# DASISH WP3.3 task 2: Keystroke Analysis for SHARE and ESS

Verena Halbherr, GESIS - Leibniz Institute for the Social Sciences, Mannheim Johanna Bristle, MEA – Max Planck Institute for Social Law and Social Policy, Munich

# Objective

Explore keystroke data and the potential of keystroke analyses and its implication for fieldwork.

# Keystroke Data

In the process of producing survey data much paradata, i.e. data about the process of survey production, are generated.

Keystroke data record actions on a keyboard and time stamps for each action taken. This allows measuring the time used to answer single questions, part of the survey (e.g. modules), or the length of the survey.

# Quality and Processing of Keystroke Data

### Availability of data depends on software system used

• PAPI and decentralised CAPI systems limit the available data for analysis.

#### Processing

- Non-rectangular structure (number of items varies over respondents)
- Technical errors produce outliers or missing data

### SHARE and ESS Data

### SHARE:

- centralized system (Blaise)
- Full potential of keystroke data available
- Data available during fieldwork

#### ESS:

- country specific CAPI system (computer assisted)
- PAPI (paper and pencil)





Potential of Keystroke Analysis to Improve Survey Quality

**Questionnaire development** 

- Understanding of questions
- Number of times an answer has been changed
- Additional time for new items
- Change of item/module length over waves
- Length of introduction items
- Length of interview (total)
- · Combining item quality with item length to decide on additional items

#### **Monitoring fieldwork**

with 25+ interviews in Germany

Interviewer performance

- Reading out times (of introduction texts)
- Speeding
- Skipping

Interviewer pace

- Interviewer fraud • Conspicuous short interviews
- **Fieldwork progress**
- Time and day of interview

# 95.3 98.1 97.8 98.7 96.9 98.0 99.1 96.3 100 83.6 99.8 97.6 98.8 97.6 98.8 98.2 100 98.9 96.1 Source: Loosveldt G and K Beullens (2013) How long will it take An analysis of interview length in the fifth round of the European Social Survey", Survey Research Methods 7 (2): 69-78 table 1)

### **Post- Survey Quality Assessment**

- Interviewer length by respondent characteristic (e.g. age, education, employment status...)
- Interview length by interviewer and neighbourhood characteristic
- Data Quality (Item nonreponse, straightlining, ...) and interview length
- Effects of interviewer workload on performance
- Larger number of respondents and interviewer allow closer checks of previous keystroke analysis

#### Lesson learned

- Keystroke analysis is a new field of analysis which offers lots of potential.
- Data preparation and analysis requires advanced skills by the data analysts and researchers.
- Information about the time used provides additional insight in survey data, which can be used in multiple phases of the survey lifecycle to inform survey managers.
- Keystroke analysis is a useful tool for the investigation of measurement error and data quality. ٠