement tests, certification, licensure)

Correlation coefficient expresses the relationship between two

---

---

---

---

---

---

---

---

---

## "Logical" Validity

---

Opinion-based means to establish validity

- Face validity
  - What the test *appears* to measure
  - Cannot be formally tested
  - More our impression or sense
- Strong face validity contributes to client "buy-in"; more sincere effort
- Content validity
  - Panel of 'experts' review content
    - Test content
    - Representative-ness of the test items to the area of interest to be measured

---

---

---

---

---

---

---

---

---

## Empirical Types of Validity

---

### Construct Validity

- Degree to which a test measures a defined construct
  - Constructs are unobservable underlying traits (IQ; personality; love)
  - Can't see the construct directly; often don't all agree on what it is
    - Construct is inferred through specific performance or behavior
  - Hypothesize & describe 'expected' test behavior or performance
  - Difficult to obtain high construct validity; seldom see in test manuals

---

---

---

---

---

---

---

---

---

## Criterion Validity

---

- Assesses whether a test reflects a set of abilities or standard of performance (criterion) in a current or future setting
- Used frequently for achievement tests, certification and licensure
- Correlate test scores results to criterion score/result obtain a correlation coefficient to express degree of relationship between two measures
- Two types of Criterion Validity
  - Predictive – predict to a future criterion
  - Concurrent – predict to criterion at same time or close to time the test is given

---

---

---

---

---

---

---

---

---

## Criterion Validity

---

- Predictive Validity
  - GRE predicts future educational performance
    - students who score well on GRE do better in post secondary education (?)
  - Attainment of B or better grade in Hamilton VE lab class predicts high success/pass rates on the CVE exam (I made this up!!! ☺)
- Concurrent Validity
  - Spatial Relations score correlated well with success in the metal fabricating training for apprentices
  - Correlate results of \*new\* test with old test
    - Weisner and Bennett Mechanical

---

---

---

---

---

---

---

---

---

## Go Figure?

---

### Coefficient of Determination

- $r^2 = .100$
- GRE predictive validity
  - $r = .65$ 
    - $r^2 = .100 = .42$
- We can attribute only 42% of the variation in the person's test score to their future success in college
- Other 58% of score is unknown sources of error

---

---

---

---

---

---

---

---

---

## Validity

---

- Ideal - validity coefficients of  $r = .80$  and above
- Realistically - few have coefficients over  $r = .60$ 
  - $.43$  is about average (Bolton, 1979)

---

---

---

---

---

---

---

---

---

## Validity

---

- *Test* itself is not valid or invalid, per se
  - it is valid or invalid for a specific population in a specific situation for specific person
  - a test that is appropriate to measure average adult IQ is completely invalid to measure IQ of Japanese-speaking child with autism
- A test cannot be valid before it's reliable
- BUT, a reliable test is not necessary a valid test

---

---

---

---

---

---

---

---

---

## Norms

---

- Allow us to compare scores with others who have taken the same test
- Most common way to use norms is to take an individual raw score and convert it into one of several scores more easily understood by others (e.g. percentiles)

---

---

---

---

---

---

---

---

---

## Standardization Sample

---

### Standardization Sample

- AKA norm group or normative sample
- Representative group of people on whom the test/assessment procedure has been standardized
- Group members have at least one common observable characteristic
  - High school seniors who aspire to go College
  - Employed sheet metal workers
- Scores of the group are standardized
- Normal distribution
  - Percentiles; Z-scores; T-scores; Deviation scores; Stanine scores

---

---

---

---

---

---

---

---

---

---

---

## Norm-Referenced Tests

---

Norm-referenced tests yield information about

- Relative standing; measure of relative rank
- Individual's score relative to the scores of others in the norm group who took the same test
- How the test-taker's score compare with the scoring distribution of the standardization sample
- Individual's score dependent on how others score

---

---

---

---

---

---

---

---

---

---

---

## What Norms To Use?

---

Norm group should be appropriate for the individual taking the test

- Representatives in the norm group 'similar' to the test-taker
  - Problematic in rehabilitation as few psychometric tests have PWD in the standardization sample
  - Discrimination results; particularly against PWD that have severe functional limitations as a result of the disability
  - ORIGINAL purpose of development of VE!
- Specialized versus General Norms
- 'Within-group' disability norms have little relevance in most vocational decision-making (Botterbusch,1985).
  - e.g. sheltered employment norms; client norms

---

---

---

---

---

---

---

---

---

---

---

## Norms

---

- Frequently referred to as ‘moving targets’ as an individual’s ‘relative’ rank can shift significantly depending on the norm group



- Michelle scored at the 87<sup>th</sup> percentile in arithmetic (grade 4 students)
- Michelle scored at the 11<sup>th</sup> percentile in arithmetic (grade 10 students)

Your question should always be?

---

---

---

---

---

---

---

---

---

---

---

## Norm Selection

---

Practitioner needs to make a *conceptual separation* about the purpose of the assessment:

- Compare an individual’s performance with members of another group
- Estimate the person’s future performance

---

---

---

---

---

---

---

---

---

---

---

## Real World Performance

---

The world of work expects people to be able to perform some job tasks based on criterion of success or precision rather than compared only to how well others perform

- Expect airline pilots to take off and land an airplane with 100% level of success, not just a little better than their pilot classmates!
- Keyboard 60 wpm
- Expect police officers to know the rules of law enforcement

---

---

---

---

---

---

---

---

---

---

---

## Criterion Referenced Tests & Scores

---

- Criterion-referenced tests have a predetermined external standard or criterion
- Work related criterion generally come from industry standards
  - Number of widgets assembled per hour
  - Keyboarding speed 60wpm 0 errors
  - MTM (Valpar)
- Measures a specific knowledge, skill, or trait and compares person's performance against the criterion
- Focus and importance of on performance rather than relative rank of how others performed

---

---

---

---

---

---

---

---

---

## *Best* Norm Group (when available)

---

- Norm to the environment you are predicting to
- e.g. employed workers, general working population, general population, education, training norms
  - 62<sup>nd</sup> percentile Form & Spatial GATB (GWP)
  - 12<sup>th</sup> percentile MN Paper Form Board (employed drafting technicians)
- Obviously, criterion-referenced tests are better indicators of performance, so if you have a choice between a norm-referenced and criterion referenced measure – choose criterion-referenced.

---

---

---

---

---

---

---

---

---

## Common Scores Used in Reports

---

- **Percentiles** - compares individual's performance to standard or group. Percentiles are most common interpretative data as they are readily understood by most people. (not to be confused with percent).
- **Percentages** raw score generally indicating number right out of total possible (8/10 = 80% correct).
- **Grade Equivalents** - compares individual's performance to average performance of others in a specific grade e.g. a 10.4 Grade Equivalent score in mathematical problem-solving suggests math abilities consistent with average performing students in the fourth month of the tenth grade of math curriculum.

See Appendix C & D

---

---

---

---

---

---

---

---

---

## The Bottom Line

---

Validity - Accuracy  
Reliability – Consistency

- Generalizability of results to employment and training environments outside the VA or VE setting
- Increase confidence in predictions of potential employment and training outcomes

---

---

---

---

---

---

---

---

---

## Interpretation: Prediction of Outcomes

---

- Reliability and validity of interpretations
- Prediction hinges on the ability to accurately ‘generalize’ results to larger populations
  - e.g. employment or training activities
- Confidence in predictions is based on the type of information and data collected
- Not all ‘data’ is of equal significance
  - Mechanical & Clinical Data
  - Degrees of inference

---

---

---

---

---

---

---

---

---

## Types of Data

---

- | Mechanical Data    | Clinical Data   |
|--------------------|---|
| □ Test scores      | □ Interviews  |
| □ Work history/TSA | □ Observations of WP/WB                               |
| □ Medical records  | □ Interpretation of scores                            |
| □ School records   | □ Analysis of other information using clinical skills |
| □ Demographic info |   |

---

---

---

---

---

---

---

---

---

## Clinical Interpretation Skills

---

- Essential clinical skills of a well trained VE practitioner is the ability to meaningfully interpret both the mechanical and clinical data
  
- Trained practitioner will provide functional information about the results of the VA or VE
  
- A VE report should **NEVER** report only test scores

---

---

---

---

---

---

---

---

---

## Functional Interpretation

---

- Report should provide more than just test scores!
  
- Functional interpretation of what scores mean in relationship to client and their vocational goal
  
- Functional interpretation provides 'durable' information that can be used for more than just vocational goal in question
  - Describe math skills that could be matched to a future job

---

---

---

---

---

---

---

---

---

## Mechanical Interpretation

---

- John has a FS-IQ of 78
- John resides in a group home and currently does not work
- John's reading comprehension was at the 4th grade level compared to 25-year old adults
- Compared to employed carpenters, John scored at the 10th percentile on the building a box from a diagram work sample

---

---

---

---

---

---

---

---

---

## Clinical Interpretation

---

- John can add, subtract, multiple and divide one and two digit numbers
- John lives in a supported group home; he is independent in all areas of activities of daily living except for budget and money management.
- John has a driver's license and independently traveled to and from the evaluation in his own vehicle
- John enjoys "building" and has volunteered his time for Habitat for Humanity projects over the past 2 years
- Compared to EMPLOYED carpenters, John was at the 10th percentile in his ability to review simple directions and follow a diagram to accurately select 2 x 4's, position them in proper locations, use tools to correctly nail the 2 x 4's into the desired formation.
- John is able to quickly and accurately compare and match sequences of numbers and letters up to and including 12 digits/letters. John has the potential to learn the basic job tasks required to stock or 'pick' inventory in a warehouse setting. OJT is the suggested training method.

---

---

---

---

---

---

---

---

---

---

---

## Triangulation of Evidence

---

- Triangulation combines multiple observations, theories, methods, and empirical materials to 'test' out hypotheses
  - (hypotheses in VA or VE = client employment outcome)
- Triangulation helps overcome the weakness or intrinsic biases and the problems that come from single method, single-observer, and single-theory approach.
- Triangulation enhances generalizability, enhances confidence in employment and outcome predicts

---

---

---

---

---

---

---

---

---

---

---

## Three Degrees of Inference

---

### Lowest degree of inference

- a single sample of behavior with consideration of the context behavior where occurred
  - Interest test results

### Moderate degree of inference

- information, observations, items are interrelated; identify consistencies (patterns) and generalizations
  - Interest test
  - Work history
  - Aptitude profile

### Highest degree of inference

- a hypothetical construct (e.g. computer drafting technician, general learning ability, spatial aptitude, motivation, depression) may be used to describe the essence of the patterns, consistencies or generalizations
    - Aptitude & achievement scores
    - Interest test
    - Expressed voc goal
    - Work history
    - Observations of WB/WP
- } agreement among results increases confidence in prediction

---

---

---

---

---

---

---

---

---

---

