

Who do they think you are?

Inconsistencies in Self- and Proxy-Reports of Education within Families.

C. Lavest ¹ M. Ichou ² M. Ferry ² P. Präg ³

¹Utrecht University

²Institut National des Etudes Démographiques

³ENSAE-CREST

September 19, 2023, ECSR Conference 2023



Table of Contents

- 1 Introduction
- 2 Theoretical background
- 3 Data and Methods
- 4 Results
- 5 Discussion
- 6 Conclusion

Table of Contents

- 1 Introduction
- 2 Theoretical background
- 3 Data and Methods
- 4 Results
- 5 Discussion
- 6 Conclusion

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...

ESS

eiscdf - Father's highest level of education, ES - ISCED	
Type	Code
Description	European survey version of ISCED. Recoded from the ESS harmonised variable EDULVLFB. For further details, please see ESS10-2020 Documentation Report, Appendix A1 Education. Greece, Hungary, Montenegro, Norway: Change in underlying questionnaire item. For more information see appendix A1 under Documents.
Location	F52
Question	Generated variable: Father's highest level of education, ES - ISCED
	0 Not possible to harmonise into ES-ISCED
	1 ES-ISCED I , less than lower secondary
	2 ES-ISCED II, lower secondary
	3 ES-ISCED IIb, lower tier upper secondary
	4 ES-ISCED IIIa, upper tier upper secondary
	5 ES-ISCED IV, advanced vocational, sub-degree
	6 ES-ISCED V1, lower tertiary education, BA level
	7 ES-ISCED V2, higher tertiary education, >= MA level
	55 Other

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...

ESS

eliscdf - Father's highest level of education

Type	Code
Description	

FR_SPISC - Spouse, partner: degree

English Translation

from the ESS harmonised variable 9-2020 Documentation Report, underlying questionnaire documents.

ES - ISCED

ISS

Question no. and text	National Language	English Translation
Q1-Q58 : Quel niveau d'études le plus élevé votre conjoint ou partenaire a-t-il/elle atteint?	Q1-Q58 : What is the education level of your spouse or your partner ?	
Non scolarisé ou école primaire non achevée	None	
École primaire uniquement		
Certificat d'études primaires		
Scolarité suivie de la 6ème à la 3ème (sans brevet)		
→ ISCED V2, higher tertiary education, >= MA level		
55 Other		

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...

ESS

ESScedf - Father's highest level of education

Type: Code

Desc: *** For all except respondent's father unknown and no foster father (PERTOT# 4) ***

F 13 / T_DIPLOP What is ["was" if PERTOT=3] the highest qualification your father or foster father obtained?

Show CARD F_1

1. No qualification..... ☐ 1 → F 14

2. Certificate of primary education or equivalent diploma..... ☐ 2 → Filter 18

3. Certificate of lower secondary education, or equivalent diploma..... ☐ 3 → Filter 18

4. Lower-secondary-level vocational training certificate or equivalent..... ☐ 4 → F 15

5. Higher-secondary vocational or technical certificate or equivalent..... ☐ 5 → F 16

6. Higher-secondary general certificate, paralegal training, or equivalent..... ☐ 6 → Filter 18

7. Two-year higher education diploma..... ☐ 7 → F 17

8. Diploma requiring more than 2 years in higher education..... ☐ 8 → F 18

98. Refusal to answer (do not read)..... ☐ 98 → Filter 18

99. Don't know (do not read)..... ☐ 99 → Filter 18

10. U - 5 grades [RF] 2. 6 - 8 grades; "Grade School" 3. 9 - 11 grades; "Junior High"

GO TO KL6

4. 12 grades; "High School" 5. 12 grades plus non-academic training

6. 13 - 14 years; Some college; Associate's Degree

7. 15 - 16 years; College; Bachelor's Degree

8. 17 or more years; College; Graduate work; Ph.D, Professional degree

GO TO KL7CKPT

DK

GO TO KL5 SIGNAL

ISS

TeO

PSID

FT

Question no. and text	Codes/ Categories
Q8-Q58: C élevé vote il/elle at	No
Non se	Yes
acher	Yes
Eco	Yes
Certain	Yes
Social (sap	Yes

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...
- ... And used in many analyses

The screenshot shows a Google Scholar search interface. The search bar at the top contains the text "socioeconomic background" and a magnifying glass icon. Below the search bar, the word "Articles" is highlighted with a blue diamond icon. To the right of "Articles", a red rectangular box highlights the text "Environ 154 000 résultats (0,11 s)". On the left side of the results, there are several filters: "Date indifférente" (with sub-options "Depuis 2023", "Depuis 2022", "Depuis 2019", and "Période spécifique..."), "Trier par pertinence" (with sub-option "Trier par date"), and "Toutes les langues" (with sub-option "Recherche dans les pages en Français"). The main results area displays two entries. The first entry is titled "[HTML] Achievement at school and socioeconomic background—an educational perspective" by S. Thomson, published in npj Science of Learning, 2018, with a link to nature.com. The second entry is titled "A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement" by R. Berkowitz, H. Moore, and B. Astar, published in a 2017 journal review on sagepub.com. Both entries include a brief abstract and links to "Enregistrer", "Citer", "Cité" (with counts), "Autres articles", "Les 7 versions", and "Web of Science".

Proxy-reports in Sociology

- Proxy-reports of socioeconomic background are in many surveys...
- ... And used in many analyses
- But how valid are they? (Lien, Friestad, and Klepp, 2001)

We have little knowledge

- Proxy reports of SES data considered reliable
 - Salient feature
 - Fixed nature
 - Nonsensitive
- Reliability of reports is difficult to assess
 - Need multiple informants or multiple sources
 - What is a "true" report?
- Little motivation due to lack of alternatives
 - Gathering information is costly
 - Register data not available everywhere

We have little knowledge

- Proxy reports of SES data considered reliable
 - Salient feature
 - Fixed nature
 - Nonsensitive
- Reliability of reports is difficult to assess
 - Need multiple informants or multiple sources
 - What is a "true" report?
- Little motivation due to lack of alternatives
 - Gathering information is costly
 - Register data not available everywhere

We have little knowledge

- Proxy reports of SES data considered reliable
 - Salient feature
 - Fixed nature
 - Nonsensitive
- Reliability of reports is difficult to assess
 - Need multiple informants or multiple sources
 - What is a "true" report?
- Little motivation due to lack of alternatives
 - Gathering information is costly
 - Register data not available everywhere

Research Question

Who can report on who and what consequences do proxy-reports of education have on social mobility estimates?

Table of Contents

- 1 Introduction
- 2 Theoretical background**
- 3 Data and Methods
- 4 Results
- 5 Discussion
- 6 Conclusion

Literature Review

- Some interest in the topic in the 80's, culminating in a literature review Looker, 1989
- Caution raised about using proxy-reported education/SES information (Engzell and Jonsson, 2015; Vries, 2006; Reynolds and Wenger, 2012; Houseworth and Fisher, 2020)
- SES itself being linked to patterns of non-response/"errors" in reporting (Engzell and Jonsson, 2015; Lien, Friestad, and Klepp, 2001; Pu, Huang, and Chou, 2011; Vries, 2006; Vereecken and Vandegehuchte, 2003)

Literature Review

- Some interest in the topic in the 80's, culminating in a literature review Looker, 1989
- Caution raised about using proxy-reported education/SES information (Engzell and Jonsson, 2015; Vries, 2006; Reynolds and Wenger, 2012; Houseworth and Fisher, 2020)
- SES itself being linked to patterns of non-response/"errors" in reporting (Engzell and Jonsson, 2015; Lien, Friestad, and Klepp, 2001; Pu, Huang, and Chou, 2011; Vries, 2006; Vereecken and Vandegehuchte, 2003)

Literature Review

- Some interest in the topic in the 80's, culminating in a literature review Looker, 1989
- Caution raised about using proxy-reported education/SES information (Engzell and Jonsson, 2015; Vries, 2006; Reynolds and Wenger, 2012; Houseworth and Fisher, 2020)
- SES itself being linked to patterns of non-response/"errors" in reporting (Engzell and Jonsson, 2015; Lien, Friestad, and Klepp, 2001; Pu, Huang, and Chou, 2011; Vries, 2006; Vereecken and Vandegehuchte, 2003)

Theoretical background

- Children studies

- Children from high SES background answer more often and are more reliable
- Parental data difficult to obtain (Engzell and Jonsson, 2015; Lien, Friestad, and Klepp, 2001)
- Minority children and children in "non-standard" families are less reliable (Looker, 1989; Pueyo et al., 2007; West, Sweeting, and Speed, 2001)

Theoretical background

- Children studies
- Adult children as proxies
 - Nonresponse?
 - Memory decay
 - **Changes in educational structure** (Breen and Müller, 2020)
 - **Social desirability bias** (Edwards, 1957)
 - **Social Mobility** (Broom et al., 1978; Hedley, 1986; Irwin, 2015; Merllié and Prévot, 1997)

Theoretical background

- Children studies
- Adult children as proxies
 - Nonresponse?
 - Memory decay
 - **Changes in educational structure** (Breen and Müller, 2020)
 - **Social desirability bias** (Edwards, 1957)
 - **Social Mobility** (Broom et al., 1978; Hedley, 1986; Irwin, 2015; Merllié and Prévot, 1997)
- What about proxy reports from parents?
 - Shared experience
 - Recent information

Summary of Hypotheses I

Hypothesis 1 - Non-response

- H1a. Parents' non-response of their children's educational attainment is lower than children's non-response of their parents' educational attainment.
- H1b. Higher educated individuals display less non-response than lower educated ones, in both self and proxy reports

Hypothesis 2 - Knowledge

Agreement on child's educational attainment (i.e. when parents act as proxy-reporters) is greater than on parent's educational attainment (i.e. when children act as proxy-reporters)

Summary of Hypotheses II

Hypothesis 3 - Social Desirability

Proxy-reports tend to overestimate the alter's level of education

Hypothesis 4 - Social Mobility

Larger educational differences between respondents and proxy reporters lead to an underestimation of the educational distance between them.

Table of Contents

1 Introduction

2 Theoretical background

3 Data and Methods

4 Results

5 Discussion

6 Conclusion

Data and Variables I

- *Trajectoires et Origines 2* dataset (2019) [▶ Additional information](#)
- Subsample of 421 pairs (and 752 individuals)
- One parent and one (or more) child(ren) interviewed

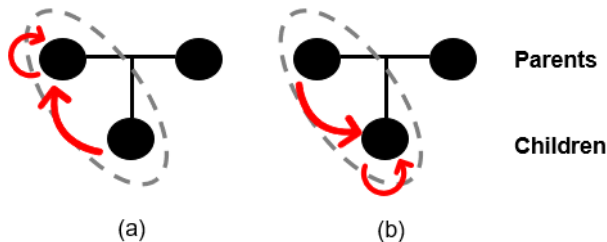


Figure 1: Composition of pairs and structure of information

Data and Variables II

- Non-response: uncodable, refusals and don't know
- Agreement: Exact agreement on 9-level Educational Attainment
- Over (Under) estimation: Proxy reports more (less) years of education than self.

Educational Attainment	Years of Education
None	0
Primary school	5
Middle school	9
Vocational Diploma (Bac-2)	10
Vocational or Technical High School Diploma (Bac)	12
Academic High School Diploma (Bac)	12
Short university degree (Bac +2)	14
Bachelor or equivalent (Bac +3/+4)	15
Master's degree or higher (Bac+5 or more)	17

Table 1: Educational attainment variable

Methods

- To study (dis)agreement:
 - Interrater reliability indices: Cohen's κ and Krippendorff's α
 - Logistic regression
- To study over and underestimation:
 - Logistic and Multinomial regressions
 - Fixed Effects models
- To study consequences on social mobility:
 - Linear regressions
 - Simulations

Table of Contents

- 1 Introduction
- 2 Theoretical background
- 3 Data and Methods
- 4 Results**
- 5 Discussion
- 6 Conclusion

Non-Response - Parents VS Children

	Child	Parent	p-val
Self-reports			<0.001
<i>Reported</i>	420 (100%)	406 (96%)	
<i>Missing</i>	1 (0.2%)	15 (3.6%)	
Proxy-reports			<0.001
<i>Reported</i>	364 (86%)	414 (98%)	
<i>Missing</i>	47 (14%)	7 (1.7%)	

Table 2: Non-Response On Self and Proxy Reports

Non-Response - Educational Differences?

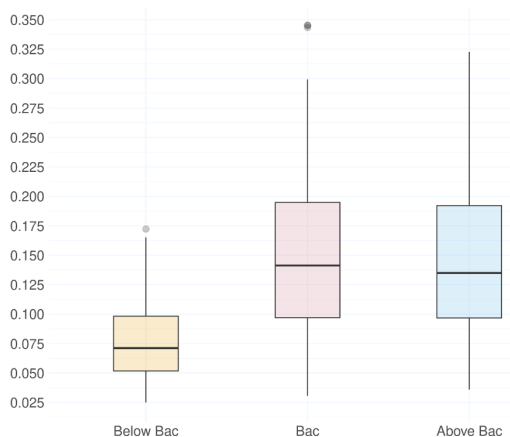


Figure 2: Probability of Non-Response depending on Child Educational Attainment

Looking at Within Pair Agreement [► More](#)

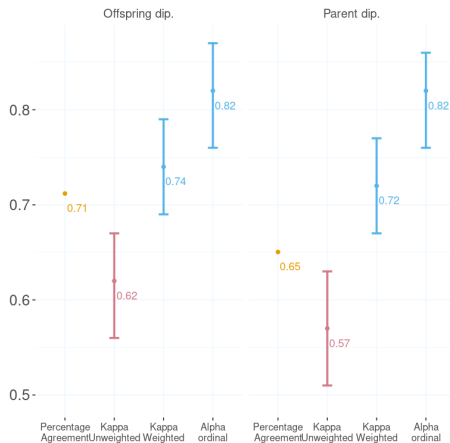
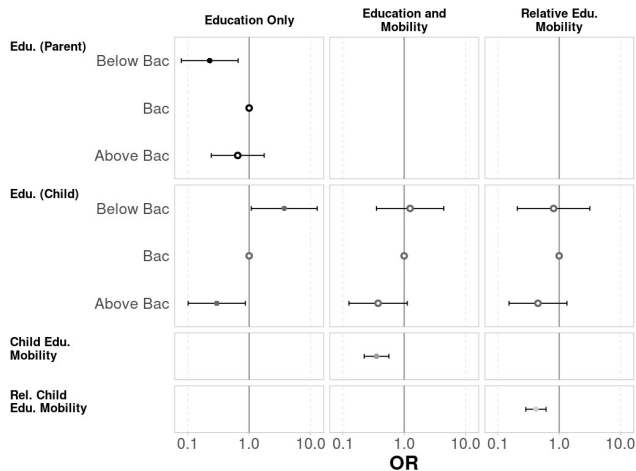


Figure 3: Inter-rater Reliability Indices

Systematic Disagreement? Overestimating Child Edu. [► More](#)



Systematic Disagreement? Underestimating Child Edu. [► More](#)

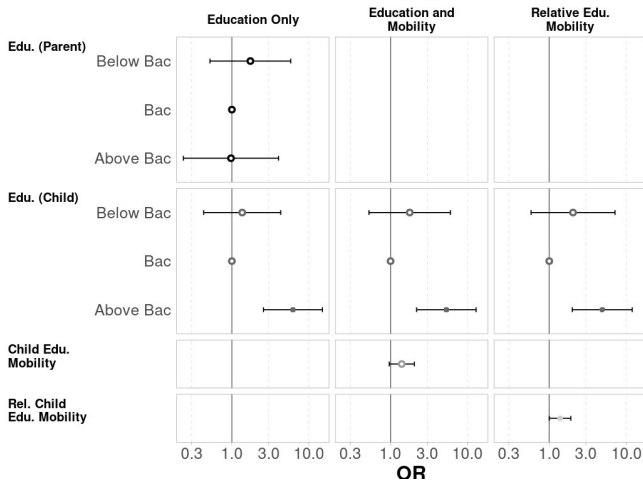


Table of Contents

- 1 Introduction
- 2 Theoretical background
- 3 Data and Methods
- 4 Results
- 5 Discussion**
- 6 Conclusion

Support for hypotheses

<i>Hypothesis</i>		<i>Results</i>
H1a	✓	Children have higher non-response on proxy reports (but not self-reports)
H1b	(✓ &) ✗	No significant difference except for parents' self-reports
H2	(✓ &) ✗	Not significantly, except when taking non-response into account
H3	✓	Parents and Children overall overestimate each other's level of education...
H4	✓	... and try to reduce the "distance" between them

Simulations of Social Mobility

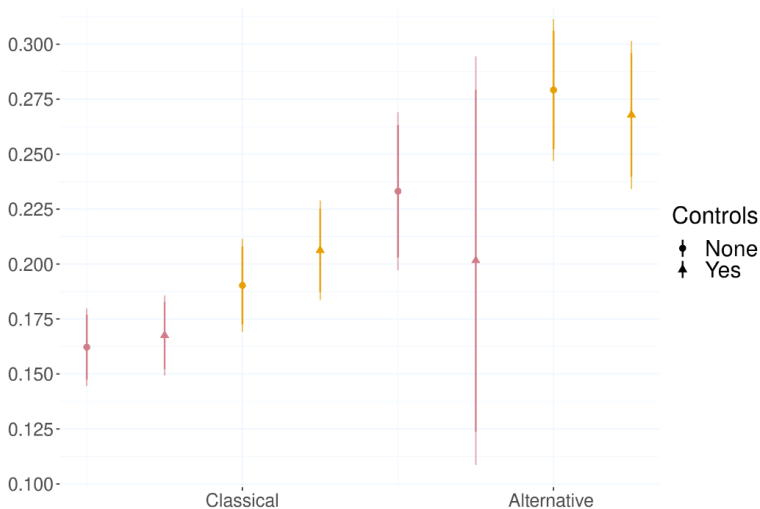
[► Matching](#)

Table of Contents

- 1 Introduction
- 2 Theoretical background
- 3 Data and Methods
- 4 Results
- 5 Discussion
- 6 Conclusion

Limitations

- (Very) Small sample with specific demographics
 - Coarsened variables
 - Lack of variability
 - Mostly immigrant background families, fairly young adults in the children sample
- Lower bounds
- Still differences in variables about parents versus children
- Need to replicate in large panel (GSOEP, BHPS...)

Conclusion

- Nonresponse affects children much more than parents
- Biases in education proxy-reports depend on own and alter's educational attainments
- (Relative) distance between ego and alter explains mistakes: We think we are closer than we are
- Significant impact on social mobility
- Possible to counteract with only self-reported data
 - More countries
 - Possible to leverage parents and children as proxies

Thank you for your attention!

Please send any question/remark to: c.lavest@uu.nl

Data and Variables

- Independent variables:
 - Educational attainment or Years of Education
 - Educational distance: Difference between years of study of child and parent (>0 when child more educated)
 - Objective: difference between two self-reports
 - Subjective: difference between one self- and one proxy-report
- Control variables:
 - sex of parent and child (M/F)
 - student status of child (1/0), age of parent and child (num)
 - cohabitation between parent and child (1/0)
 - status towards migration of child (G2, G2.5, G3+)
 - time lag between parent and child interviews (less than 6 months, between 6 and 12 months, more than 12 months)
 - highest diploma of parent obtained abroads (1/0) [▶ Back](#)

Non-Response - Gender differences

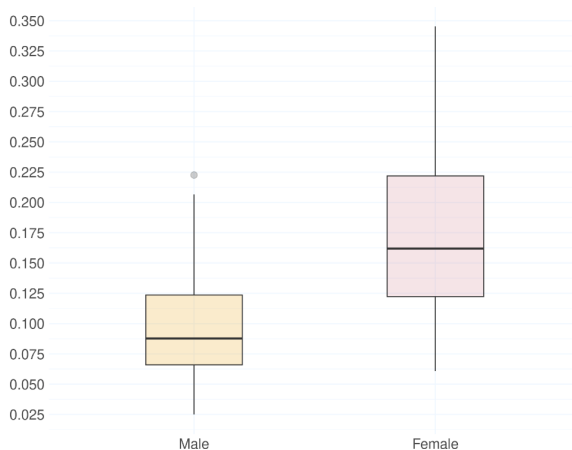


Figure 4: Probability of Non-Response depending on Child Sex ($p < 0.1$)

Agreement - Including Non-Response as a Level [▶ Back](#)

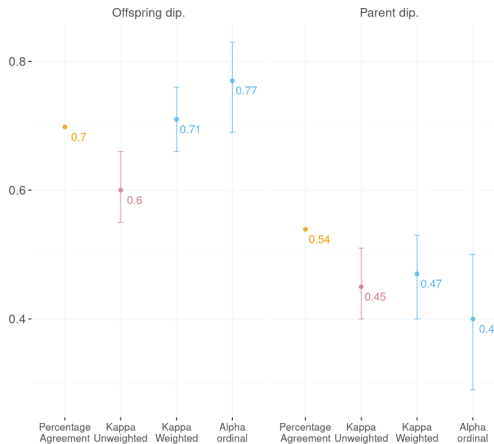
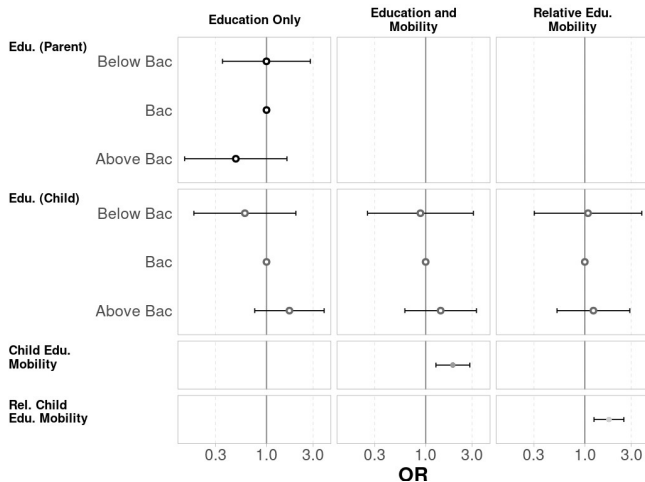
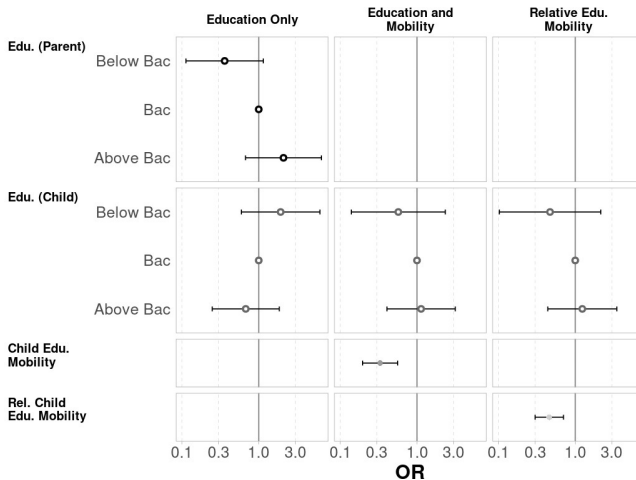


Figure 5: Inter-rater Reliability Indices

Overestimation of Parent Education

[▶ Back](#)

Underestimation of Parent Education

[▶ Back](#)

Social Mobility on Matched Sample [▶ Back](#)

