

Identifying and Developing Predictors of Job Performance

Patrick J. Curtin

Deborah L. Whetzel

Kenneth E. Graham

Caliber Associates

Pre-Conference Workshop, IPMAAC

June 22, 2003

Agenda

- Introductions
- Test Plan
- Cognitive Ability
- Training and Experience
- Biodata
- Interviews

Introduction

- Who are we?
- Who are participants?

Test Plan – Overview

- Summarizes information from job analysis and literature review of assessment methods
- Provides a rationale for the use of particular test and assessment methods
- Provides a blueprint for the development and validation of tests and assessments

Test Plan - Process

- Steps
 - Conduct job analysis to identify KSAs
 - Identify potential measurement methods
 - Assess selection methods against psychometric and operational criteria
 - Create a KSA by method matrix
 - Determine how best to assess relevant KSAs

Test Plan – Centerpiece is the Matrix

- The matrix is used to link KSAs to measurement methods
 - KSAs are rows
 - Measurement methods are columns
 - Rater judgments are in the cells

Test Plan – Matrix Example

KSA ↓ / Selection Method	Cognitive Ability	Training and Experience	Biodata	Interviews
Knowledge of business management (such as quality control, human resource management)				
Knowledge of math including addition, subtraction, multiplication, and division				
Skill in dealing with all types of people				
Skill in managing own time and coordinating with others				
Ability to have an impact on others in the organization and display energy and leadership				
Ability to communicate information and ideas orally so others will understand				

Test Plan – Identifying KSAs

- Knowledge - A body of information applied directly to the performance of a function.
- Skills - A present, observable competence to perform a learned psychomotor skill.
- Abilities - A present competence to perform an observable behavior or a behavior which results in an observable product.

Example KSAs

- Example Managerial Knowledges
 - Knowledge of business management (such as quality control, human resource management)
 - Knowledge of math including addition, subtraction, multiplication, and division
- Example Managerial Skills
 - Skill in dealing with all types of people
 - Skill in managing own time and coordinating with others
- Example Managerial Abilities
 - Ability to have an impact on others in the organization and display energy and leadership
 - Ability to communicate information and ideas in speaking so others will understand

Test Plan – Identifying measurement methods

- Example measurement methods:
 - Cognitive ability
 - Training and experience
 - Biodata
 - Interviews

Test Plan

- Methods need to be assessed against Psychometric & Operational criteria
- Psychometric criteria
 - Reliability evidence
 - Validity evidence
 - Content validation support
 - Subgroup differences

Test Plan

- Operational criteria
 - Applicant acceptance
 - Resistance to compromise
 - Consistency of administration and scoring
 - Development cost
 - Development time
 - Operational cost

Test Plan - Selection Tools by Criteria Example

EVALUATION OF SELECTION TOOLS AGAINST CRITERIA														
Criteria ↓	Selection Tools													
	Measures of Basic Attributes		Measures of Background, Training, and Experience					Measures of Declarative Knowledge and Procedural Knowledge and Skill						
	Cognitive Ability	Personality	Accomplishment Record/Portfolio	Interview	Biodata	GPA	Task-based Questionnaire	Job Knowledge	SJT	Work Sample/Job Tryout	Role Play	Leaderless Group	In-basket	Essay
Validation Evidence	H	M	M	M	M	L	M	H	H	H	L	L	M	L
Content Validation Support	M	M	H	H	M	L	H	H	H	H	H	H	H	M
Reliability	H	H	M	M	M	H	H	H	M	H	M	M	M	M
Subgroup differences	H	L	L	L	L	H	M	M	M	L	L	L	M	H
Applicant acceptance	M	L	M	H	L	M	M	H	H	H	H	M	M	M
Resistance to compromise	H	L	H	M	M	H	M	H	M	H	L	M	M	M
Consistency of administration and scoring	H	H	M	M	H	H	H	H	H	M	L	L	M	M
Development cost	L	L	M	M	H	L	L	M	H	H	H	M	H	M
Development time	L	L	M	M	H	L	L	M	H	H	H	M	H	M
Operational cost	L	L	M	H	L	L	L	L	L	H	H	H	H	M

Test Plan Exercise: Completing the Matrix

KSA ↓ / Selection Method	Cognitive Ability	Training and Experience	Biodata	Interviews
Knowledge of business management (such as quality control, human resource management)		✓		✓
Knowledge of math including addition, subtraction, multiplication, and division	✓			
Skill in dealing with all types of people		✓		✓
Skill in managing own time and coordinating with others			✓	✓
Ability to have an impact on others in the organization and display energy and leadership				✓
Ability to communicate information and ideas orally so others will understand		✓		

Test Plan – Final Review

- Do the selection methods provide “coverage” of the KSAs?
- Are the Psychometric and Operational aspects of the chosen methods appropriate given the restrictions of a particular testing scenario?
- In what order should the methods be used?

Test Plan - Compromises

- To balance among competing concerns such as adverse impact and operational costs, the following should be considered:
 - Multiple hurdle approach
 - Whole job measurement approach

Test Plan – In Summary

- The final test plan
 - identifies suitable measurement methods given a host of concerns and limitations
 - demonstrates the expected relationships between KSAs and measurement methods.
 - provides a blueprint for the development and validation of personnel selection procedures.

Cognitive Ability

- Definition
 - Cognitive ability has been conceptualized in more than one way. The two main conceptualizations are based on cognitive ability being thought of as a unified whole “g” or thought of as a collection of intelligence facets.

Cognitive Ability

- Examples of Off-the-shelf Cognitive Ability Tests
 - The Wonderlic Personnel Test
 - Employee Aptitude Survey – a basic skills battery
 - Watson-Glaser Critical Thinking Appraisal
 - Wechsler Adult Intelligence Scale III
 - Raven's Progressive Matrices
 - Stanford-Binet Intelligence Scale
 - Woodcock Johnson III Tests

Cognitive Ability

- Psychometric Characteristics
 - Reliabilities can range from the mid .80s to the low .90s (Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Cognitive Ability

- Psychometric Characteristics
 - Validity of cognitive ability tests to predict performance is approximately .51 (Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Cognitive Ability

- Psychometric Characteristics
 - Subgroup differences -- A difference of approximately one standard deviation has been repeatedly found between African Americans and White subgroups on cognitive ability testing (Hunter and Hunter, 1984; Jensen, 1980; Schmitt, Clause, & Pulakos, 1996).

Cognitive Ability

- How to Select a Cognitive Ability Test
 - Mental Measurements Yearbooks (University of Nebraska Press)
 - Tests in Print (University of Nebraska Press)
 - Test Critiques (Pro-Ed – Austin, TX)
 - Tests (Pro-Ed – Austin, TX)
 - A.P.A. (www.apa.org)
 - Test Publishers

Cognitive Ability

- Issues to Consider
 - Administration time
 - Special accommodations
 - Retesting policy

Cognitive Ability

- Reasons for developing a new cognitive ability test
 - Alternate forms are required
 - Tests become outdated
 - Client wants a proprietary test
 - Test for newly hypothesized or specialized ability

Cognitive Ability

- Steps for developing a new cognitive ability test
 - Develop test specifications
 - Create items
 - Conduct technical reviews
 - Conduct sensitivity reviews
 - Create test administration instructions
 - Conduct pilot test of items
 - Analyze item data
 - Select items and assemble test
 - Write documentation

Training and Experience

- Definition
 - Training and experience (T&E) selection instruments are based on the notion that past behavior and experience are powerful predictors of future behavior (Ash, Johnson, Levine, & McDaniel; 1989).

Training and Experience

- Past behavior and experience captured by T&E instruments include job-relevant tasks, job performance, training, and overall knowledge, skills, and abilities (KSAs) that an individual may have gained from previous jobs.

Training and Experience

- Holistic Judgment
 - Not a formally scored T&E evaluation method but rather a more general overview of an individuals resume.
 - Lacks both structure and objectivity.

Training and Experience

- Traditional Point Method
 - Consists of a mechanical formula set out in a formal schedule.
 - The number of months or years of different types of relevant training, education, and experience are allotted a certain number of points.

Training and Experience

- Improved Point Method
 - Examines an applicants' specific job behaviors/duties as indicators of job-related KSAOs.
 - Less arbitrary and a more adequate approach to assigning fairly specific point scores to applicants based on their individual backgrounds.

Training and Experience

- Grouping Method
 - Applicants are divided into a small number of groups on the basis of simultaneous consideration of training and experience.
 - Rather than being ranked over a continuous range of scores, applicants assigned to each group are assigned to the same score.

Training and Experience

- Behavioral Consistency Method
 - Has been referred to as the accomplishment record method (Hough, 1984 cited in Ash et al., 1989).
 - Rank order applicants on the basis of the type of achievement behaviors that are necessary for optimum performance in the target job.

Training and Experience

- Self-rating T&E Rating Method
 - Composed of three different methods: Task-based, KSAO-based, and The Illinois Job Element Method.
 - Require applicants to evaluate and judge their own past tasks that they have performed or KSAOs that they have acquired.

Training and Experience

- Psychometric Characteristics
 - Past research has found interrater reliability estimates of T&E ratings were found to be around .80, which is quite desirable (Gatewood and Field, 1988).

Training and Experience

- Psychometric Characteristics
 - Meta-analyses yielded the following average estimated validities: behavioral consistency (.45), self-rating KSAO (.20), self-rating task (.15), and point method (.11).
 - These results, with the exception of the point method, are based on few studies (McDaniel, Schmidt, and Hunter, 1988).

Training and Experience

- Psychometric Characteristics
 - Because the method yields higher scores to those applicants with the largest amount of work experience and education; females, minorities, and young applicants may not have had the time or opportunity to accumulate as many experience and training points as white, male and older applicants.

What is an Experience Questionnaire?

- Task-based questionnaire (TBQ)
 - Measures experience with tasks from job analysis
 - Assumes experience performing job tasks
 - Not usually appropriate for entry-level jobs
 - Minimal written communications skills required
 - Applicants indicate experience performing job tasks

What is an Experience Questionnaire?

- KSA based questionnaire (KSABQ)
 - Measures experience as indicator of KSAs.
 - No need for experience with specific tasks
 - Can be used for entry-level jobs
 - Minimal written communications skills required
 - Applicants indicate experience performing activities related to the job.

Steps to Develop TBQ

- Identify tasks
 - Create counterfeit items (faking)
- Develop scoring system
- Pilot test form

Identify Tasks

- Include tasks that are:
 - Critical
 - Needed-at-entry
- Limit TBQ to tasks that qualified applicants could have experience performing.

Develop Scoring System

- Award points for increasing amounts of experience performing tasks.
 - Number of times.
- Reflect the asymptotic relationship between experience and performance.
 - Estimate asymptotes based on job analysis and MQ data.
 - SMEs review and adjust asymptotes.
 - Assign points linearly below asymptote.

Example Scoring System

Points Associated with Levels of Experience	
Points	Level of Experience
0	0% of performance asymptote
1	20% of performance asymptote
2	40% of performance asymptote
3	60% of performance asymptote
4	80% of performance asymptote
5	100% of performance asymptote or more

Pilot Test TBQ

- Examine readability:
 - Flesch-Kincaid
 - Linguist
- SMEs participate in “think-aloud” workshop.
- SMEs try to identify counterfeit tasks.

Challenges

- Faking
- Collecting accurate/reliable information
- Setting performance asymptote
- Measuring education/training-based experience

Faking

- Three levels of protection:
 - Certification of information accuracy
 - Statement certifying veracity of responses.
 - TBQ lists penalties (e.g., dismissal) for falsification.
 - Verifiers
 - References that can support accuracy of information.
 - Counterfeit items
 - Nonsensical tasks that applicant could not perform.
 - Include multiple opportunities to “correct” responses.

Collecting Accurate/ Reliable Information

- Number of times is best predictor, but hard to estimate.
 - Calculate from frequency and duration.
 - More verifiable.
 - Difficult when tasks/activities vary in frequency.

Collecting Accurate/ Reliable Information

Skill in repairing clocks				
Please rate your experience with each of the activities below:	How much work experience do you have performing this activity?			Who can verify your work experience with this activity?
	Duration		Frequency	
	Years	Months		
1. Replacing the batteries in clocks.			① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
2. Replacing the hands on a clock.			① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
3. Replacing the gears in a clock.			① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩

Collecting Accurate/ Reliable Information

- Number of times is best predictor, but hard to estimate.
 - Provide anchoring examples.
 - More flexible.
 - Less verifiable.
 - Applicants may overlook them.

Collecting Accurate/ Reliable Information

Skill in Repairing Clocks					
Activity	Frequency	Examples			Verifier(s)
	How many times have you performed this activity? (Mark the option closest to your level of experience).	<i>To perform this activity the number of times listed in the option to the left, you'd have to perform the task...</i>			Who can verify your work experience with this activity? (Mark all corresponding verifier numbers that apply)
		<i>every day for about...</i>	<i>or every week for about...</i>	<i>or every month for about...</i>	
1. Replacing the batteries in a clock.	0 times	--	--	--	
	19 times	1 month	4 months	2 years	① ⑥
	38 times	2 months	9 months	3 years	② ⑦
	58 times	3 months	13 months	5 years	③ ⑧
	77 times	4 months	18 months	6 years	④ ⑨
	96 times or more	5 months or more	22 months or more	8 years or more	⑤ ⑩

Setting Performance Asymptote

- Difficult to estimate asymptote:
 - Some tasks may not have one.
 - Interpersonal KSAs.
 - Complex or rapidly changing jobs.
 - Many different routes.
 - Experience.
 - Education/training.
 - Hybrid (e.g. apprenticeships).
 - Does not account for individual differences.
 - Intelligence.
 - Openness to experience.

Measuring Education/Training-Based Experience

- Difficult to equate education/training:
 - Different schools.
 - Different instructors.
 - Different course materials.
 - Person X situation interaction.
- Collect information on activities *performed* as part of education or training.

Training and Experience

- Strengths and Weaknesses
 - The holistic judgment and traditional point methods, which are the two most used methods tend to result in decisions that lack validity.
 - Behavioral consistency, improved point method, and grouping methods will be effective once organizations decide to use these in place of the holistic judgment and traditional point methods. Cost more to develop and implement than the other procedures due to increased job analysis data requirements.

Biodata

- Definition

- Biodata consists of background information from various life history domains, including education, work experiences and accomplishments, developmental experiences, hobbies, and social interactions.
- Biodata is typically elicited through the use of questionnaire items, which ask people to recall and report typical behaviors and experiences that have occurred in the past.
- The focus of biodata is to assess a pattern of historical experiences that are *manifestations* of predictor constructs (e.g., self confidence).

Biodata

- Example item for quality and quantity of work

About how many hours in a day do you spend in constructive work?

- a. Less than 8
- b. 8-10
- c. 11-13
- d. 14-16
- e. More than 16

Biodata

- Example item for quality and quantity of work

How often do you do a task over and over again until it is perfect?

- a. always
- b. often
- c. sometimes
- d. almost never
- e. never

Biodata

- Example item for problem solving

About how many non-fiction books have you read in the past year?

- a. none
- b. one
- c. two
- d. three
- e. four or more

Biodata

- Example item for problem solving

How often have you invented something to serve a needed purpose?

- a. never
- b. one or two times
- c. three to five times
- d. six to ten times
- e. more than ten times

Biodata

- Scaling methods
 - Empirical scaling procedures select and weight items based on their ability to differentiate membership in higher and lower-performing criterion subgroups (Hogan, 1994)

Biodata

- Scaling methods
 - Rational scaling procedures involves developing items based on an individual difference variable (based on job analysis). Item selection is based on judgments of relevance of item to construct (Hough & Paullin, 1994).

Biodata

- Psychometric Characteristics
 - Empirical scales – test-retest reliability .60-.96
 - Rational scales – internal consistency low .70s; coefficient alpha .66-.78

Biodata

- Psychometric Characteristics
 - Empirical scales – Meta-analysis .36 with Ability to perform; .34 with Performance of duties
 - Rational scales – Initial validities in low .40s; shrinkage roughly .05-.10.

Biodata

- Psychometric Characteristics
 - Low adverse impact
 - But may be negatively biased due to range restriction (pre-selection)
 - Correlation with cognitive ability .51

Biodata

- How to generate items
 1. Identify the purpose of the instrument.
 2. Precisely define the criteria of interest (e.g., performance ratings, absenteeism, turnover).
 3. Specify the variables to be measured by identifying job tasks/behaviors and knowledge, skills, abilities, motivation, and other characteristics necessary to effectively perform job behaviors.

Biodata

- How to generate biodata items
 4. Develop items that reflect hypotheses about the influence of prior experience, behaviors, and/or situations on the knowledge, skills, abilities and other important characteristics needed for successful job performance.

Biodata

- How to generate biodata items
 5. A second panel of SME's should screen the items and eliminate those that are not objective, not relevant, or prone to socially desirable responding. Experienced item developers who are not familiar with the items should review them at this stage.

Biodata

- How to generate biodata items
 6. The final instrument will require 8–15 items per construct, there should be a sufficient number of items at the pretest stage to evaluate.
 7. Administer screened items to a try-out sample of 100 – 200.

Biodata

- How to generate biodata items
 8. Conduct statistical screening of items using the following criteria
 1. Items must show sufficient variability (standard deviation of at least .80)
 2. Item distributions must be consistent with the hypotheses for which they were generated.
 3. A coefficient alpha of .50 or above is suitable for a rationally constructed biodata scale. Support for the reliability of the scales may be demonstrated by lack of shrinkage across samples.

Biodata

- Faking controversy
 - Correcting for social desirability does not improve criterion-related validity of personality scales (Ones, Viswesvaran & Reiss, 1996).
 - Criterion-related validity decreases when people are told to fake (Graham, McDaniel, Douglas & Snell, 2002).

Employment Interview

- Definition
 - Interviews involve the exchange of information between an interviewer and an interviewee, either face to face, by telephone, or in writing.

Employment Interview

- Types of Interviews
 - Structured
 - Unstructured

Employment Interview

- Types of Questions
 - Situational
 - Past behavior

Employment Interview

- To improve quality of evaluation
 - Base questions on a job analysis
 - Ask exact same questions
 - Do not allow questions from candidate till end of interview
 - Rate each answer or use multiple scales
 - Take notes
 - Use multiple and same trained interviewers across candidates

Employment Interview

- Administration modes
 - Telephone interviews
 - Validity of a telephone-administered interview is comparable to face-to-face interviews.
 - Require no travel for the interviewer or interviewee and can result in significant cost savings for geographically dispersed employers.

Employment Interview

- Administration modes
 - Written structured interviews
 - Candidates provide written responses to questions; administered in a group setting
 - Require no travel for the interviewer or interviewee and can result in significant cost savings for geographically dispersed employers.

Employment Interview

- Psychometric Characteristics
 - Reliability .67 structured interviews
 - Reliability .84 job-related structured interviews
 - Reliability .68 job-related unstructured interviews

Employment Interview

- Psychometric Characteristics
 - Validity
 - Structured .44
 - Unstructured .33
 - Situational .50
 - Job-related .39
 - Psychological .29

Employment Interview

- Psychometric Characteristics
- Correlations with other constructs
 - High structure interviews focused more on applied mental skills, direct job knowledge, applied social skills and organizational fit.
 - Low structure interviews focused more on general mental ability, background credentials, some aspects of personality, and physical attributes.

Employment Interview

- Psychometric Characteristics
 - Research on subgroup differences has yielded mixed results

Employment Interview

- Psychometric Characteristics
- Incremental validity beyond GMA
 - Higher levels of structure result in greater incremental validity of GMA than lower levels of structure
 - Correlation of .40 between interviews and GMA shows that 16% of variance in interview represents cognitive ability

Employment Interview

- How to develop structured interviews
 1. SMEs review KSAs and describe incidents of effective and ineffective performance in writing.
 2. SMEs sort incidents into KSAs and rate the behaviors described in incidents for effectiveness (retranslation).
 3. Select incidents allocated to KSAs by 75% of SMEs and that have low SD on effectiveness (< 1.0).

Employment Interview

- How to develop structured interviews
 4. Write questions from the incidents retained from Step 3.
 5. SMEs think of people whose performance on the job they would rate as outstanding, average, and poor, and provide written narratives describing how those people would respond to each question.

Employment Interview

- How to develop structured interviews
 6. SMEs then rate the simulated responses and those responses on which there is a high degree of rater agreement are retained as anchors for a rating scale.

Summary - what we covered

- Questions
 - Test Plan
 - Cognitive Ability
 - Training and Experience
 - Biodata
 - Interviews

Identifying and Developing Predictors of Job Performance

Patrick J. Curtin

Deborah L. Whetzel

Kenneth E. Graham

Caliber Associates

Pre-Conference Workshop, IPMAAC

June 22, 2003