Using Tests to Study People's Responses:
What Do the Scores Mean?

Ariadna Angulo-Brunet
Open University of Catalonian, Spain

Oscar Lecuona
Rey Juan Carlos University, Spain

EXECUTIVE SUMMARY

In applied research in communication and journalism, as well as in other related sciences, it is common to use tests to assess unobservable constructs. The scores of these tests frequently need to be given meaning and interpreted, and their psychometric properties need to be reported as part of the study process or because the peer review procedure requires it. This chapter reviews the validity evidence required to give meaning to the scores of a study. It also provides practical examples from the literature, material for understanding the techniques to be applied, and an overview of best practices when using tests.

INTRODUCTION

In applied research, it is common to enquire about the reception of attitudes to or narratives around media content and formats. This usually requires the use of questionnaires, which when designed to elicit opinions on a subject based on a series of variables, are generally described as surveys. However, there are other common types of questionnaires aimed at measuring unobserved (or latent) qualities or characteristics. When these questionnaires involve standardized procedures, they are defined as tests (depending on the context we can refer to them as scales, subscales, etc.). This chapter focuses on giving meaning to the scores obtained in these questionnaires, that is, on the validity of test scores.

According to the Standards for Educational and Psychological Testing (SEPT; American Educational Research Association [AERA], et al., 2014, p. 11), "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests." In a nutshell, this entails pro-

viding evidence to support the use of certain test scores for a given intended purpose, such as describing responses or experiences or assessing the properties of particular groups of people or media contents. These interpretations of test scores are usually deemed "constructs" and are used to understand the assumed latent phenomena in tests. For a correct interpretation of test scores, it is important to detail the evidence to be able to make an interpretation according to the evidence presented. Depending on the context in which the research is developed, it will be necessary to present one type of evidence or another. In any case, following the recommendations given in this chapter will lead to a thorough understanding of the validity of the scores, and will therefore lead to a correct interpretation of them.

This chapter focuses on the kind of research evidence that can be used to develop a psychometric test, or to use a test that was developed in previous research. This chapter is intended as guidelines for those who want to use tests in their scientific work and need to account for the validity of their scores. Examples from the published literature are provided, and complementary materials are recommended in order for readers to implement the procedures described herein in their own work.

After presenting the background, the chapter is divided into two main blocks. The first focuses on the three qualitative sources of validity evidence (1) evidence based on test content, (2) evidence based on response processes, and (3) evidence based on the consequences of testing. The second section focuses on quantitative evidence: (4) evidence based on internal structure and (5) evidence based on relationships with other variables.

BACKGROUND

Despite the naïve assumption that the inclusion of a test is a routine action, the use of tests for research is much more complex than one might imagine. Researchers often believe that as long as a test has been validated and published in the past (either in a test editorial or in previous research articles) there is no reason not to trust it. However, this rationale exposes researchers and practitioners to lack of psychometric guarantees in their selected tests. This potential thoughtlessness towards measurement is broader that just test selection. The recently named "credibility revolution" (Vazire, 2018; also called "replication crisis," Nelson et al., 2018) signals that the scientific community is beginning to question the credibility of their findings, in which perhaps one of the biggest problems is related to the measurement of concepts (Flake & Fried, 2020; Scheel et al., 2021).

Or, put another way, are we sure we measure what we want to measure? Within this revolution, questionable or sub-optimal practices are used to measure unobservable concepts, which encompass from uninformed or not exhaustive review of psychometric literature to thoughtless application of statistical techniques (e.g., Viladrich et al., 2017; Navarro et al., 2018, chapter 8; an applied example in Lecuona et al., 2020). Thus, those who are doing the measuring need to ask what they can do to improve the process. Every decision that is made both when developing a test and when adapting it to certain research can jeopardize the interpretation of the scores. This is not a new issue. Forty years ago, McCroskey and Young (1979) were already warning about the misuse of certain techniques in communication studies (especially in the context of SEM [Structural Equation Modeling]).

Validating an instrument consists of providing evidence to support the use of the scores in the way they are intended to be used. This process requires, on the one hand, a qualitative work on which the whole reaction and revision of the test will be based. On the other hand, it also requires quantitative evidence to help prove or disprove the quality of the test. This validation process, to a greater or lesser

Related Content

Classification and Regression Trees
Johannes Gehrke (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 192-195).*
www.irma-international.org/chapter/classification-regression-trees/10819

Statistical Data Editing
Claudio Conversanoand Roberta Siciliano (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1835-1840).*
www.irma-international.org/chapter/statistical-data-editing/11068

Utilizing Fuzzy Decision Trees in Decision Making
Malcolm J. Beynonm (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2024-2030).*
www.irma-international.org/chapter/utilizing-fuzzy-decision-trees-decision/11097

Extending a Conceptual Multidimensional Model for Representing Spatial Data
Elzbieta Malinowskiand Esteban Zimányi (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 849-856).*
www.irma-international.org/chapter/extending-conceptual-multidimensional-model-representing/10919

Modeling Score Distributions
Anca Doloc-Mihu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1330-1336).*
www.irma-international.org/chapter/modeling-score-distributions/10994