

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

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 Psicología (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:

- a. databases with adequate amounts of relevant publications: Eric, ISOC-Psicología, Pascal, PsycINFO and PSYNDEX
- b. journal articles focusing on the validation of psychological tests
- c. 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?

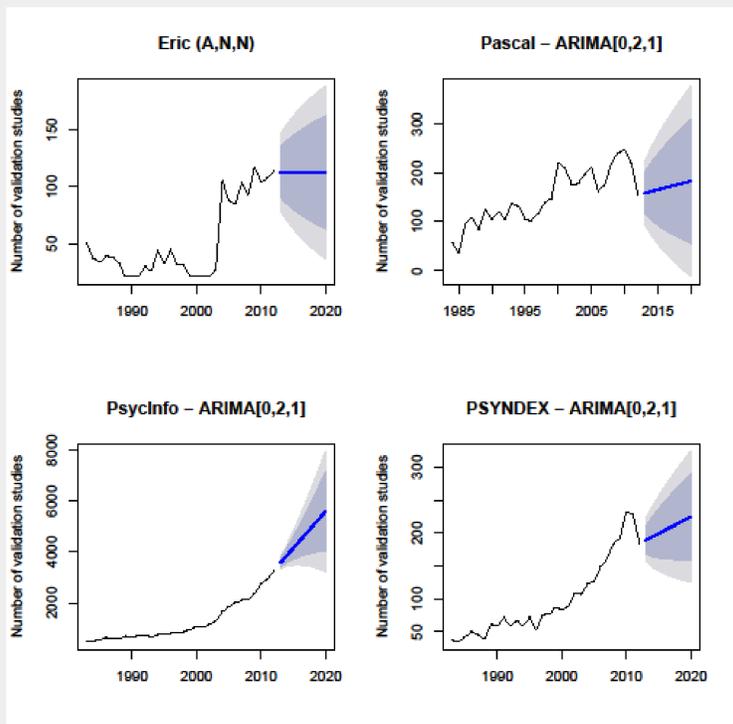
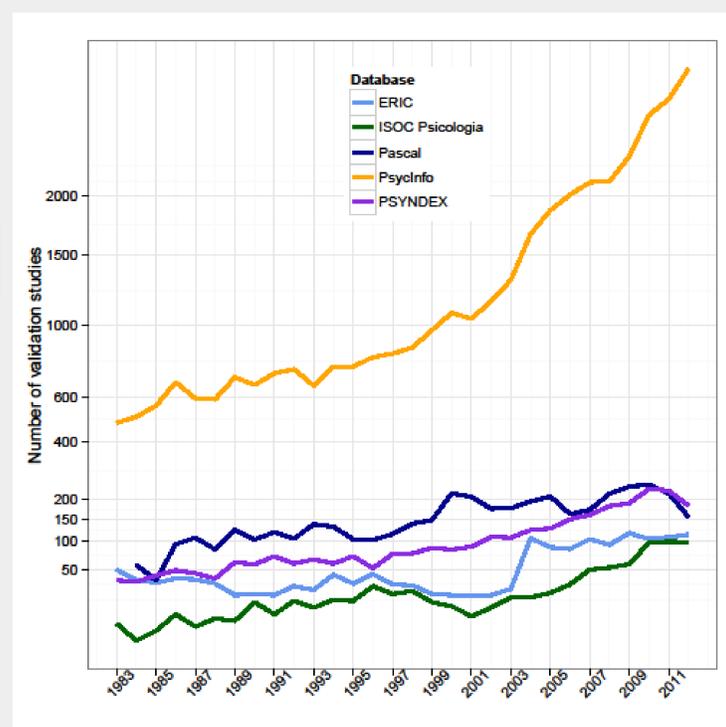
Results

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-Psicología

ARIMA modeling:

- the ARIMA[0,2,1] model can best describe the time series of Eric, ISOC Psicología, PsycINFO & PSYNDEX
- ➔ 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
- 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-Psicología
- 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-Psicología

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 Psicología).

Literature

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for Educational and Psychological Testing. American Educational Research Association.

Burholt, V. et al. (2007). Reliability and Validity of the Older Americans Resources and Services (OARS) Social Resources Scale in Six European Countries. Commentary. The Journals of gerontology. Series B, Psychological sciences and social sciences, 62 (6).

International Test Commission (2001). International Guidelines for Test Use, International Journal of Testing,

1(2), 93-114.

Kersting, M. (2008). DIN Screen, version 2. Guideline to control and optimize the quality of psychological tests and their use in vocational aptitude testing. In M. Kersting, Qualitätsicherung in der Diagnostik und Personalauswahl - der DIN Ansatz (S. 141-210). Goettingen: Hogrefe.

Naescher, S., Schroth, J., Krampen, G. & Schui, G. (2013). Testtrends: Entwicklung klinisch-psychologischer Testverfahren aus dem deutschsprachigen Bereich [Testtrends: Development of clinical-psychological tests from the German-speaking countries]. Poster presented on the DGPPN-Congress, Berlin.