

gesis

Leibniz Institute  
for the Social Sciences



# Open-Ended Questions

## GESIS Survey Guidelines

*Cornelia Züll*

These slides are based on the GESIS Survey Guideline paper about open-ended questions:

Züll, C. (2016). Open-Ended Questions. GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg\_en\_002

Please cite the slides as:

Züll, C. (2017). Slide Set: Open-Ended Questions. GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences.

A complete list of all references used on these slides can be found in the above mentioned Survey Guideline paper.

# Open-ended Questions

- What are open-ended questions?
  - ▶ Survey questions **without any response options**.
  - ▶ Respondents formulate an answer in **their own words** and express it verbally or in writing.
  - ▶ Example:

Can you tell me what you associate with the term "left"?

⇒ Record the response given precisely!!

☒ : \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

# Area of Application

- Open-ended questions **to gather information** are used in the following cases:
  - ▶ Knowledge measurement
  - ▶ Unknown range of possible answers
  - ▶ Avoidance of excessively long lists of response options
  - ▶ Avoidance of directive questions
  - ▶ Cognitive pretesting
- Open-ended questions **to motivate** the respondents:
  - ▶ Give the respondents the opportunity to express their opinion freely and to express criticism or make comments.

# Who Answers Open-ended Questions?

Answering open-ended questions is a great burden for respondents: instead of selecting a category they have to formulate an answer in their own words.

Research showed, for example, that:

- All respondents **are able to answer** this type of question.
  - ▶ Important is the respondent's **level of interest** and
  - ▶ the **relevance** of the topic for them.
- Studies have shown that dissatisfied respondents are more likely than satisfied respondents to answer. They are more interested in expressing their dissatisfaction (for example, Borg & Zuell, 2012; Poncheri et al., 2008).

# Design of Open-ended Questions (1)

- Clarity and expected length of the answer
  - ▶ Respondents must understand **very clearly** what is expected of them.
  - ▶ Avoid addressing more than one topic per question (e.g., if the aim is to ask about positive and negative aspects it is better to ask two questions).
  - ▶ The length of the answer should be clear to the respondents (e.g., only keywords, one item of information, a small essay, etc.).
  - ▶ In Web surveys the size of answer boxes and motivation texts can help the respondents.

## Design of Open-ended Questions (2)

- Increasing respondents' motivation to answer
  - ▶ Open-ended questions should be used sparingly to avoid overburdening respondents.
  - ▶ The importance of the information for the researcher and his/her work should be emphasized.
- Motivation and survey mode
  - ▶ Face-to-face interviews: interviewers can help respondents and motivate them (Dillman & Christian, 2005).
  - ▶ Web and postal surveys: respondents can be motivated by providing motivation texts and designing the answers boxes appropriately (see, for example, Smyth et al., 2009; Emde & Fuchs, 2012; Züll at al., 2014).

# Coding of Open-ended Questions: Quantitative Content Analysis

One or more coders code the open responses on the basis of a predefined categorization scheme.

The coding process consists of the following steps:

1. Development of a categorization scheme describing all relevant coding categories. Categories can be derived from theory or directly from text data.
2. Coder training and pilot testing of the categories.
3. Coding of all answers.
4. Reliability: a sample of answers is coded by a second coder and a measure of coding reliability is calculated.
5. Data management: matching the coded data and other data of the survey.

# Coding of Open-ended Questions: Computer-assisted Content Analysis (1)

## Coding procedure

1. Coding rules for each category are formulated as lists of words/phrases.
2. Responses will be automatically coded on the basis of a content-analytic dictionary (corresponding to the “categorization scheme” in content analysis).
3. Quality of the coding in computer-assisted content analysis needs to be tested.
4. Data management: matching the coded data and other data of the survey.

# Coding of Open-ended Questions: Computer-assisted Content Analysis (2)

Advantages/disadvantages in comparison to content analysis

- Coding of large volumes of data
  - ▶ In one survey with many respondents or
  - ▶ in several surveys in which the same question was fielded.
- Quick and reliable coding.

**But**

- Definition and validation of suitable lists of words is resource-consuming.
- Applicable only for categories where words are unambiguous indicators for the category (for example: topics).

# Coding of Open-ended Questions: Computer-assisted Content Analysis (3)

- **TextQuest**

<http://www.textquest.de>

- **MaxDictio, an add-on module of Maxqda**

<http://www.MaxQDA.de>

- **WordStat**

<https://provalisresearch.com/products/content-analysis-software/>

# Other Approaches to Coding and Analysis (1)

## Co-occurrence analysis

- Besides the coding of the answers, there are a number of different procedures for directly analysing responses without assigning one or more codes to each individual response. These procedures focus on words that occur together within a response. Based on this co-occurrence a similarity or distance matrix is calculated and further analysed, for example by subjecting it to cluster, correspondence, or multidimensional scaling analysis.
- Programs for the analysis
  - ▶ TLab ([www.tlab.it/default.php](http://www.tlab.it/default.php))
  - ▶ Alceste (<http://www.image-zafar.com/Logicieluk.html>)

# Other Approaches to Coding and Analysis (2)

## Semi-automatic coding (based on supervised machine learning)

- On the basis of a manually coded text sample (“a training set of answers”) further texts can be automatically coded. The program “learns” from the manually coded texts how particular responses should be coded (see, for example, Giorgetti & Sebastiani, 2003; Schonlau & Couper, 2016).
- An alternative semi-automated procedure is the structural topic model (see Roberts et al., 2014), which is based on a machine learning approach to the analysis of textual data. Information about the respondents, such as their sex, age, or political affiliation, is incorporated into the analysis.

gesis

Leibniz Institute  
for the Social Sciences

Member of  
*Leibniz*  
Leibniz  
Association