ORL 5524

A Validation Plan tied to Assessment Context Specifications

Directions: Think about the instrument you wish to design and your construct measures.

Follow the Process Model to specify the assessment context (Phase I-Construct(s), population and units, the measure-based inferences and specific uses you intend). Note this was your Assignment 1.

Create a draft of a validation plan using the “Unitarian” framework building on Assignment 1. Enter 2-5 types of validity and reliability evidence you would need, and say why.

See the example below from my forthcoming book ( Chatterji, in press, Guilford) Read the box and the table side by side.

**Box 10.2.**

The Unitarian View of Validity Applied to a Scenario

|  |
| --- |
| **Key Principle of the Unitarian View:** “All of validity is construct validity”  **Assessment Scenario:** A global technology corporation adopted, and now employs a standardized, mental ability test to select only the top 16% of candidates from the distributions of job applicants every year. To inform their decision-making processes, corporate leaders have chosen to interpret and use the normalized Standard Scores produced by the test and its sub-tests. As the cut-point for identifying the most able candidates, the decision-makers have set a z score= +1.0 using the distributions of three selected sub-tests’ and the overall test’s scores. Once recruited, candidates are placed in specialized jobs around the world.  **What is the Context of Assessment Use?**  ***Construct(s)?*** Mental ability levels in verbal, mathematical and technology domains, as relevant to specific jobs  ***Population?*** College-educated adults, ages 22+, English-speaking and demographically representative of the population in the multi-national organization  ***Population Units****?* Individual applicants *(units for making measure-based inferences)*  **What are the Assessment Purposes? *Intended uses of results:***  To recruit the most able applicants in verbal, mathematical and technology domains relevant for jobs, with the highest potential for succeeding over time  ***Intended interpretations of Standard Scores***:   1. To infer applicant ability levels accurately and reliably in three job-relevant sub-domains, using Standard Scores in sub-tests at the entry point. 2. To make accurate, comparative inferences of applicants’ overall ability levels at the entry point using Standard Scores on the total test. 3. To make valid and reliable categorizations of able versus less able applicants at the entry points using the cut-score point set. 4. To identify applicants with the highest potential for future success at the firm (at least a 2 year period).   ***Primary users:*** Human Resource (HR) and organizational leaders at the firm  **Validation Questions:**   * To be able to support the desired measure-based inferences and uses from applicants’ test scores, **w**hat kinds of evidence would we need to compile and evaluate as a whole? * What evidence is *essential*, or minimally needed to justify the proposed inferences and assessment uses? * What evidence is *most relevant* to the context of assessment use?   **See Table 10.1.** |

*Note.* Readers should review Box 10.2 and Table 10.1 together

**Table 10.1.**

Rationalizing and Tying Validity Evidence to Assessment User Needs: A Unified Validation Plan

|  |  |  |
| --- | --- | --- |
| **Type of**  **Validity Evidence** | **What is it?** | **Why is it necessary? Is it essential and relevant to the user context?** |
| 1. **Content-based validity of scores** | Evidence showing  *Content Relevance* and  *Content Representativeness*  of the items and the overall assessment operations | Assures that measure-based inferences reflect the theoretical definition of the constructs in terms of content and operations.  **Essential**? This evidence is essential to support the user-desired interpretations #1- #2 in Box 10.2  **Relevance?** This evidence will show the degree to which the test will yield scores relevant to, and representative of, the key indicators of job performance in valued domains. |
| 1. **Validity of norms and norm-referenced scores** | Evidence showing that the norms are recent, representative of the population, and relevant to the construct and user context | Assures validity of comparative inferences and decisions made with norm-referenced scores.  **Essential?** This evidence would support user-desired interpretations #1-2 in Box 10.2.  **Relevant?** As Standard Scores are norm-referenced scores, this evidence is highly relevant for the proposed inferences and assessment uses. |
| 1. **Predictive validity of scores** | Evidence showing that assessment results will predict performance of examinees sufficiently on valued criterion measures in future | Assures validity of aptitude-based, long-term inferences on the examinees’ potential to succeed in valued domains.  **Essential?** This evidence is essential for justifying the user-desired interpretation #4 in Box 10.2.  **Relevant?** This evidence is relevant to support examinee selection and recruitment decisions that expect long-term performance outcomes. |
| 1. **Score reliability** | Evidence showing that the total scores and sub-test scores of individuals yield consistent and replicable information. | Assures that individual measures/scores can be replicated within tolerable margins of error, even when conditions change.  **Essential?** It is essential for interpretations #1- 4 in Box 10.2.  **Relevant?** Reliability evidence is necessary for *all* test scores or derived measures. |
| 1. **Validity of categorical classifications and decision consistency levels around cut-score points** | Evidence showing that the cut-point set by users places candidates accurately in high versus low ability groups, and that the categorical placements are reliable. | Assures decision accuracy levels at the cutoff score point so that the most able recruits are selected, and the selections are reliable.  **Essential?** This evidence is essential to support desired interpretation #3.  **Relevant?** This evidence is relevant whenever users plan to make criterion-referenced, categorical selections of examinees based on the test scores |

*Note*. Readers should review Box 10.2 and Table 10.1 together