

# gesis

Leibniz Institute  
for the Social Sciences

## Checklist for the Documentation of Measurement Instruments in ZIS

*(Last update: September 2019)*



Open-access measurement instruments for the social sciences  
[www.gesis.org/en/zis](http://www.gesis.org/en/zis)

# 1 Information on the instrument

## Instruction and instrument

### *Description of*

- ...the instructions for the measurement instrument
- ...the wording of the items of the measurement instrument
- ...to which subscales which items belong to (only if there are subscales)

## Response specifications

### *Description of*

- ...the response categories
- ...the numerical codes of the response categories
- ...the labels of the response categories

## Scoring

### *Description of*

- ...how numerical values are assigned to different response categories and if and which items have to be recoded
- ...which items can be combined to form a scale-score/index
- ...whether there are subscales that have to be analysed separately
- ...whether simple or weighted total score(s)/index(es) are formed
- ...how to handle unanswered items if non-response is possible (item non-response)

## Application field

### *Description of*

- ...the purpose of the measurement instrument
- ...the survey mode (e.g., web-based, paper & pencil, or verbal interviewing, PAPI, CAPI, CATI, CASI) typically used for the measurement instrument and to which mode the samples in the study refer to

### *Description of*

- ...the target population for which the measurement instrument has been developed and in which population the measurement instrument has been psychometrically tested
- ...the duration of the processing time including the time for instruction when using a particular survey mode (mean/median and range or 25/75 percentile or the author's experience)

## **2 Information of the theoretical background**

### *Description of*

- ...why the measurement instrument is relevant to the field
- ...the theory the underlying construct is based on
- ...any relevant literature

## **3 Information about measurement instrument development**

- Specification of the statistical software and the version (e.g., Mplus, R packages, Stata, SPSS) with which the analyses were conducted

### **Item generation and selection**

#### *Description of*

- ...how the items were generated
- ...from which source they were taken over
- ...to which criteria items were selected (e.g., item difficulty and selectivity)

### **Sample(s)**

#### *Description of*

- ...the sample(s) for development and evaluation of the measurement instrument
- ...the recruitment of the sample(s) (simple random sampling, stratified sampling, cluster sampling, ad-hoc samples; participation with or without payment)
- ...the year of the sampling

### *Description of*

- ...the characteristic features of the sample(s) (e.g., gender, age, educational achievement, mother language, socioeconomic status, geographic region)
- ...whether the sample(s) are suitable or not for use as norming sample(s)

### **Item analyses**

#### *Description of*

- ...the dimensionality of the measurement instrument (if applicable)
- ...the evaluation methods you used to explore the dimensionality of the measurement instrument (if applicable)

- If you used exploratory factor analyses

#### *Description of*

- ...the method of extraction and rotation, the Eigenvalues, at least a matrix of the factor loadings, and communalities

- If you used structural equation models (e.g., confirmatory factor analysis, exploratory structural equation models)

#### *Description of*

- ...the dealing with missing data
- ...the estimator (e.g., ML(R), WLSMV, Bayes)
- the model specifications (e.g., which items belong to which factor)
- ...indicators for the global model fit (at least:  $\chi^2$ , CFI, RMSEA) and ideally for aspects of the local model fit (modification indices or residuals).

### **Item parameter**

#### *Description of*

- ...characteristic parameters that allow to rate the item quality  
[For example, sign and size of path coefficients (from construct to item) or means (of the items) of a structural equation model, the item discrimination parameters and threshold of an IRT (item response theory) model can be presented, or, alternatively, means, standard deviations, and selectivity of the manifest items.]

## 4 Information about quality criteria

### Objectivity

#### *Evaluation of*

- ...the application of the measurement instrument
- ...the evaluation of the measurement instrument
- ...the interpretation of the scale-score(s)/index(es)

### Reliability

#### *Description of*

- ...the estimates of reliability. Confidence intervals are an option and may be presented in addition to point estimates. Note that reliability indices pertain to the sample and not to the measurement instruments!
  - If models of classical test theory were used, Cronbach's alpha, split-half, or test-retest correlations are appropriate if their prerequisites are met (e.g., essential tau-equivalence for alpha, parallel test forms for split-half). Otherwise the scale reliability according to Raykov (1997) or one of the McDonald's omega should be presented.
  - If measurement models of the latent state-trait theory (Steyer, Schmitt, & Eid, 1999) were used, estimations of reliability, consistency, specificity of time of measurement (and, if applicable, specificity of methods) are presented.
  - If IRT models were used, Andrich's reliability or the scale reliability according to Raykov, Dimitrov, and Asparouhov (2010) can be presented.

### Validity

#### *Description of results that indicate*

- ...content validity (if applicable)
- ...factorial validity (if applicable)
- ...construct validity
- ...criterion validity of the measurement instrument (applicable)

### **Further quality criteria**

*Description of ratings or results to rate*

- ...test economy
- ...fakeability and response bias
- ...test fairness, inter alia, measurement invariance testing across groups (e.g., gender)
- ...measurement invariance testing over time or groups

### **Descriptive statistics**

*Description of*

- ...means, standard deviations, skewness, and kurtosis of scale score(s)/index(es)
- Norm tables (if intended and applicable)

## References

- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*, 173–184. doi:10.1177/01466216970212006
- Raykov, T. Dimitrov, D., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling, 17*, 265–279. doi:10.1080/10705511003659417
- Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13*(5, Spec Issue), 389–408. doi:10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.0.CO;2-A