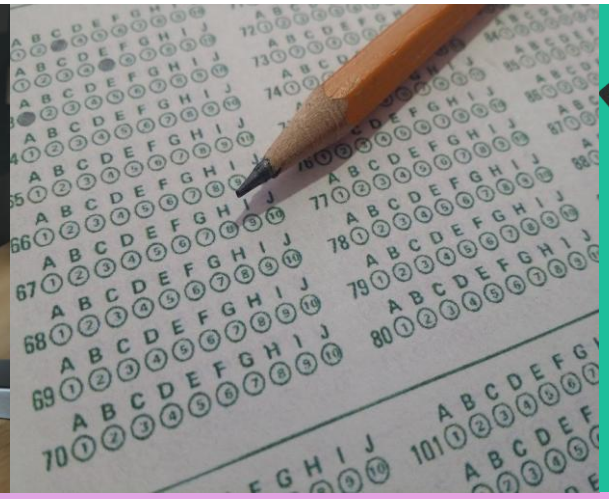
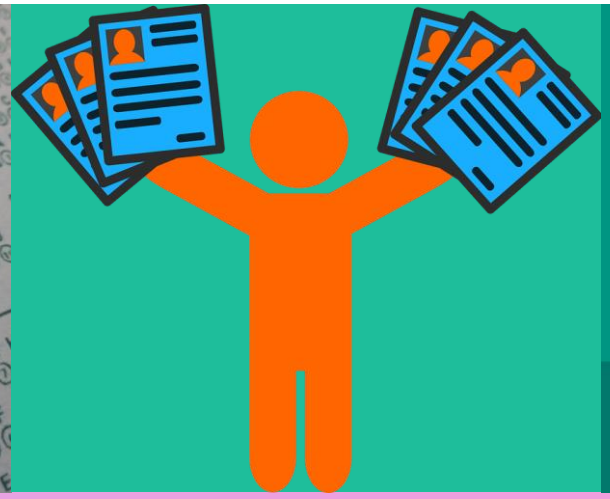
Employment Tests: Validity & Validation

OFCCP Annual Enforcement Conference

September 2023

Agenda

- What is a test?
- Job analysis
- Psychometric properties
- Types of validity
- Aggregation
- Physical requirements
- Transported validity
- Examining reports
- Interesting trends


Introduction

Purpose of this presentation is to provide you with a foundation in understanding what validity is and why it is important.

This presentation is NOT intended to be exhaustive of the subject.

Background on Validity

- What is a test?
- Job analysis
- Psychometric properties

What is a test?

- A test is an objective and standardized procedure for measuring a psychological construct using a sample of behavior. (Conte & Landy, 2018)
- An assessment systematic method and procedure for ascertaining work-related knowledge, skills, abilities, or other characteristics of an individual or group of individuals, or the performance of an individual or group of individuals. (ISO, 2020)
- An assessment is a test.

Job analysis

- Job analysis is a process for determining the essential job duties of a particular job title.
- The purpose of a job analysis:
  - Document the important job duties,
  - How they are performed, and
  - Human attributes are necessary to successfully perform the job duties.
- Required attributes include:
  - Performance expectations (properties of the job in the context of the organization's expectations) and
  - The abilities, knowledge, experience, skill, and personal characteristics necessary to meet those expectations.
- *Job analysis should be the first step in developing a test.*

Psychometric properties of tests

- Reliability (Consistency in scores)
  - Test-Retest
  - Interrater reliability
  - Internal consistency
- Validity
  - Refers to how well a test measures what it was intended/designed to measure.
  - A test is valid if it accurately measures the construct or trait it was supposed to measure.
- Standardization
  - Refers to the process of developing uniform procedures for administering and scoring a test.
  - Standardization guarantees that everyone is given the same instructions, has fair time limits, and has the same criteria used to score their tests.
    - Uniform Guidelines on Employee Selection Procedures (UGESP)

Types of Validity

- Criterion
- Content
- Construct

Criterion validity

- Criterion validity is the extent to which a test predicts with a specific outcome/criterion (e.g., a job performance duty).
- To determine if a test has criterion validity, data is collected for both the test scores and the criterion measure (e.g., job performance ratings).
  - A correlation coefficient describes the relationship between the test scores and the criterion measure.
  - The validity coefficient ranges from -1.00 to +1.00, with higher values indicating a stronger relationship.
  - If the correlation is positive and statistically significant, it indicates that the test is a valid predictor of the criterion measure.
- There are two types of criterion validity: concurrent validity and predictive validity.
  - Concurrent validity is assessed when the test and criterion measure are collected at the same time.
  - Predictive validity is assessed when the test is administered before the criterion measure is obtained.

Content validity

- Content validity assesses the extent to which a test or measure covers a representative sample (i.e., includes all the relevant and important aspects) of the construct (i.e., domain) it is was designed to measure.
- Using the dimensions or subcomponents (i.e., required knowledge, skills, abilities, and other performance factors) found from the job analysis, test items (questions) that cover each of these KSAOs.
  - Subject matter Experts review the test items to ensure that they are relevant to, and representative of, the construct being measured.
- If a test or measure does not have good content validity, it does not accurately measure the construct it is intended to measure, which leads to inaccurate or unfair decisions.
  - This is true even if a test demonstrates criterion validity.

Construct validity

- Construct validity evaluates the degree to which a test actually measures the construct or characteristic it is intended to measure. [Note: This is the standard definition of what validity IS. –Refer back to Slide 8]
- Construct validity requires the collection of evidence from multiple sources that confirm that the test measures measuring the intended construct.
  - Convergent validity is a measure of how well the test correlates with other tests of the same construct.
  - Discriminant validity is a measure that confirms that the test does not correlate with tests of unrelated constructs.
  - Known-groups validity is the degree to which the test can distinguish between groups known to differ on the construct being measured.
- If there is a lack of construct validity, then applicants are being unfairly evaluated against an outcome that would not be expected to occur for the job in question.
  - Content validity and criterion validity are improperly used as a stand-in/proxy for construct validity.

Frequent Issues Regarding Validity

- Aggregation vs. Disaggregation
- Physical requirements
- Transported validity
- Examining validity reports

Aggregation vs. disaggregation

- Aggregation is combining test data from multiple sources (e.g., employees from multiple locations).
- The benefits are having enough statistical power to detect an effect because of a larger number of employees being including in the analyses.
  - This is why employers tend not to like aggregation.
- It is not always appropriate to aggregate (e.g., work is really different or the applicant pools are very different).
  - In this case, aggregating could mask differences.
    - This is why employers might present aggregated data.

Physical requirements

- For jobs that require strength, flexibility, and/or stamina, employers may elect to give physical agilities tests (e.g., police and fire).
- However, aspects of the test must still be job-related.
  - Tests should examine fitness for duty and not likelihood of workplace injury.
- These tests can be unfair to females, older workers, and persons with disabilities.

Transported validity

- Transportability is permitted by UGESP (Section 7B)
  - Criterion-related validity evidence gathered in one setting to be borrowed or "transported" to another setting.
  - Must meet the following requirements:
  - The jobs are similar (e.g., same KSAOs, similar work conditions).
  - The original sample must show "clear" validity evidence.
  - The test is "fair" (i.e., no evidence of adverse impact against any protected group).

Examining validation reports

- Things to look for when reviewing reports:
  - Job analysis report
  - Basic description of the test
  - Validation data
    - Types of validation data provided
    - Actual statistics (range, means, reliability, validity coefficients)
    - Clearly stated sources of the data (who took the test and when)
    - Indications of adverse impact (or lack of adverse impact)
    - How the statistics were calculated

Interesting trends in validity

- Lee et al. (2021) advocated for the use of expectancy tables to help set score cut-offs.
- Schäpers et al. (2020) found that a lack of situation descriptions had no impact on situational judgement tests' predictive ability of in-role job performance.
- Rottman et al. (2023) advocated for doing predictor-level analyses to reduce/eliminate adverse impact in trained models, while preserving predictive validity.
- Oh et al. (2023) argued that recent suggestions of not correcting for range restriction (restrictions of scores) in concurrent validation studies is problematic and can lead to underestimation.
  - Oh et al. did not consider that, when there is ROR of the predictor due to adverse impact, that the validity coefficients are likely to be overestimation and not underestimated.
- Götz et al. (2023) argued that test items can be generated via AI (NPL –The Psychometric Item Generator -PIG).
  - Götz et al. say that reliability and validity should be "carefully evaluated" when developing a test with PIG.

Thank you

References

- Conte, J. M. & Landy, F. J., (2018). Work in the 21st century: An introduction to industrial and organizational psychology (6th ed.). John Wiley & Sons.
- Equal Employment Opportunity Commission, Department of Labor, Department of Justice, and the Office of Personnel Management. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43(247), 38290-38315
- Götz, F. M., Maertens, R., Loomba, S., & van der Linden, S. (2023). Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development. Psychological Methods. https://doi.org/10.1037/met0000540.supp (Supplemental)
- International Organization for Standardization. (2020). ISO 10667-2:2020 Assessment service delivery – Procedures and methods to assess people in work and organizational settings – Part 2: Requirements for service providers. https://www.iso.org/standard/74716.html
- Lee, S. D., Walmsley, P. T., Sackett, P. R., & Kuncel, N. (2021). A method for displaying incremental validity with expectancy charts. Applied Measurement in Education, 34(4), 251–261. https://doi.org/10.1080/08957347.2021.1987902
- Oh, I.-S., Le, H., & Roth, P. L. (2023). Revisiting Sackett et al's (2022) rationale behind their recommendation against correcting for range restriction in concurrent validation studies. Journal of Applied Psychology, 108(8), 1300–1310. https://doi.org/10.1037/apl0001078.supp (Supplemental)
- Rottman, C., Gardner, C., Liff, J., Mondragon, N., & Zuloaga, L. (2023). New strategies for addressing the diversity–validity dilemma with big data. Journal of Applied Psychology, 108(9), 1425–1444. https://doi.org/10.1037/apl0001084.supp (Supplemental)
- Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (2020). The role of Situations in Situational Judgment Tests: Effects on construct saturation, predictive validity, and applicant perceptions. Journal of Applied Psychology, 105(8), 800–818. https://doi.org/10.1037/apl0000457
- Schmidt, F. L., & Hunter, J. E. (1980). The future of criterion-related validity. Personnel Psychology, 33(1), 41–60. https://doi.org/10.1111/j.1744-6570.1980.tb02163.x

Citation for this presentation

- El Kharzazi, R.J. (2023, September 21). Validity and Validation [Conference presentation]. Annual OFCCP Enforcement Conference, District of Columbia.