gesis
Leibniz Institute for the Social Sciences

Data Quality of Proxy Reports

Inconsistent Educational Information in the German Microcensus Panel
Simon Börlin, July 16th, 2019

Leibniz Association

Overview

- What are proxy interviews?
  - ▶ Target person (P) does not participate directly in survey
  - ▶ Third person (R) provides information
- Advantages:
  - ▶ Increase of the response rate
  - ▶ Reduction of survey costs, field time & respondent burden
- Disadvantages:
  - ▶ Reduced data quality?

  (Moore 1988, Cobb 2018a, Cobb 2018b)
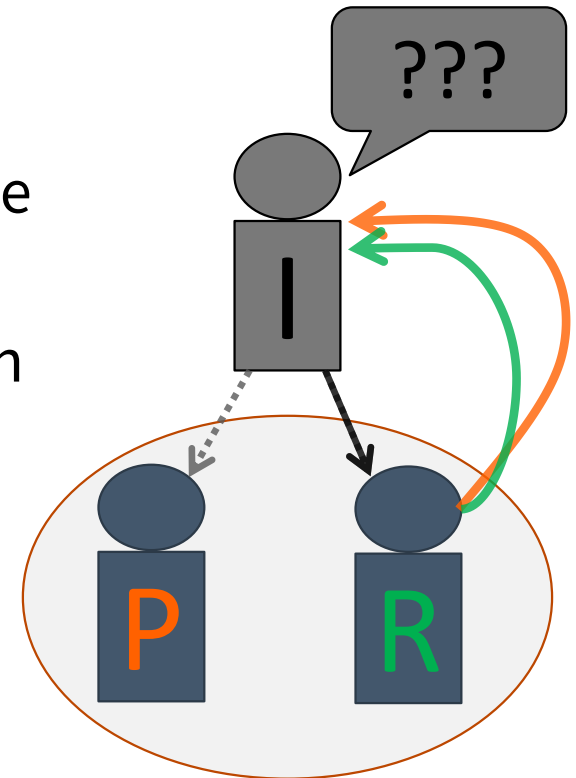
Problematic when analyzing data quality:

- Separation of *Selection Effect* & *Measurement Effect* (e.g. Moore 1988; Stark 2006)

- Possible solutions:

  1. Surveys with random proxy selection
     (e.g. Lee, Mathiowetz & Tourangeau 2004)

  2. Panel data: Investigation of (relatively) time-constant characteristics over time (test-retest method)(e.g. Zühlke 2008)

- Usually no external information available, assumption: self-report = "true" value (cf. Moore 1988)

Constellations regarding the type of reporting in t1 & t2

Constellation	t1	t2	Dimension of data quality
1	**Self**	**Self**	Reliability
2	Self	Proxy	Validity
3	Proxy	Proxy	Reliability
4	Proxy	Self	Validity

- Application of the test-retest method

- Verification of data quality using educational information:

  ▸ Educational attainment is a central variable in social science research

  ▸ *Education degree* & *year of graduation* are relatively constant from a certain age onwards

- Research question:

  To what extent does the respondent type (self or proxy report) affect inconsistent educational information?

Previous Research

Proxy reports ≠ self-reports on educational information

- Telephone follow-up survey
  - Parents > Spouses > Children (Dawe & Knight 1997)
- Children about parents (Kreuter et al. 2010)
- German microcensus (North Rhine-Westphalia): (small) differences (Zühlke 2008)

Meta-analysis on proxy reports:

Methodology & thematic range of research should be extended (Cobb 2018b)

4 Components of the Response Process (Tourangeau, Rips & Rasinski 2000; Lee, Mathiowetz & Tourangeau 2004):

1. ## Comprehension
   - Respondents answer question for themselves & proxy

2. ## Retrieval
   - Less / 2nd hand information
   - Motivation differences?
   - Relationship respondent & proxy decisive

3. ## Jugdment

4. ## Response
   - Differences in social desirability?

Hypotheses

- **Item-Nonresponse**:
  - ▸ H1a: Item-Nonresponse is higher for proxy reports
  - ▸ H1b: Relationship between respondent & proxy
    - ▪ Spouses less item-nonresponse than children or other persons

- **Inconsistent Educational Information**:
  - ▸ H2a: Reliability
    - ▪ Proxy-proxy entail more inconsistencies than self-self
  - ▸ H2b: Validity
    - ▪ Self-proxy (& proxy-self) more inconsistencies than self-self
  - ▸ H3: Relationship between respondent & proxy
    - ▪ Spouses less inconsistencies than children or other persons

Measuring Inconsistent Educational Information

2012		2013					
		0	1	3	4	5	9
0	Person w/o a general school leaving certificate (0) / finished after a max. of 7 years of school (6)	++	-	-	-	-	0
1	Lower secondary school-leaving certificate (*Hauptschulabschluss*)	--	++	-	-	-	0
3	Intermediate secondary school leaving certificate (*Realschulabschluss*)	--	--	++	-	-	0
4	Advanced technical college entrance qualification (*Fachhochschulreife*)	--	--	--	++	-	0
5	Upper secondary school certificate (*Abitur*)	--	--	--	--	++	0
9	No answer	0	0	0	0	0	00

Table: Comparison of Educational Degrees in the Years 2012 & 2013

Legend:

++	=	Same certificate in both years
-	=	Possible change of certificates
--	=	Lower / impossible combination of certificates
0/00	=	No answer (item-nonresponse)

Educ. Information used for the Analyses:
1) Highest achieved educational degree
2) Highest achieved vocational qualification
3) Year of educational degree
4) Year of vocational qualification
5) ISCED (Version 1997)

10

gesis Leibniz Institute for the Social Sciences

- German Microcensus Scientific Use Files of 2012 & 2013
  - ▶ Annual household sample survey w/ sampling fraction of 1% of the population in Germany
  - ▶ Collected by 14 statistical offices of the German states & prepared by the Federal Statistical Office in Germany (Statistisches Bundesamt - Destatis)
  - ▶ Obligation by law to provide information for majority of questions *(Mikrozensusgesetz)*

- Exclusion for analysis:
  - ▶ Missings
  - ▶ Only private households considered
  - ▶ Persons in general education schools
  - ▶ Age <20 years
  - ▶ Only persons with information about proxy respondents in both years

- N ≈ 120,000 persons

*Prop. of **Item-Nonresponse** of Educational Variables by Respondent Type*

Proxy in 2012						
	Self Report		Proxy Report		Total		Chi2
Item-Nonresponse in…	Obs	%	Obs	%	Obs	%	Signif
… Educational Degree	119	0.1	62	0.2	181	0.2	11.573
Total	90,577		27,879		118,456		***
… Voc. Qualification	205	0.3	73	0.3	278	0.3	2.100
Total	77,730		22,732		100,462		.
… Year of Educ. Degree	1,120	8.0	602	11.0	1,722	8.8	44.481
Total	14,086		5,495		19,581		***
… Year of Voc. Qualification	2,340	3.1	1,739	7.8	4,079	4.1	968.669
Total	76,502		22,412		98,914		***
… ISCED	238	0.3	107	0.4	345	0.3	9.796
Total	92,157		28,854		121,011		**
… at least One Educ. Variable	3,678	4.0	2,421	8.6	6,099	5.1	929.006
Total	90,823		27,997		118,820		***

Source: RDC of the Federal Statistical Office and Statistical Offices of the Laender, Mikrozensus **2012**; own calculations. Only persons with information about proxy respondent; only private households (>=2 persons) and persons older than 19 years.
Attention: Non-missing-category has been omitted for each variable.

Results of Logistic Regressions on Item-Nonresponse in at least One Educ. Variable for the Year 2012 - Average Marginal Effects

	Model 0 AME	Model 1 AME	Model 2 AME	Model 3 AME
0 Self Report	*0,000*	*0,000*		
1 Proxy Report	0,045***	0,060***		
0 No Proxy Report			*0,000*	*0,000*
1 Spouse			0,037***	0,040***
2 Daughter/Son (-in-law)			0,138***	0,069***
3 Mother/Father (-in-law)			0,014***	0,050***
4 Others			0,051***	0,079***
Pseudo R2	0,016	0,160	0,019	0,157
BIC	46881	40373	46804	40553
N	118530	118530	118530	118530

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; only private households (>=2 persons), only persons 20 years or older & not attending general school

Source: RDC of the Federal Statistical Office and Statistical Offices of the Laender, Mikrozensus **2012**; own calculations

Reference categories are *italic*. Reduced sample: only persons with information about proxy respondent in both years.

Models 1 & 3 controlling for: Sex, age, size of household, employment status, country of birth, citizenship, **mode**, region, German states.

Results of Logistic Regressions on Inconsistent ISCED - AME

	Model 0 AME	Model 1 AME	Model 2 AME	Model 3 AME
Self/Self	*0,000*	*0,000*		
Self/Proxy (Proxy/Self)	0,025***	0,015***		
Proxy/Proxy	0,018***	-0,004		
0 No Proxy Report			*0,000*	*0,000*
1 Spouse: Wife (pp)			-0,019***	-0,025***
2 Spouse: Husband (pp)			0,008	-0,003
3 Daughter/Son (-in-law) (pp)			0,052**	-0,016
4 Mother/Father (-in-law) (pp)			0,089***	0,030***
5 Others (pp)			0,029***	0,001
6 Spouse: Wife (sp/ps)			0,013**	0,011*
7 Spouse: Husband (sp/ps)			0,018***	0,016***
8 Daughter/Son (-in-law) (sp/ps)			0,133***	0,068***
9 Mother/Father (-in-law) (sp/ps)			0,071***	0,013
10 Others (sp/ps)			0,040***	0,015*
Pseudo R2	0,001	0,028	0,003	0,029
BIC	104272	101783	104129	101818
N	116699	116699	116699	116699

Models 1 & 3 controlling for:

Age, size of household, employment status, workload, **foreign certificate,** *citizenship, mode, region, German states*

Source: RDC of the Federal Statistical Office and Statistical Offices of the Laender, Mikrozensus 2012 & 2013; own calculations;

* p < 0.05, ** p < 0.01, *** p < 0.001; Only Private Households (>=2 Persons), only Persons 20 years or older & not attending school/voc. training; Reference Categories are *italic*.

Reduced Sample: Only Persons with Information about Proxy Respondent in both years.

14

Item-Nonresponse

- Proxy reports higher item-nonresponse on most educational variables
- Spouses less worse respondents than other persons

Inconsistencies in educational information

- Appear in all types of respondents
- Change of respondent type -> higher inconsistent rates for educational information
- Reliability: self reports slightly higher than proxy reports
- Validity: proxy reports lower data quality
- With regard to relationship of the proxy respondent:
  - spouses are as good as self reports & wives slightly better than husbands

Thank you for your attention!

For further information do not hesitate to contact me!

Data Quality of Proxy Reports: Inconsistent Educational Information in the German Microcensus Panel

Simon Börlin, July 16th, 2019

GESIS – Leibniz Institute for the Social Sciences

Simon.Boerlin@gesis.org