Can Survey Items Relevant to Mode-Specific Measurement Error Be Coded Reliably?

A Case Study on Eleven Official Dutch Surveys

Frank Bais, Barry Schouten, Vera Toepoel
F.Bais@uu.nl

Universiteit Utrecht

Centraal Bureau voor de Statistiek

INTRODUCTION

In current survey practice, various modes are used to collect data to increase response rates and reduce costs. However, measurement error may differ between modes. The occurrence of this mode-specific measurement error is partly influenced by the characteristics of the items of the survey (Tourangeau, Rips & Rasinski, 2000). To investigate this relationship, we need to code survey items on their characteristics. A first step is to check to what extent the coding of items can be done *reliably* by multiple coders.

RESEARCH QUESTIONS

1) To what extent can survey items be coded *reliably*?
2) How can we *explain* low intercoder reliability?
3) How do we *cope with* low intercoder reliability?

METHOD

Eight experienced survey researchers coded a selection of item characteristics for all items (2470!) of the Dutch Labour Force Survey (LFS) administered by Statistics Netherlands and the ten core studies of the Longitudinal Internet studies for the Social Sciences (LISS) (Assets, Housing, Income, Personality, Health, Politics & Values, Religion & Ethnicity, Family & Household, Work & Schooling, and Social Integration & Leisure) administered by CentERdata. Characteristics particularly relevant to mode-specific measurement error were coded by three coders. All other characteristics were coded by two coders.

Figure 1. An example of the characteristic 'sensitive information': A respondent starting to feel the tendency to give a socially desirable answer due to an intimidating survey mode.

Table 1. The Item Characteristics and Intercoder Reliabilities.

Item Characteristic (Campanelli et al., 2011; Gallhofer et al., 2007; Saris & Gallhofer, 2007)	1) Intercoder Reliability
Time reference	0.85
Conditions	0.89
Memory	0.85
Hypothetical situation	0.98
Calculations	0.94
Ambiguity	0.96
Mismatch	0.98
Formulation	0.57
Clarification	0.71
** Content of the question	0.56
** Difficult language usage	0.61
** Emotional charge	0.75
** Presumption of filter question	0.62
** Sensitive information	0.53
** Centrality	0.59
** Response complexity	0.91
** *coded by three coders*	

2) EXPLAINING LOW INTERCODER RELIABILITY

- The difficulty of defining the item characteristics.
- The subjectivity of coding the item characteristics.

3) COPING WITH LOW INTERCODER RELIABILITY

1) Excluding survey items for which no consensus was reached.
2) Redefining item characteristics.
3) Computerizing the definitions of the item characteristics.
4) Using scales consisting of different degrees of applicability for the item characteristics using all three coders.

CONCLUSION

Item characteristics relevant to mode-specific measurement error cannot be coded reliably. To investigate the relation between mode-specific measurement error and characteristics of survey items, potential ways of coping with low intercoder reliability need to be explored.

Many thanks to: Utrecht University, Statistics Netherlands, and CentERdata, and Peter Lugtig, Judit Arends-Toth, Salima Douhou, Natalia Kieruj, Mattijn Morren, Corrie Vis, Rodinde Pauw, and Rachel Vis-Visschers.