Validating psychological tests: Trends and international differences. A bibliometric study

Jennifer Schroth, Saskia Naescher & Gabriel Schui

Background: Professional psychological testing and assessment is only possible when existing scientific standards and guidelines on test construction and test use are considered. Well-established examples are the German DIN norm 33430 (Kersting, 2008), the Standards for Educational and Psychological Testing (AERA, APA, & CME, 1999), and the International Guidelines for Test Use (ITC, 2001). In addition to several important aspects, the continuous examination of test validity represents one of the essential issues in all these standards. The study of Burholt et al. (2007) shows the significance of country-specific validity studies for social resource scales due to differences across countries.

Many validation studies are documented in the database records of the psychological databases from different countries, e.g. PsycINFO (US) and databases contained in the information retrieval system PubPsych: PSYNDEX (Germany), Pascal (France), ISOC Psicologia (Spain), ERIC (US), and Medline (US).

Research Questions:
1. Are there country-specific differences in the amount of documented validation studies over the years?
2. What future development(s) can be predicted for the total and country-specific amount of validation studies?

Methods

Bibliometric Method: Analysis of database records focusing on the validation of psychological tests. Included were PsycINFO and suitable databases contained by the information retrieval system PubPsych.

Inclusion criteria:
- databases with adequate amounts of relevant publications: Eric, ISOC-Psicologia, Pascal, PsycINFO and PSYNDEX
- journal articles focusing on the validation of psychological tests
- publication year between 1983 and 2013

Time Series Analysis: Analysis of the empirical time series of each database
1. Do the databases differ in respect to their documented validation studies over the years?
2. Which (ARIMA-) model fits best to describe the time series in each database?
3. What future development(s) can be predicted?

Results

Descriptive characteristics:
- high autocorrelations (eACF) and non-stationary processes
- no seasonal components
- trends can only be assumed for PsycINFO & PSYNDEX
- additive error terms for the series of Eric & ISOC-Psicologia

ARIMA modeling:
- the ARIMA[0,2,1] model can best describe the time series of Eric, ISOC Psicologia, PsycINFO & PSYNDEX
  1. i.e. the model contains only normally distributed and non-significant residuals as well as only significant parameters; the information criteria (AIC, BIC) are smallest
  2. however, the series of Eric is difficult to model!

Forecasting models:
- different forecasting techniques reveal similar results for the series of Eric, Pascal, PsycINFO & PSYNDEX but different results for the series of ISOC-Psicologia
- forecasts based on ARIMA-models for Pascal, PsycINFO & PSYNDEX but based on exponential smoothing for Eric (remember: there is no appropriate ARIMA-model for the Eric series)
- no clear picture for the series of ISOC-Psicologia

Discussion

- The increasing amount of validation studies in almost all databases leads to hypothesize that the perceived importance of country-specific validation also increases.
- There are distinctions between the databases from different countries in respect to the longitudinal development, depending on the thematic emphases of the particular databases.
- Along with the overall decrease of psychological tests in the German-speaking countries during the last years (Naescher, Schroth, Krampen & Schui, 2013), the focus of current research documented in PSYNDEX appears to be less on the construction of new instruments but rather on the further validation of existing ones. The model-similarities of PsycINFO and PSYNDEX (but with a much higher increase in PsycINFO) could also hint in this direction or just show similarities with regard to content and structure.
- Time series modeling suggests a continued development in this direction in total and for each database (except for the databases Eric and ISOC Psicologia).

Literature


Presented on the 9th Conference of the International Test Commission, July 2014, San Sebastian (Spain)