PREDICTORS AND TESTING in Personnel Selection
Personnel Psychology – subfield of I/O psychology focusing on the management of human resources

- Recruitment ✓
- **Selection**
- Placement
- Psychometrics ✓
- **Individual differences**
- Criterion development
- Performance appraisal

- Training and development
- **Team selection and training**
- Legal issues - Equal Employment Opportunity, Title VII, CRA 1991
Applicant/employee flow through and interactions with various PERSONNEL systems
• why do we use tests and implement selection systems?
  ▪ low selection ratios
  ▪ cost of full probationary selection models

• predictor/test ⇨ any assessment tool that is used to make inferences about [future] job-related behavior
  ▪ some measure of individual differences
What is a predictor?

- any variable used to forecast a criterion
  - personality
    - FFM
  - self-efficacy
  - general mental ability/cognitive ability
  - integrity
  - creativity
whereas specific predictors and their sequencing is variable, the typical selection process may take the form illustrated below:
Psychometric properties of predictor or test scores

• “Whenever a worker in psychology or education desires to measure some quality in a group of individuals, he faces the problem of choosing the best instrument for his purpose” –Thorndike, 1955
Psychometric properties of predictor or test scores

- reliability of scores
- validity of scores
Reliability

• consistency or stability of scores obtained from a measure
• consistency of scores obtained by the same person when examined with the same test (or equivalent forms) on different occasions, time, places, etc.
• measurement error!!
• reliability, like validity, is based on correlations
Methods for assessing the **reliability** of test scores

- test-retest (temporal consistency/stability)
- alternate-form (temporal consistency/stability, and inter-form consistency/equivalence)
- split-half (internal consistency)
- coefficient alpha (Cronbach’s alpha) (inter-item [and also internal] consistency)
- scorer or inter-rater reliability or agreement
Methods for assessing the reliability of test scores

• test-retest reliability (temporal consistency/stability)
  - involves the repeated administration of the same test to the same sample
    - intervening events
      - time lag
      - exposure to the test or measure
      - maturation
      - experience
      - temporally unstable constructs (i.e., mood)
• alternate-form reliability (temporal consistency/stability, and inter-form consistency/equivalence)
  ▪ a measure of the extent to which 2 separate forms of the same test are equivalent
    – difficult to construct 2 tests
    – impact of nonequivalent forms in high-stakes testing
• split-half, odd-even (or random split) reliability (internal consistency)
  ▪ the primary issue here is one of obtaining comparable halves
    – issues?
    – Spearman-Brown correction
• Coefficient alpha (Cronbach’s alpha) (inter-item [and also internal] consistency)
  ▪ these are measures of inter-item consistency, (i.e., the consistency of responses to all items on the test)
  ▪ b/c average of all possible split-halves is equal to the coefficient alpha, they are also considered to be measures of internal consistency
  – issues??
issues

- indication of the extent to which each item on the test measures the same thing as every other item on the test
- the more homogeneous the domain (test), the higher the inter-item consistency
• scorer or inter-rater reliability or agreement
  ▪ extent to which 2 or more raters are consistent or agree
  ▪ also extent to which same rater is consistent over time (intra-rater reliability)
    – issues??
Validity

• the extent to which a test measures what we designed it to measure, all of what we designed it to measure, and nothing but what we designed it to measure

• self-evident measurement tools

• dubious measures
  – require evidence
Validity

• paper-and-pencil measure of agreeableness
  ▪ marking statements on a paper is far removed from actually exhibiting agreeableness
  ▪ must establish inferential links between marks on the page and agreeableness
    • measures as record of a sample of behavior
Validity

• appropriateness of **inferences** drawn from scores

• for specified use, it is an assessment of **WHAT** the test measures and **HOW WELL** it does so

• validity depends not only on the predictor, but also the criterion
  ▪ agreeableness and moving violation tickets
  ▪ openness and training outcomes
  ▪ conscientiousness and job performance
Strategies for assessing the validity of test scores (i.e., validation techniques/strategies)

- criterion-related
- content-related
- construct-related
- meta-analysis & validity generalization
- synthetic validity/j-coefficients
- face validity
Predictor Measure

Criterion Measure

Criterion-related validity

Content-related validity

Construct-related validity

Construct validity

Predictor Construct Domain

Performance Domain

Job analysis

Performance Construct Domain

Job analysis

Criterion-related validity

Content-related validity

Construct-related validity

Construct validity
Predictor Measure

Criterion Measure

Criterion-related validity

Content-related validity

Construct-related validity

Construct validity

Job analysis

Job analysis

Predictor Construct Domain

Performance Domain

PSYC 353 11C predictors & testing 10/18/11 [Arthur]
Criterion-related validity

• effectiveness of a test in predicting an individual's behavior in specific situations
  - test score is intended as an indicator or predictor of some other behavior (that typically will not be observed until some future date)
  - test or predictor score is correlated with a criterion (i.e., a direct and independent measure of that which the test is designed to predict)
Criterion-related validity

- criterion-related, validity coefficients can range from $-1.0$ to $+1.0$
- absolute value is used to compare different validity coefficients
• criterion-related validation designs
  ▪ predictive
  ▪ concurrent
  ▪ postdictive
<table>
<thead>
<tr>
<th>Predictor Measure</th>
<th>Criterion Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion-related validity</td>
<td></td>
</tr>
<tr>
<td>Content-related validity</td>
<td></td>
</tr>
</tbody>
</table>

**Construct-related validity**

- Predictor Construct Domain
- Performance Domain

**Job analysis**

- PSYC 353 11C predictors & testing 10/18/11 [Arthur]
Content-related validity

- effectiveness of a test in predicting an individual's behavior in specific situations
  - adequacy with which a specified content domain is sampled
  - degree to which a predictor covers a representative sample of the behavior being assessed (e.g., knowledge-based tests such as classroom tests)
Content-related validity

- systematic examination of test content to determine whether it covers a representative sample of the behavior domain being measured
- typically rational and nonempirical, in contrast to criterion-related validity which is empirical
- content domain to be tested should be fully described in advance in very specific terms
Predictor Measure

Criterion Measure

Criterion-related validity

Content-related validity

Construct validity

Predictor Construct Domain

Performance Domain

Job analysis

Construct-related validity

4
Construct-related validity

• extent to which the test may be said to measure a theoretical construct or trait
  - **convergent validity**—different measures of the same construct should be correlated or related to each other
  - **discriminant validity**—different measures of different constructs should not be correlated or related to each other
## Multi-Trait/ Multi-Method Matrix (MTMM) approach

<table>
<thead>
<tr>
<th>Trait</th>
<th>Paper-and-pencil</th>
<th>Assessment Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>NEO-FFI</td>
<td>Assessor Ratings</td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>Wesman Personnel Classification Test</td>
<td>Assessor Ratings</td>
</tr>
</tbody>
</table>
• face validity
• meta-analysis and validity generalization
Summary of $r_{xx}$ and $r_{xy}$

- $r_{xx} = \text{property of scores and not tests}$
- choice of specific $r_{xx}$ metric depends on:
  - facet of consistency of interest
  - construct of interest
  - resources available
- good level of $r_{xx}$?
  - $\geq .80$ for applied decision making
  - $\geq .70$ (or .65?) for research purposes
- $r_{xx} = \text{a necessary but not sufficient condition to justify or support specified test use}$
Summary of $r_{xx}$ and $r_{xy}$

- a test can be reliable but not valid
- however, a test **cannot** be valid but not reliable—a test that does not correlate with itself cannot be expected to correlate with anything else
- thus, reliability is a necessary but not sufficient condition for validity
  - reliability sets the upper limit of validity
    - square root of a test’s reliability is the upper limit of its validity
Summary of $r_{xx}$ and $r_{xy}$

- test with unknown reliability and validity is to be avoided
- reliabilities **cannot** be negative; but validity coefficients can be negative
- $r_{xy} = $ use of the test and not an inherent property of the test
  - appropriateness of inferences
Some characteristics/features of tests and inventories (self-report measures)

- difference b/n tests and inventories (self-report measures)
- features and characteristics
  - speed vs. power
  - individual vs. group
  - paper-and-pencil, performance, computer-administered, video-based, oral — [stimulus and response]
  - adaptive tests (usually computerized)
  - typical vs. maximal performance
  - objectively vs. subjectively scored
  - sign vs. sample
  - proctored vs. unproctored
  - construct vs. method
### Construct/Method Matrix

<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>C1</td>
<td>A</td>
</tr>
<tr>
<td>C2</td>
<td>E</td>
</tr>
<tr>
<td>C3</td>
<td>I</td>
</tr>
<tr>
<td>C4</td>
<td>M</td>
</tr>
<tr>
<td>PREDICTORS</td>
<td>VALIDITY</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Work sample tests</td>
<td>0.54</td>
</tr>
<tr>
<td>GMA tests</td>
<td>0.51</td>
</tr>
<tr>
<td>Employment interviews (structured)</td>
<td>0.51</td>
</tr>
<tr>
<td>Peer ratings</td>
<td>0.49</td>
</tr>
<tr>
<td>Job knowledge tests</td>
<td>0.48</td>
</tr>
<tr>
<td>T &amp; E behavioral consistency method</td>
<td>0.45</td>
</tr>
<tr>
<td>Job tryout procedure</td>
<td>0.44</td>
</tr>
<tr>
<td>Integrity tests</td>
<td>0.41</td>
</tr>
<tr>
<td>Employment interviews (unstructured)</td>
<td>0.38</td>
</tr>
<tr>
<td>Assessment centers</td>
<td>0.37</td>
</tr>
<tr>
<td>Biographical data measures</td>
<td>0.35</td>
</tr>
<tr>
<td>Conscientiousness tests</td>
<td>0.31</td>
</tr>
<tr>
<td>Reference checks</td>
<td>0.26</td>
</tr>
<tr>
<td>Job experience (years)</td>
<td>0.18</td>
</tr>
<tr>
<td>T &amp; E point method</td>
<td>0.11</td>
</tr>
<tr>
<td>Interests</td>
<td>0.10</td>
</tr>
<tr>
<td>Years education</td>
<td>0.10</td>
</tr>
<tr>
<td>Graphology</td>
<td>0.02</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

**Mean**

- **mean**: 0.33
- **SD**: 0.16

<table>
<thead>
<tr>
<th>Construct</th>
<th>Computer-based</th>
<th>Paper-Pencil</th>
<th>Questionnaire</th>
<th>Interview</th>
<th>Assessment Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples of some commonly used constructs and methods in personnel selection

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledge</strong></td>
<td><strong>Ability</strong></td>
</tr>
<tr>
<td>Job knowledge</td>
<td>General mental</td>
</tr>
<tr>
<td>Education/training</td>
<td>Verbal</td>
</tr>
<tr>
<td>Experience/work history</td>
<td>Quantitative/numerical</td>
</tr>
<tr>
<td><strong>Skill</strong></td>
<td>Psychomotor</td>
</tr>
<tr>
<td>Communication</td>
<td>Perceptual</td>
</tr>
<tr>
<td>Reading</td>
<td>Spatial</td>
</tr>
<tr>
<td>Writing</td>
<td>Mechanical</td>
</tr>
<tr>
<td>Listening</td>
<td>Speed of information processing</td>
</tr>
<tr>
<td>Speaking</td>
<td>Memory</td>
</tr>
<tr>
<td>Critical thinking</td>
<td>Attention</td>
</tr>
<tr>
<td>Social perceptive</td>
<td>Physical</td>
</tr>
<tr>
<td>Organizing and planning</td>
<td></td>
</tr>
<tr>
<td>Time management</td>
<td><strong>Attitudinal</strong></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Values</td>
</tr>
<tr>
<td>Leadership</td>
<td>Drive/motivation</td>
</tr>
<tr>
<td>Persuasion/negotiation</td>
<td>Vocational interests</td>
</tr>
<tr>
<td><strong>Personality</strong></td>
<td>Hobbies/other interests</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
</tr>
<tr>
<td>Emotional stability</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
</tr>
<tr>
<td>Locus of control</td>
<td></td>
</tr>
<tr>
<td>Integrity</td>
<td></td>
</tr>
<tr>
<td><strong>MULTIPLE CONSTRUCT METHODS</strong></td>
<td></td>
</tr>
<tr>
<td>Biodata or Biographical information blanks</td>
<td></td>
</tr>
<tr>
<td>Person-organization fit</td>
<td></td>
</tr>
<tr>
<td>Background and credit checks</td>
<td></td>
</tr>
<tr>
<td>References and letters of recommendation</td>
<td></td>
</tr>
<tr>
<td>Minimum qualifications</td>
<td></td>
</tr>
<tr>
<td>Drug testing</td>
<td></td>
</tr>
</tbody>
</table>
### General overview of the assessment center process

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Participants work through exercises.</td>
</tr>
<tr>
<td>2.</td>
<td>Assessors observe and record behaviors, focusing on the participants to whom they have been assigned (see Figure 4). Assessors are typically assigned to observe two participants.</td>
</tr>
<tr>
<td>3.</td>
<td>Assessors sort and organize observations and may generate initial dimension-level scores for the exercise.</td>
</tr>
<tr>
<td>4.</td>
<td>After all the exercises are completed, assessors meet (typically for hours) to generate participants' dimension-level scores. See the Scoring section of this chapter for a detailed description of this process.</td>
</tr>
<tr>
<td>5.</td>
<td>Depending on the purpose of the assessment center, assessors may meet with and provide one-on-one feedback to each participant. Participants may also later receive a formal, written feedback report.</td>
</tr>
</tbody>
</table>
Some common constructs measured by assessment centers

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>.33</td>
</tr>
<tr>
<td>Consideration/Awareness of others</td>
<td>.25</td>
</tr>
<tr>
<td>Drive</td>
<td>.31</td>
</tr>
<tr>
<td>Influencing others</td>
<td>.38</td>
</tr>
<tr>
<td>Organizing and planning</td>
<td>.37</td>
</tr>
<tr>
<td>Problem solving</td>
<td>.39</td>
</tr>
</tbody>
</table>
Example of situational judgment test item

You have to be at work by 7:30 a.m. It is typically a 20-minute drive to work. You leave your apartment at 7:00 a.m., go down to start your car, and it does not start.

- **response instructions**
  - what *would* you do?
    - behavioral tendency
  - what *should* you do?
    - knowledge-based

- **response formats**
  - select the best and worst
  - rate effectiveness of each option
  - rank order options in terms of effectiveness
Example of situational judgment test item

An employee has been repeatedly mistreated by a coworker. Rate the effectiveness of each action in terms of what the employee should do in response to the situation.

- Take items the mean coworker is responsible for handling, to get them in trouble
- Remove valuable personal items from the mean coworker’s area
- Stop working with the mean coworker
- Keep a returned item and give it to the mean coworker to get on their good side
- Stop socializing with the mean coworker
Some common constructs measured by situational judgment tests
(Christian, Edwards, & Bradley, 2010)

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job knowledge and skills</td>
<td>.19</td>
</tr>
<tr>
<td>Interpersonal skills</td>
<td>.25</td>
</tr>
<tr>
<td>Teamwork skills</td>
<td>.38</td>
</tr>
<tr>
<td>Leadership</td>
<td>.28</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.24</td>
</tr>
</tbody>
</table>
Interviews

- interviews are predictor methods that are able to measure various constructs
Interviews

• interviews are both predictor methods and recruitment and socialization tools
Interviews

• degree of structure
  ▪ questions
  ▪ answers

• types of questions
  ▪ situational, hypothetical, future-oriented
  ▪ experience and behaviorally-based
Interviews

- structure increases validity
- experience-based vs. situational
- interviewer experience
- contrast effects, rating errors
- personal factors
- note-taking
Interviews

- structured interviews = .51 (validity)
- unstructured interviews < .20 (validity)
- applicant also collects information (RJP)
- 2 key functions
  - fills in information gaps
  - assess factors only measurable in face-to-face interaction
- first impressions
- verbal and nonverbal cues
Some common constructs measured by interviews
(Huffcutt, Conway, Roth, & Stone, 2001)

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Overall</th>
<th>Low structure</th>
<th>High structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>General intelligence</td>
<td>.24</td>
<td>.26</td>
<td>.11&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Job knowledge &amp; skills</td>
<td>.42</td>
<td>.49</td>
<td>.33</td>
</tr>
<tr>
<td>Extroversion</td>
<td>.33</td>
<td>.22</td>
<td>.40</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.33</td>
<td>.24</td>
<td>.37</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.51</td>
<td>.25&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.53</td>
</tr>
<tr>
<td>Openness</td>
<td>.16&lt;sup&gt;A&lt;/sup&gt;</td>
<td>–</td>
<td>.16&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>.47</td>
<td>.18&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.56</td>
</tr>
<tr>
<td>Communication skills</td>
<td>.26</td>
<td>.05&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.31</td>
</tr>
<tr>
<td>Interpersonal skills</td>
<td>.39</td>
<td>.31</td>
<td>.40</td>
</tr>
<tr>
<td>Leadership</td>
<td>.47</td>
<td>.11&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.18&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>General physical attributes</td>
<td>-.18&lt;sup&gt;A&lt;/sup&gt;</td>
<td>-.18&lt;sup&gt;A&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>Job-related physical skills</td>
<td>.15&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.10&lt;sup&gt;A&lt;/sup&gt;</td>
<td>.19&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note. <sup>A</sup>Sample-weighted correction only.
Good Interview Practices

- link questions to job analysis
- ask same questions of each candidate
- anchor rating scales
- use interview panels w/ multiple interviewers
- combine ratings mechanically
- provide training to interviewers
- document development procedures
- institute feedback system to interviewers
Interview Questions That Should **NOT** be Asked

- Are you married?
- Do you have children?
- Who takes care of your children?
- What does your spouse do?
- How long have you been divorced?
- Do you have credit cards?
- How old are you?
- Have you ever been arrested?
- What is your religious affiliation?
- What is your family background? Do you have brothers and sisters?
- Have you ever received a dishonorable discharge from the military?
ADA-Related **UNLAWFUL** Interview or Hiring Questions

- Have you ever been treated for any of the following conditions?
- Have you ever been hospitalized? If so, for what conditions?
- Have you ever been treated for any mental condition?
- Is there a health-related reason you may not be able to perform the job for which you are applying?
- Have you had a major illness is the past 5 years?
- Do you have any physical conditions that preclude you from performing certain kinds of work?
- Are you taking any prescribed drugs or medications?
- Have you ever been treated for drug addiction or alcoholism?
Faking

• potential threat to all nonability, self-report measures
• response distortion, socially desirable responding
  ▪ self-deception
  ▪ impression management
• individuals can increase scores by .5 $SD$
• can be reduced with instructions and a host of design features as well
• does not seem to significantly adverse the validity
• can faking be a good thing?
Keep your profile private and set appropriate filters.
Facebook has a plethora of filters to control who can see what information about you, although many users are unaware of some of these privacy features (Cain, 2008). Some faculty members now conduct internet searches of graduate school applicants prior to sending out interview invitations — and have chosen not to invite certain applicants because of what they found. There are reports of employers using Facebook as a hiring tool, as well as stories of student suspensions and arrests due to information posted on social networking sites (Brady, 2006). Students, colleagues, mentors, and clients will Google you, and it is up to you to monitor what information the internet has to offer.

• select-in vs. select-out predictors
Evaluation of Predictors

• reliability
• validity
• fairness
• applicability
• cost
• utility
Sources of Information About (Standardized) Tests

• Mental Measurement Yearbook
• Tests in Print
• Publisher Catalogs
• Comprehensive Handbook of Psychological Assessment
Ethical Standards in Testing

- APA Guidelines
- APA’s *Standards for Educational and Psychological Testing*
- SIOP’s *Principles for the Validation and Use of Personnel Selection Procedures*
- Invasion of privacy
  - collect only job relevant information that is not intrusive
- Confidentiality — who should have access to the test results?