Insuring Content Validity in a Litigious Environment

2010 IPAC Conference – Making the Case for Assessment

Newport Beach, CA

Kyle E. Brink

St. John Fisher College

Jeffrey L. Crenshaw

Brian L. Bellenger

Centrus Personnel Solutions

Types of Validity Evidence

Uniform Guidelines

- Types of validity studies
  - Content
  - Criterion-oriented
    - Predictive
    - Concurrent
  - Construct
    - Convergent
    - Discriminant

APA Standards & SIOP Principles

- Sources of validity evidence
  - Test content
  - Relations to other variables
    - Test-criterion relationship
      - Predictive vs. concurrent
    - Convergent & discriminant
    - Validity generalization
  - Response processes
  - Internal structure of the test
  - Consequences of testing

Content Validity

- Data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. (UGESP)

- Evidence based on test content may include logical or empirical analyses that compare the adequacy of the match between test content and work content, worker requirements, or outcomes of the job. Test content includes the questions, tasks, format, and wording of questions, response formats, and guidelines regarding administration and scoring of the test. (Principles)

Content Validation Steps

- Job analysis

- Test development

- Test administration

- Test assessment

- Data analysis & scoring

Required Documentation (UGESP)

- Users of selection procedures should maintain and have available for each job information on adverse impact
- If selection process has adverse impact, must maintain and have available validity evidence including:
  - User(s), location(s) and date(s) of study
  - Problem and setting
  - Job analysis - Content of the job
  - Selection procedure and its content
  - Relationship between the selection procedure and the job
  - Alternative procedures investigated
  - Uses and applications
  - Contact person
  - Accuracy and completeness

Required Documentation (Principles)

Reports of validation efforts should include enough detail to enable a researcher competent in personnel selection to know what was done, to draw independent conclusions in evaluating the research, and to replicate the study. The following information should be included:

- Identifying Information
- Statement of Purpose
- Analysis of Work
- Search for Alternative Selection Procedures
- Selection Procedures
- Relationship to Work Requirements
- Criterion Measures (When Applicable)

- Research Sample
- Results
- Scoring and Transformation of Raw Scores
- Normative Information
- Recommendations
- Caution Regarding Interpretations
- References

Job Analysis Defined

- Systematic
- process of discovery of the nature of a job
- by dividing it into smaller units,
- where the process results in one or more written products with the goal of describing what is done in the job or what capabilities are needed to effectively perform the job
  - Brannick, Levine & Morgeson (2007)

Job Analysis Overview

- Methods of collecting data

- Sources of data

- Content & unit of analysis

Job Analysis Methods

- Methods of gathering job analysis data
  - Review of prior information
  - Observation
  - Interview
  - Panel/focus group
  - Questionnaire

Job Analysis Methods

- What are the advantages and disadvantages to using multiple methods of job analysis for a particular job?

- Utilizing multiple methods strengthens the job analysis and increases legal defensibility
(Thompson & Thompson, 1982; Veres, Lahey & Buckly, 1987)

Job Analysis Methods

Job description/specification

- Product of job analysis
  - Do not provide the depth or breadth of information needed to support content validity

- **Are often out of date** (Barrett, 1998; Noe, Hollenbeck, Gerhart & Wright, 2008)

- Rarely documentation regarding their development

- Not a substitute for job analysis!

Job Analysis Sources

- Job analyst

- Job incumbents

- Supervisors

- Subject matter experts (SMEs)

- How many?

Job Analysis Content & Unit of Analysis

- Work behaviors (duties) and/or tasks
  - Not too general, not too specific, but just right
- Knowledge
  - Operationally defined as a body of learned information required for performing work behaviors
- Skills & abilities
  - Operationally defined in terms of work behavior

Identifying KSAOs

- Always start job analysis with O*NET
  - See O*NET Content Model
    - https://cliffie2.sjfc.edu/exchweb/bin/redir.asp?URL=http://www.onetcenter.org/dl_files/ContentModel_DetailedDesc.pdf
  - Abilities: based on Fleishman taxonomy
  - Knowledge & skills: not sufficiently specific
- Fill in gaps with job analysis methods/sources

Rating Behaviors & KSAs

- Work behaviors and/or tasks
  - Importance/criticality
  - Frequency
  - Needed upon entry
- Knowledge, skills & abilities
  - Importance/criticality
  - Needed upon entry
  - Distinguishing value
  - Memorization (knowledge only)
- Behavior to KSA linkage

Common Issues in Job Analysis

- Jobs evolve
- Assuming job description/specification is accurate, current, and sufficient
- Failure to sufficiently customize off-the-shelf job analyses
- Failure to rely on multiple methods/sources
- Failure to gather linkage, needed at entry, and memorization ratings
- SME capabilities
- SME motivation & resistance

Test Development

- ## What do we mean by "valid" test?
  - Accuracy of prediction
    - The extent to which performance on the measure is associated with performance on the job
  - Accuracy of measurement
    - Degree to which a measure truly measures the attribute it is intended to measure

Source: Heneman & Judge (2009)

Accuracy of Measurement
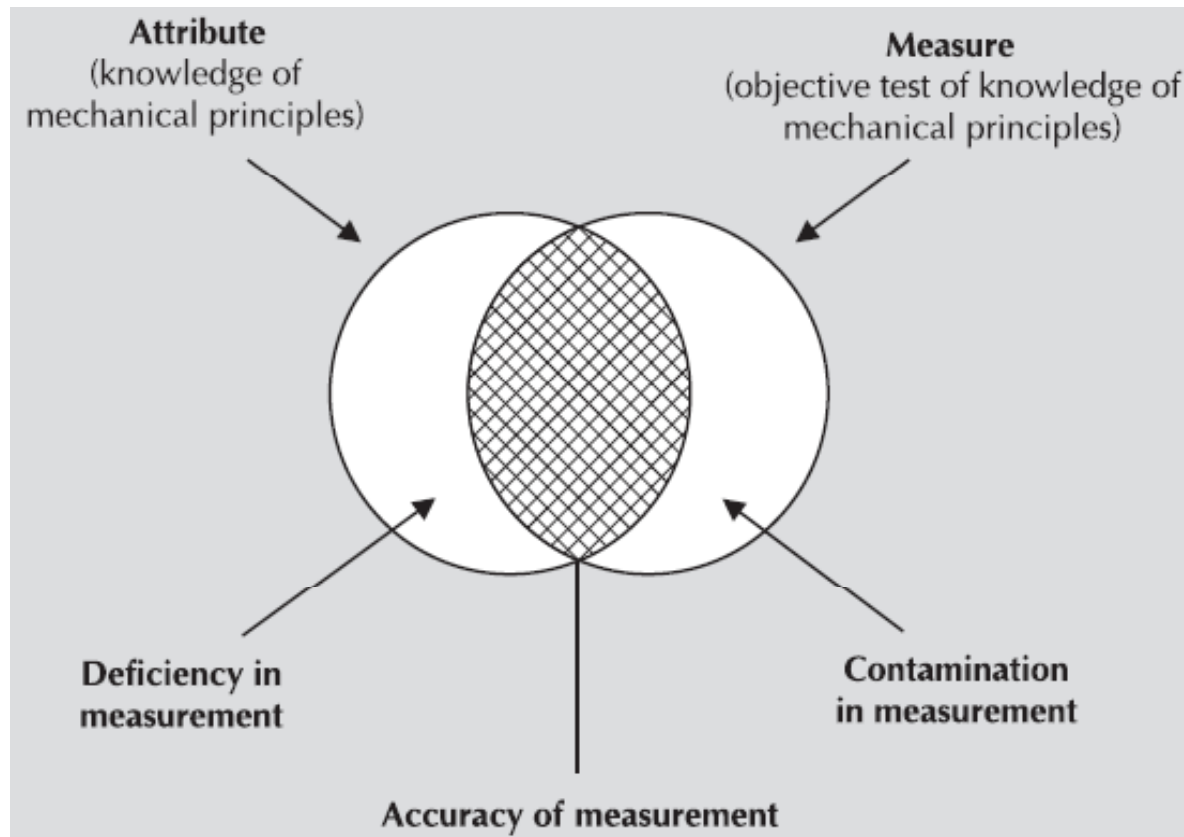
Source: Heneman & Judge (2009)

Test Development

- What do we mean by "content valid" test?
  - Representative of the job
  - Resemble the job

- How do we develop a content valid test?
  - Review job analysis results
  - Develop a selection plan
  - Gather critical incidents
  - Develop exercise/questions, benchmarks, and rating scale

Test Development

- Review job analysis results (duties/tasks & KSAs)
  - Only assess what is important and needed upon entry
  - If assessing knowledge, appropriately use open vs. closed book testing depending on whether knowledge is referenced vs. memorized
    - Failing to assess knowledge that can be referenced is a systematic bias in content

Test Development

- Develop the selection plan
  - What are the most important KSAs?
  - What is the best method (alternative measures) for assessing these KSAs?
    - Adverse impact
    - Validity
    - Face validity/applicant reactions
    - Correlations with other predictors/incremental validity
    - Utility
    - Relationship to the job duties
    - Others?

Test Development

- Gather critical incidents

1. Choose a duty or KSA upon which to base the critical incident (*e.g., customer service*)

2. What was the critical incident?
   - *While waiting tables at a restaurant, a server did not place a customer's order into the computer system right away. After 15 minutes the customer questions the server as to why their food had not arrived. At this point the server realized the mistake. As a consequence, the customer left no tip for the server.*

Test Development

- **Gather critical incidents** (continued)

3. What was the situation leading up to the critical incident?

   – *The server was waiting tables at a restaurant. The restaurant was very busy and two of the scheduled servers called in sick that morning.*

4. What were the actions or behaviors of the person of interest in the incident?

   – *The server did not place the customers order immediately.*

5. What were the results or outcomes of those actions?

   – *The server angered the customer and received no tip.*

6. What level of performance does this event represent?

Test Development

- ## Gather critical incidents (continued)

7. Develop a question/exercise based on this critical incident

    – *Suppose you are waiting tables at a restaurant and you forgot to place a customer's order. After 15 minutes, the customer asks you why the food has not arrived. At this point you realize your mistake. What would you do? What would you say to the customer?*

8. Develop a scoring rubric/benchmarks based on this question and critical incident

Test Development

- **Develop questions, benchmarks, and rating scales**
  - Make sure they are: related to the job, relevant to the scenario/question, related to the KSA being measured, and behaviorally-based
  - Examples:
    - Apologized for the error
    - Confirmed that they had the order correct
    - Arranged for a free appetizer to be brought out while the customers waited

- **Once developed:**
  - Have SMEs review the test
  - Have SMEs document (e.g., via ratings) their judgments regarding the validity of the exam

Test Administration

- Lack of consistency in treatment of applicants is a major factor contributing to discrimination
- Must be standardized
  - Equal opportunity to prepare for test
  - Equal access to preparation/orientation materials
  - Identical test content
  - Identical test administration
  - Opportunity for appeal
- Problems
  - Unstructured interview
  - Orientation only during 1 shift
  - Variability in role-players
  - Variability in interviewers/administrators

Test Administration

- Video and computer interviews offer
  - Cost savings
  - Increased standardization

Situational Interview Question

- Suppose you are a Bailiff and the trial you are working at is on a short recess. You hear voices beginning to get louder and louder outside the back of the courtroom so you walk over to see what is going on. You see that a man and woman are arguing loudly. The woman is accusing the man of lying and trying to get the kids. The man is accusing the woman of trying to take his money. The exchange is very heated and they are saying mean things. You also observe the woman shove the man.

- After observing this situation, what would you do?
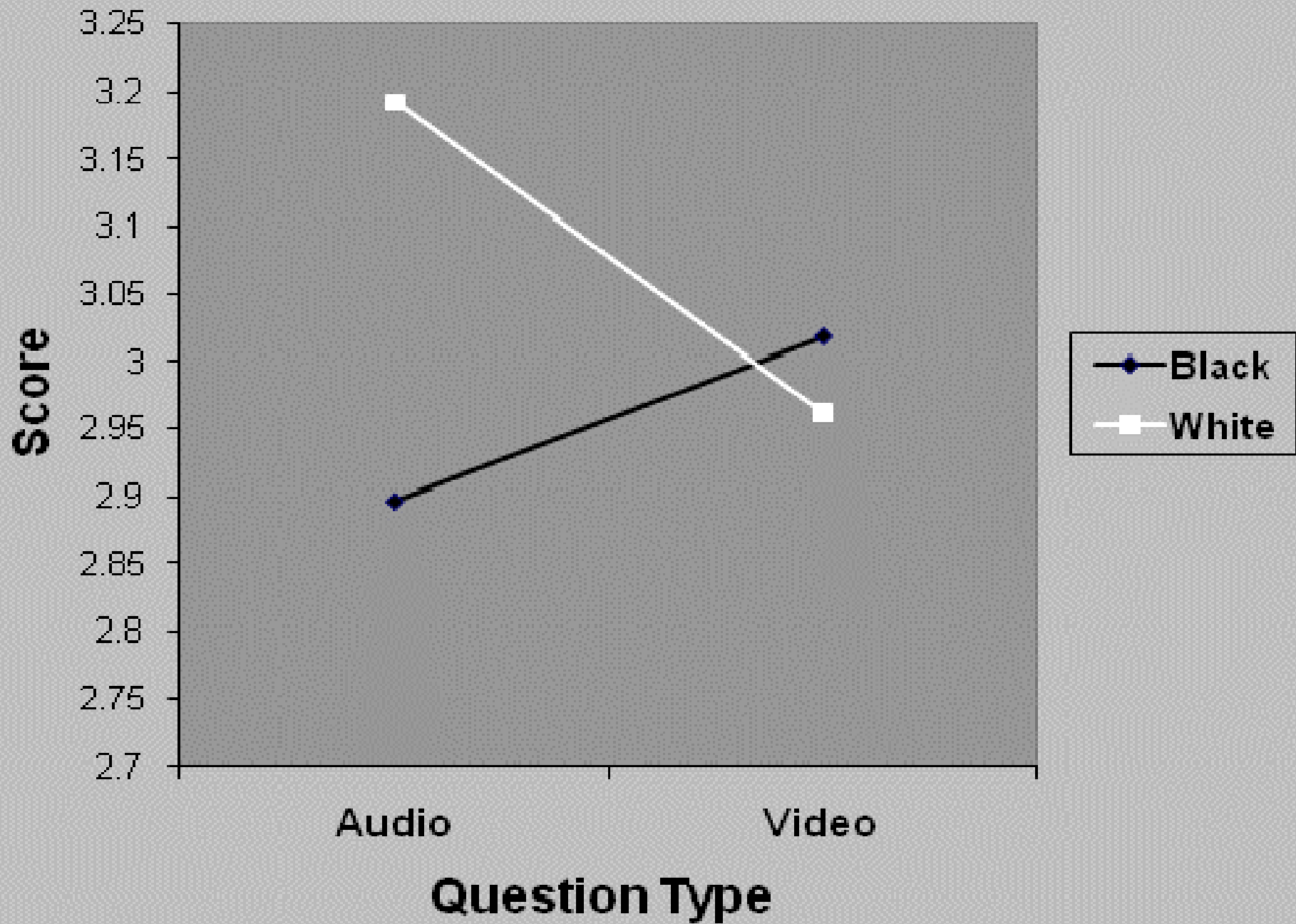
Test Assessment

- Ensure standardized assessment
  - Identical scoring rubric/evaluation criteria across all candidates

- Ensure reliable and accurate ratings
  - Assessor training
    - How to assess candidates
    - Familiarity with questions, rating scales
    - Consistency in evaluation
  - Procedural issues
    - Panels (number of assessors)
    - Internal v. external
    - Demographics of assessors?
    - Rotation of assessors?
    - Quality checks?

Data Analysis & Scoring

- **Minimize human error**
  - Data entry, data analysis, data interpretation
- **Report**
  - Central tendency & dispersion
  - Item analyses (difficulty, discrimination)
  - Group differences & adverse impact

Data Analysis & Scoring

- Reliability: Definition
  - The degree to which a measure of physical or cognitive abilities, or traits, is free from random error
    - Actual score = true score + error
  - Consistency, dependability, or stability of measurement of an attribute
    - A measure is reliable to the extent it provides a consistent set of scores to represent an attribute

Data Analysis & Scoring

- Sources of unreliability
  - Test content
    - Item sampling
  - Test structure
    - Chance response tendencies, guessing, item format
  - Standardization of test administration
    - Standardization of measure
    - Situational or environmental variables
    - Person variables
  - Standardization of assessment
    - Rater error

Data Analysis & Scoring

- **Estimating reliability**
  - Test-retest reliability
    - Concerned with stability of measurement
  - Coefficient alpha
    - Concerned with interrelationships among items
  - Interrater agreement
    - Concerned with agreement among raters

Data Analysis & Scoring

- ## Reliability standards (Nunnally & Bernstein, 1994)
  - Early stages of research $r_{xx} > .70$
  - If making important decisions w/ respect to specific test scores ("high stakes testing")
    - $r_{xx} \geq .90$ is the bare minimum
    - $r_{xx} \geq .95$ should be considered a desirable standard

Data Analysis & Scoring

- **Reliability implications**
  - As reliability increases,
    - More confidence in individuals' obtained scores
    - More confidence in differences between individuals' scores
- **Proof (FYI):**
  - Standard error of measurement
    - Since only one score is obtained from an applicant, the critical issue is how accurate the score is as an indicator of an applicant's true level of knowledge/ability

    $$SEM = SD \sqrt{1 - r_{xx}}$$

    - As reliability increases, SEM decreases
    - There is a 95% chance a persons true score falls between $\pm$ 2 SEM
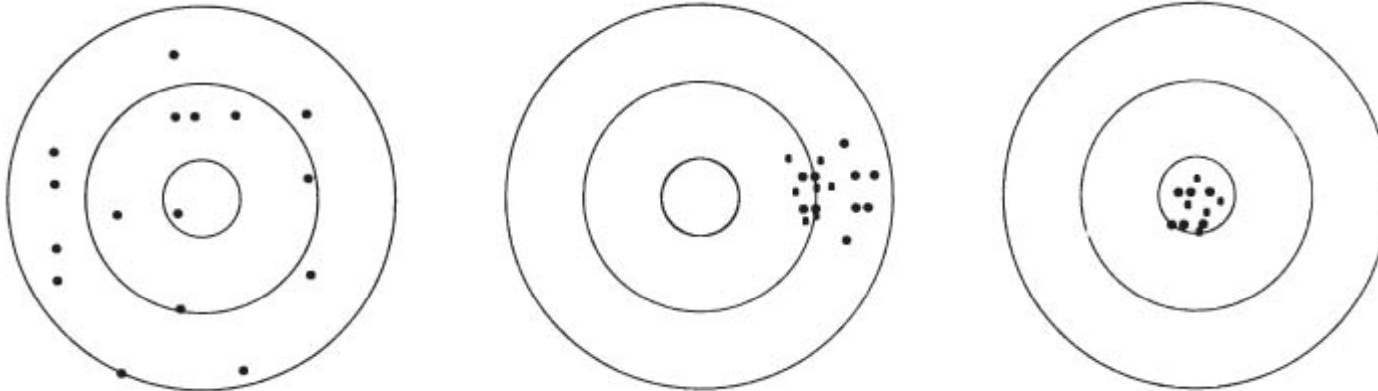
Data Analysis & Scoring

- **Reliability implications**
  - Relationship to validity
    - Reliability of a measure places an upper limit on the possible validity of a measure
    - Reliability does not guarantee validity - it only makes it possible
    - A highly reliable measure is not necessarily valid

$$r_{xy} = r_{x\infty y\infty}\sqrt{r_{xx}r_{yy}}$$

  - Proof (FYI):
    - r_{xy} represents the empirically determined validity coefficient
    - $r_{x\infty y\infty}$ represents the correlation between the theoretical true predictor (i.e., the test) and true criterion scores (e.g., job performance)
    - r_{xx} represents the reliability of the predictor
    - r_{yy} represents the reliability of the criterion

Reliability & Validity

Data Analysis & Scoring

- Methods for combining and weighting items and predictors
  - When should a compensatory model be used? When should a multiple hurdles model be used?

- Critical vs. cut score
  - What are the positive and negative consequences of using a high predictor cut score?

- Top-down vs. random vs. banding
  - What are the advantages of ranking vs. random selection?

Questions?

References & Source Materials

- American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.
- Barrett, R. S. (1996). *Fair employment strategies in human resource management*. Westport, CT: Quorum Books.
- Barrett, R. S. (1998). *Challenging the myths of fair employment practices.* Westport, CT: Quorum Books.
- Biddle, D. (2006). *Adverse Impact and Test Validation* (2nd ed.). Hampshire, England : Gower Publishing Limited.
- Brannick, M. T., Levine, E. L., & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Cascio, W. F. (1998). *Applied Psychology in Human Resource Management* (5th Edition). Prentice Hall, Upper Saddle River, NJ.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). *Uniform guidelines on employee selection procedures.* Federal Register, 43, 38290-38315.
- Gatewood, R. D., Feild, H. S., & Barrick, M. (2008). *Human resource selection* (6th ed.). Mason, OH: South-Western Cengage Learning.

References & Source Materials

- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Heneman, H. G., & Judge, T. A. (2009). *Staffing organizations* (6th ed.). Madison, WI: Mendota House/Irwin.
- Noe, R. A., Hollenbeck, J., Gerhart, B., & Wright, P. (2008). *Human resource management: Gaining a competitive advantage* (6th ed.). New York: McGraw-Hill.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Phillips, J. M., & Gully, S. M. (2009). *Strategic Staffing.* Upper Saddle River, NJ: Pearson Prentice Hall.
- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH: Society for Industrial and Organizational Psychology, Inc.
- Thompson, D. E., & Thompson, T. A. (1982). Court standards for job analysis in test validation. *Personnel Psychology, 35*, 865-874.
- Veres, J. G., Lahey, M. A., & Buckly, R. (1987). A practical rationale for using multi-method job analyses. *Public Personnel Management, 16,* 153-157.