I Introduction to Construct Measurement

Scales concisely show how a theoretical construct has been empirically measured. Their reliability and validity ratios give a notion of the quality of the scale. Since most scales are developed by the respective authors and are therefore in a first version, they can be used as a basis for supplementation and further development.

A multilevel procedure was applied to develop and evaluate these scales, as described below. The calculation of the reliability and validity ratios and their necessary ranges for high-quality designs is also presented.

The chosen approach is mainly based on Homburg's[1] guidelines and Homburg/Giering's[2] concept for conceptualizing and operationalizing complex constructs.

Many empirical studies examine relations between complex constructs and latent variables that cannot be measured directly.[3] As these constructs cannot be measured directly they have to be measured indirectly using indicator variables. Indicators are observed variables that are operationalizations of the latent variables and are therefore formally associated with the constructs. In order to perform a causal analysis with data from an empirical survey, the latent variables are first conceptualized and then operationalized. Conceptualization is the formulation of the relevant construct dimensions. In the subsequent process of operationalization, the measurement instrument in question is developed with its indicator variables.[4]

After conceptualization and operationalization, it is necessary to assess on the basis of empirical data whether or not the developed measurement instruments or scales fulfill the psychometric requirements of reliability and validity.[5]

Reliability refers to consistency of scores on a particular measurement instrument or scale. Methods that assess the reliability of a measurement instrument are in most cases based on classical test theory.[6] According to classical test theory each item *i* can be decomposed into a linear combination of a true score (common to all items) and a random error:

$$X_i = T + E_i$$

The random error is assumed to be uncorrelated with the true score. Additionally, researchers generally assume that the measurement error is not correlated with measurement errors of the other items. The total score is the sum of all *k* indicators or items that constitute the measurement instrument or scale:

$$X = X_1 + X_2 + \ldots + X_k$$

---

[1] Homburg (2000).
[2] Homburg/Giering (1996).
[3] Bagozzi/Phillips (1982), p. 465; Long (1983), p. 11.
[4] Churchill Jr. (1979), p. 66; Bagozzi/Baumgartner (1994), p. 388.
[5] Carmines/Zeller (1979), pp. 11-13.
[6] Lord/Novick (1968).

The reliability of the scale is the ratio of the true score variance to the total variance:

$$Rel(X) = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

The higher the reliability of an instrument, the closer the true scores will be to the total scores for that instrument. In case of high reliability only a minor proportion of shared variance of different measurements (indicators, items) can be attributed to an incorrect measurement process.[7] The association with the underlying constructs can instead explain an important amount of shared variance of the indicators.

Conceptually, reliability is consistency. Reliability is therefore frequently assessed in terms of internal consistency of a measure using Cronbach's alpha. Other methods for the assessment of an instrument's reliability are retest reliability and parallel test reliability.[8]

Construct validity is one of the most central concepts in psychometrics. Researchers typically establish construct validity by estimating correlations between a measure of a construct and other measures that should be associated with it (convergent validity) or vary independently of it (discriminant validity)[9]. A measurement instrument is valid if it is as free of random errors as possible and additionally conceptually correct.[10] This is why the reliability of a measurement is a necessary prerequisite for its validity.[11] The validity concept, however, goes beyond the reliability concept because it includes conceptual accuracy.[12] A measurement instrument is valid if it measures what it is supposed to measure. Various quantitative methods are used to verify that measurement models comply with the criteria of reliability and validity. Fornell (1982) differentiates between 1st generation and 2nd generation methods.[13]

1st generation criteria for reliability and validity derive from psychometrics and are based on classical test theory. Among the 1st generation criteria described in the construct part of this book are item-to-total correlations, Cronbach's alpha and exploratory factor analysis.[14]

2nd generation criteria for reliability and validity are based on modern test theory, i.e. structural equation modeling using confirmatory factor analyses (CFA). Another measurement theory that assesses reliability and validity in a different manner is item response theory.[15] In various respects, evaluation methods based on modern test theories are more powerful than 1st generation criteria.[16] The advantages of CFA models include estimates of reliability and validity as well as an overall test of the fit of the model to the data.

---

[7] Churchill Jr. (1991), p. 495.
[8] Moosbrugger (in press); Schermelleh-Engel/Werner (in press).
[9] Westen/Rosenthal (2003).
[10] Churchill Jr. (1979), p. 65.
[11] Peter (1979), pp. 6-10.
[12] Homburg/Baumgartner (1995).
[13] Fornell (1982), p. 217.
[14] Anderson/Gerbing (1988).
[15] Van der Linden/Hambleton (1997).
[16] Long (1983), pp. 11-19.