Methodological approach for the development of equivalent rating scales for comparative cross-national surveys

anna andeenkova

CeSSI, Russia

natalja menold

GESIS, Germany

Comparability of rating scales

To draw reliable comparisons based on cross-national comparative data, survey data should satisfy requirements of comparability on measurement level.

The quality of measurements obtained may depend on the **verbal labels** attached to response options (e.g. Edwards, 1952; Krosnick and Fabrigar, 1997; Krosnick & Presser, 2010).

The lack of equivalence in rating scales and verbalization of rating scales can limit cross cultural comparability (Mohler et al., 1998), prevents from drawing reliable conclusions or can even mislead the analysis

We are looking for the approach to at least party solve this problem

Broadly used question design for measuring attitudes, values and opinions of general population = questions with rating scales with verbal labels as answer categories.

Use of Verbalized Rating scales in comparative surveys

Analysis of large cross-national surveys

	Total number of questions	Questions with verbal rating scale
WVS (2009-2010)	260	149
ISSP (2009)	73	45
ESS (core module)	148	42
Asiabarometer	167	105

gesis
Leibniz Institute
for the Social Sciences

Cessi

Verbalization of rating scales

Labels in rating numerical scales:

- Designed to define the **meaning** of the whole scale

  Agree-disagree, in favor-oppose, support-against, favorable-unfavorable, like-dislike, good-bad etc.

- Define the **interval** (spectrum)

  From Like very much to Dislike very much

  Or from Like extremely – Do not like

- Define the **distance** between points of scale

  Very much-somewhat-not very much-not at all

  A lot – mostly – half-half -slightly

- Define **type of the scale**

  unipolar-bipolar, balance-unbalanced, extremity of the scale

SHOULD BE EQUIVALENT ON THESE PARAMETERS IN DIFFERENT LANGUAGES

Cessi

Perception of verbalized rating scales in different cultures/ countries

MAIN ELEMENTS

- Language
- Culture
- Social context, institutions

IMPACT

- Translation issues
- Adaptation issues (cultural symbols, norms, history)
- Response formatting and presentation (e.g. acquiescence, positively bias, extremity of opinions or avoiding the definite answer = response styles)
- Communication during the interview
- Interpretation in analysis stage

gesis
Leibniz Institute
for the Social Sciences

Cessi

Differences in designing equivalent labels in different countries/ languages

Construction of agree-disagree scale in ESS (example of 6 countries)

GB	DE and CH	RU	UA	EE	IL
Agree strongly	Stimme stark zu	Полностью согласен	Полностью согласен(а)	Полностью согласен (на)	Абсолютно согласен
Agree	Stimme zu	Согласен	Согласен(а)	Согласен (на)	Согласен
Neither agree nor disagree	Weder noch	Где-то посередине	Насколько согласен(а), настолько и не согласен(а)	Ни согласен(на) ни против этого	В чем-то согласен, а в чем-то нет
Disagree	Lehne ab	Не согласен	Не согласен(а)	Не согласен(на)	Не согласен
Disagree strongly	Lehne stark ab	Совершенно не согласен	Совершенно не согласен(а)	Совершенно не согласен(на)	Абсолютно не согласен

Differences in designing equivalent labels in different countries/ languages

Analysis of Differences in Realization of the scale

✓ Polarity (some are unipolar, some are bipolar grammatically)

✓ Interpretation of middle category (neither-nor, partly, in-between (middle)

✓ Extremety (strength – strongly, absolutely or completeness (fully)

✓ Reference to gender (neutral, masculine only, masculine-feminine)

gesis
Leibniz Institute
for the Social Sciences

Cessi

Differences in designing equivalent labels in different countries/ languages

Conclusions from the analysis of configural, metric and scalar invariance using MGCFA test (Menold, Andreenkova, 2012)

- **Descriptive results**
  - Differences were found between bipolar and unipolar scales in choosing disagree category
  - Differences between countries in the choice of middle category in the case of "neither/nor" were low

- **Overall results**
  - Limited comparability across all countries
  - Higher comparability across German/English than within Russian language countries
  - Higher comparability if middle category in one form (neither/ nor is used)

gesis
Leibniz Institute
for the Social Sciences

Cessi

Approach to construct equivalent rating scales for cross-national surveys - Idea of new project

- **Task:** construct equivalent rating scales for comparative surveys of different countries and languages based on rigorous methodological experimental design

- **First step:** CRS-GR project (Comparable Rating Scale between German and Russian project)

- **Main method:** simultaneous construction of rating scales in different languages and parallel evaluation of the equivalence of these scales ("language-invariant" generation of scales, Osgood et al.1974).

Example: labels for extremes

ENGLISH	RUSSIAN	GERMAN
Entirely	Полностью	Voll
Absolutely	Совершенно	Voll und Ganz
Totally	Абсолютно	Ganz
Completely	Очень	Völlig
Wholly	Чрезвычайно	Vollkommen
Certainly	Крайне	Total
Strongly	Определенно	Vollends
Highly	Вполне	Absolut
Very much	Всецело	Durchaus
Really	Безусловно	Gar
		Sehr

gesis
Leibniz Institute
for the Social Sciences

Cessi

CRS-GR project design

- **First stage.** Explorative analysis

  of available verbalizations (lexica) and practice of their usage in surveys

  linguistic analysis (linguistic characteristics of each label including strength, emotional power, fixed or unfixed nature, etc).

- **Second stage.** Cognitive interviews to empirically explore the perceived meaning of each item of rating scale for selected central social science concepts, define subjective distance between items, major characteristics of each labeled item for respondents from different social groups, education, gender and age.

  Based on the analysis of cognitive interviews, few rating scales will be selected for quantitative test to obtain the subjective score (distance) for each label obtained and the comparison of the most equivalent labels between different languages.

gesis
Leibniz Institute
for the Social Sciences

Cessi

CRS-GR project design

- **Third stage.** Experimental test is used to evaluated several alternative verbal scales which have the potential for being highly comparable between languages to check inner consistency, concept equivalence and concept comparability.

EXPECTED RESULT: few rating scales for different topics will be constructed with empirically proved equivalence between countries and languages

other researchers to make informed choice of rating scales for further surveys.

Discussion

- Translation procedures without cross-cultural cognitive pretests and obtaining measurement equivalence/cross-cultural „measurement quality" are not enough to ensure the comparability

- Lack on empirical research whether the limited comparability is due to items, or rating scales, or both

- Lack on empirical research with respect to the possible standardization of cross culturally equivalent rating scales independent from the content of items/questions

- Extension of the project to other languages/cultures.