

Validity and reliability: reading the evidence

Consistency, interpretation, and use: concepts for evaluating a technical claim in context.

Reliability concerns the consistency and precision of scores under specified conditions. Validity concerns the evidence and theory supporting their interpretation for a particular use. Both require knowing the instrument, population, and study method.

Reliability: consistency of measurement

Measurement is subject to variation. Studying reliability helps examine the consistency of scores under defined conditions. Some methods examine relationships among items; others investigate stability across administrations.

These methods answer different questions. Internal consistency alone does not describe stability over time, predictive performance, or suitability for every population.

Validity: support for an interpretation

The AERA, APA, and NCME standards place validity in score interpretations for proposed uses. It therefore matters which conclusion is being proposed and what evidence supports it.

An analysis may consider content, response processes, internal structure, and relationships with other variables. The relevance of each source of evidence depends on the question and use; one number cannot summarize all of this work.

Why a coefficient is not a percentage of correct predictions

A consistency coefficient describes a property of scores in a study. Converting it into a percentage of correctly classified people changes its meaning.

Anticipating an external outcome requires a different design: a defined criterion, population, observation period, and performance analysis. Internal consistency and prediction should be communicated separately.

Context that should accompany a number

- Instrument name and version.
- Dimension or scale studied.
- Population, sample size, and selection criteria.
- Method, conditions, and analysis period.
- Result, uncertainty, and limitations.
- Intended use of the interpretation.

When a number belongs to a specific scale, retain that scope. It does not automatically apply to all dimensions or later versions.

Comparisons that require care

Coefficients from different populations can vary with sample composition, administration conditions, and study design. Before claiming that one version is better, check whether the comparison addresses an equivalent question.

Similarly, an association between variables does not establish causation. Method and limitations are part of the result, even when the number looks straightforward.

How to request useful evidence

Ask a concrete question about the use you need to support and its context. Request the complete report and check that conclusions match the design described.

- Does the evidence match the version I will use?
- Is the study population relevant to my context?
- Which interpretation does it support, and which are outside its scope?
- What further information does the decision team need?

What to keep in mind

An assessment can produce consistent scores and still be unsuitable for a particular decision. Check whether its supporting studies address how you plan to use it.

Frequently asked questions

Does high reliability establish predictive performance?

Not by itself. Prediction requires evidence involving an external criterion and an appropriate design.

Is validity a permanent label attached to an instrument?

It should be examined in relation to an interpretation, a use, and the evidence available for that context.

Sources for this reading

AMITAI Honestidad foundations manual, pages 27-31.

AERA, APA y NCME · Estándares para Pruebas Educativas y Psicológicas
<https://www.testingstandards.net/open-access-files.html>

Reidl-Martínez · Confiabilidad en la medición
<https://riem.facmed.unam.mx/index.php/riem/article/view/440>

An AMITAI educational guide. It reproduces no individual reports and does not replace documentation for the version used.