
MEASUREMENT ERROR IN SELF- AND PROXY-REPORTS OF EDUCATIONAL QUALIFICATIONS: A VALIDATION USING ADMINISTRATIVE DATA

13^{ES} JOURNÉES DE MÉTHODOLOGIE STATISTIQUE DE L'INSEE, 12-14 JUIN 2018

CAROLINE ROBERTS (PROFESSEURE ASSISTANTE, UNIL) ET FIONA MÜLLER (COLLABORATRICE SCIENTIFIQUE, OFS SUISSE)



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la statistique OFS

UNIL | Université de Lausanne

Faculté des sciences
sociales et politiques



BACKGROUND AND MOTIVATION

BACKGROUND

- Growing interest in official statistics in the opportunities presented by combining data sources
 - Survey data and administrative data – but also other sources
 - Hope is to:
 - Improve the quality of the existing data
 - Increase the amount of information available for a unit of interest
 - Reduce burden on survey respondents
 - Answer new research questions



BACKGROUND MOTIVATION

- Opportunities for *methodological* research lie in possibility to validate survey estimates and assess **data quality** in record check studies
 1. To assess the magnitude of different forms of survey error (distance from the 'true' value)
 2. To understand the causes of errors and address them
 3. To adjust / correct for error – or provide alternative, more accurate data sources
- More generally, to improve survey practice

'Fitness for use'

Relevance

Timeliness

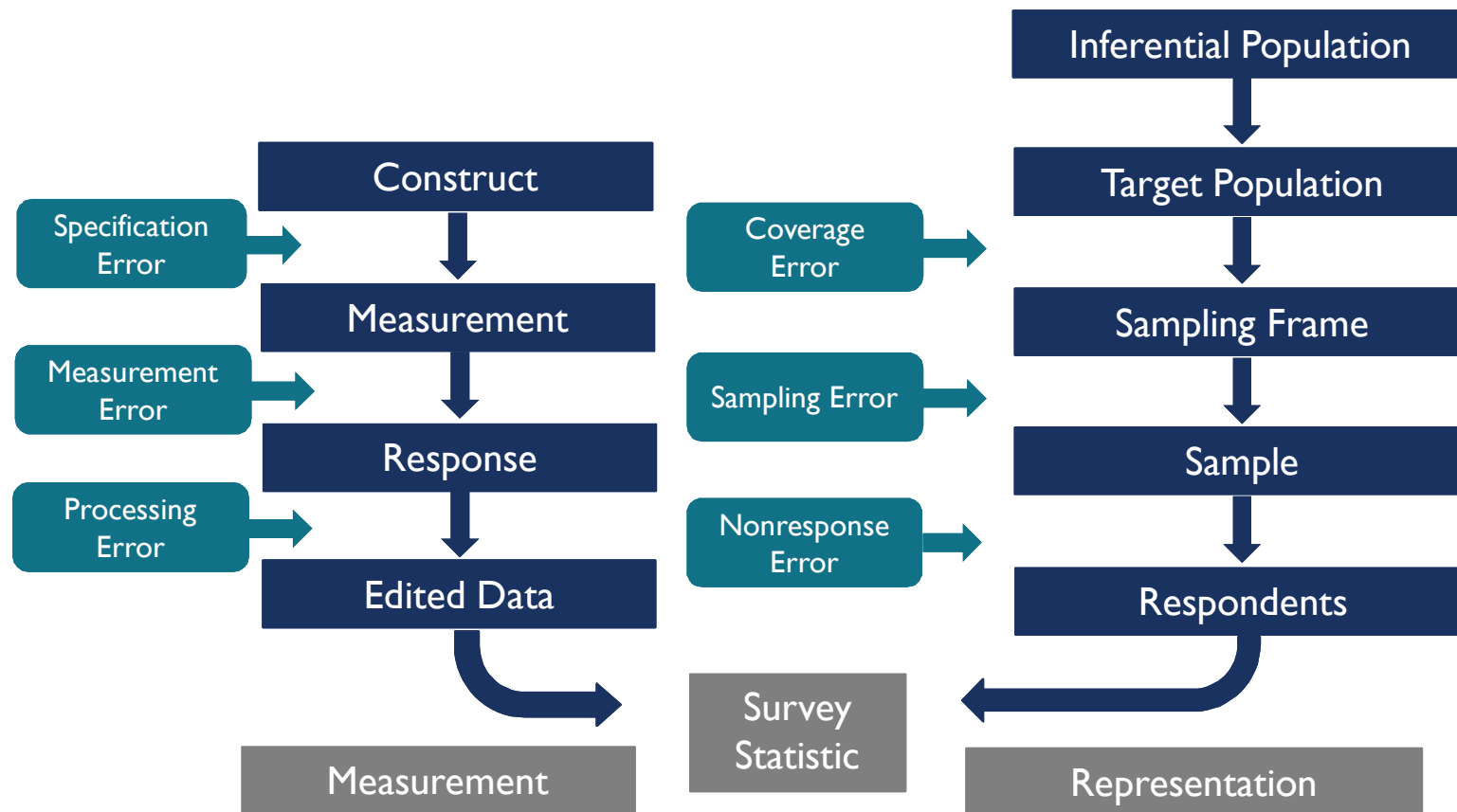
Comparability

Completeness

Accuracy

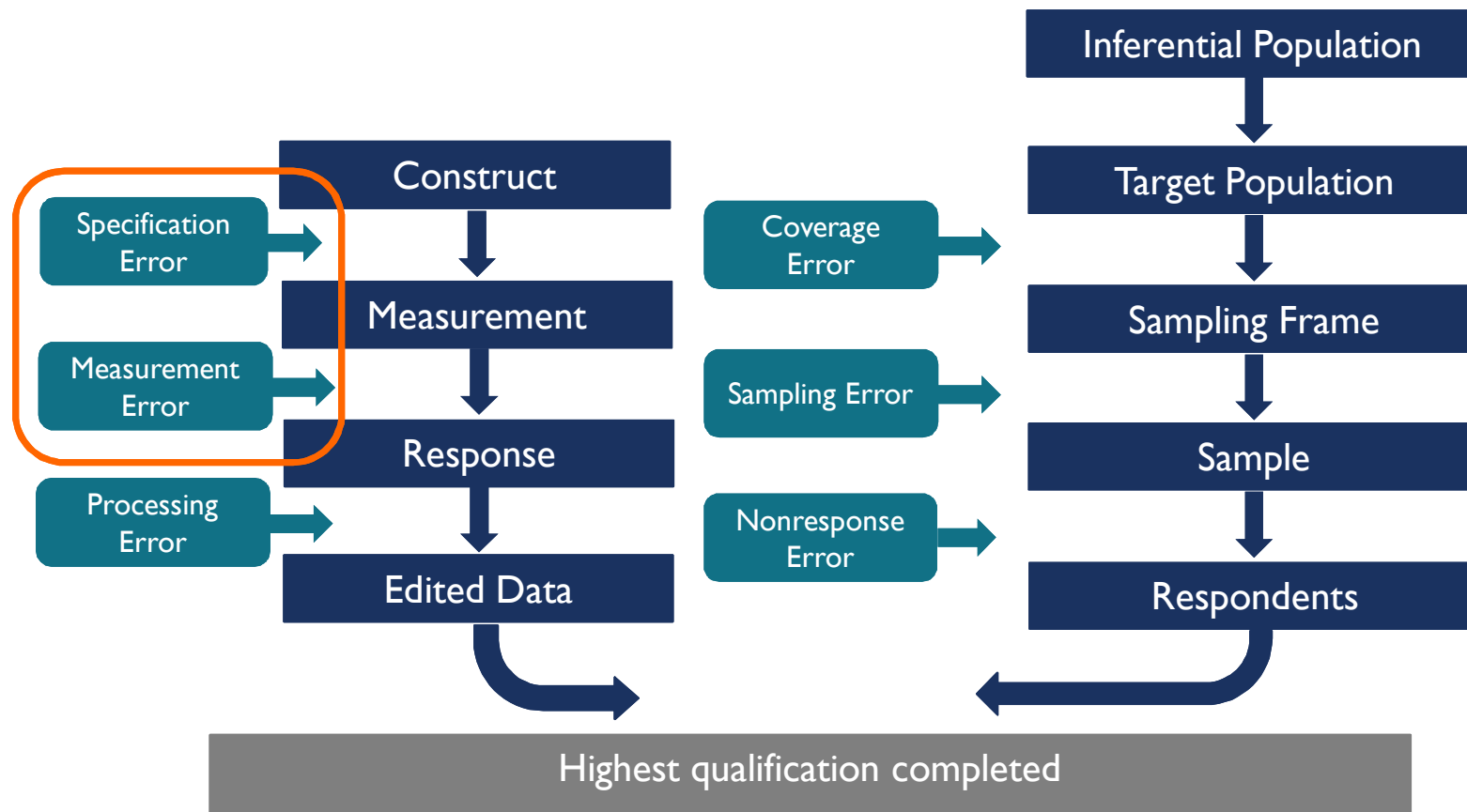
BACKGROUND

TOTAL SURVEY ERROR



BACKGROUND

TOTAL SURVEY ERROR



BACKGROUND

MEASUREMENT ERROR

- “One of the most damaging sources of error” (Biemer, 2011; p.283)
- Particularly in socio-demographic variables like educational qualifications, commonly used:
 - To describe populations (and evaluate survey samples)
 - To adjust survey data (e.g. in survey weights and imputation)
 - As control variables
 - As explanatory variables for diverse phenomena
- Self-reports and proxy reports
- Often assumed to be ‘error-free’ – an assumption rarely verified



BACKGROUND

SOURCES AND TYPES OF MEASUREMENT ERROR

1. Specification errors –

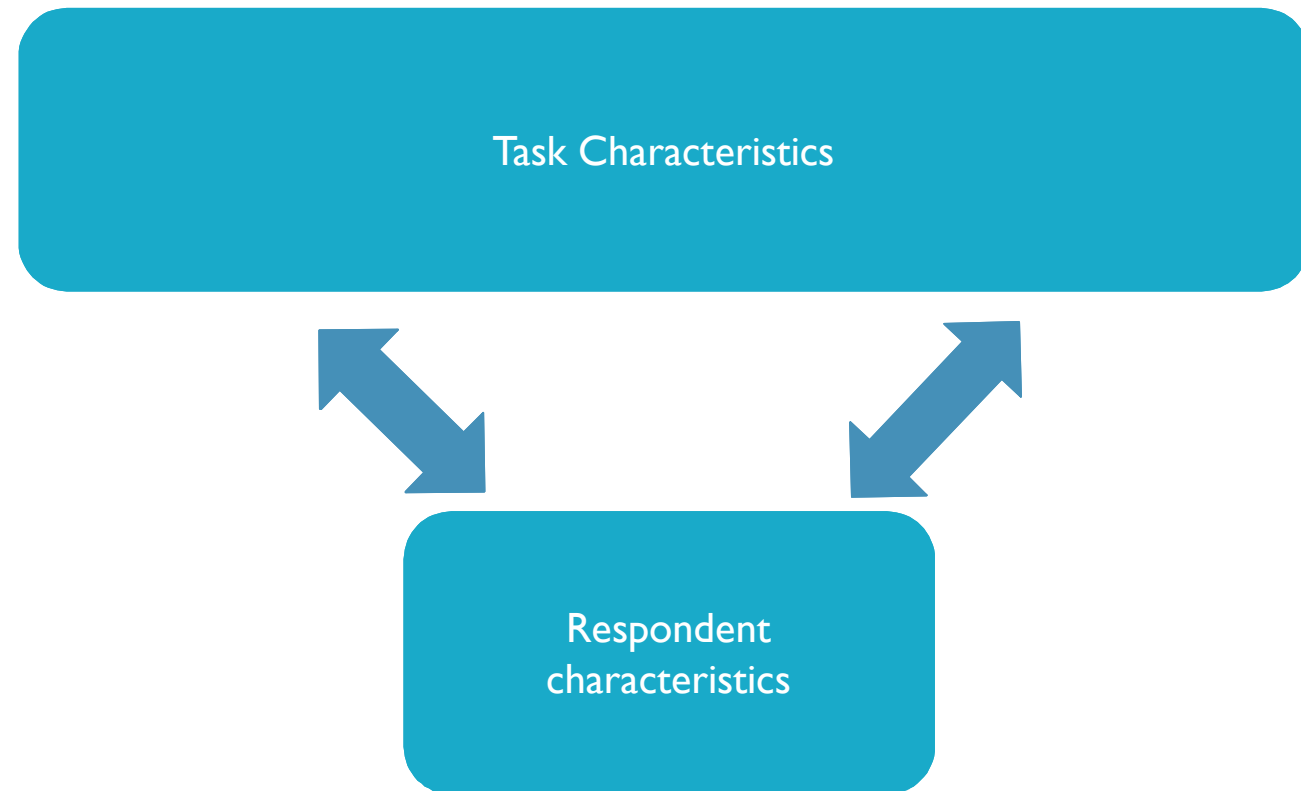
- E.g. specific qualifications, changing nomenclature over time

2. Misclassification errors –

- Education
- Unfamiliarity with Swiss qualifications; language barrier
- Length of recall period (age)

3. Motivated misreporting –

- Social desirability bias;
- Forward telescoping?



RESEARCH QUESTIONS

1. Do people misreport educational qualifications (self- and proxy reports)?
 2. If so, what is the extent and nature of bias introduced by misreporting?
 3. What are possible mechanisms that may account for misreporting?
 - Task vs. respondent characteristics?
- **Investigation of measurement error in self- and proxy reports of educational qualifications through a *forward record check* study (Groves, 2004), comparing survey responses to administrative data on respondents and their household members**



METHODS

METHODS

DATA

Two data sources linked on individual level by social security number (AHVN13):

- **Survey data:** Structural Survey of the Swiss population census (Relevé structurel), years 2012-2014
- **Administrative data:** LABB, register data collected and harmonised for longitudinal analyses in the area of education («Längsschnittanalysen im Bildungsbereich»), years 2012-2014

METHODS

SURVEY QUESTIONS ON EDUCATIONAL QUALIFICATIONS

8. Indicate <u>all</u> education /training you have completed in column a)... (tick all qualifications obtained)	a) Education /training completed	b) Current education /training
9. ... and in column b) enter the education or training in which you are currently engaged. (only one answer)		
1 none	1 <input type="checkbox"/>	1 <input type="checkbox"/>
2 up to 7 years of compulsory education	2 <input type="checkbox"/>	2 <input type="checkbox"/>
3 compulsory education (lower secondary school, middle education, special needs school)	3 <input type="checkbox"/>	3 <input type="checkbox"/>
4 1-year preliminary course, 1-year general education school, 1-year vocational course, language school (at least 1 year) with certificate of completion, social work year, bridging courses or similar	4 <input type="checkbox"/>	4 <input type="checkbox"/>
5 2-3 years higher secondary general education not giving access to universities or similar		
6 basic vocational education and training (2-4 year apprenticeship or full-time vocational school, commercial diploma, technical college, national vocational qualifications NVO 1-3 or similar)		
7 high school-leaving certificate giving access to universities (general baccalaureate) or similar		
8 vocational baccalaureate		
9 advanced technical and professional training (vocational associate's degree, national vocational qualifications NVO 4), higher technical or commercial college or similar		
10 advanced technical and professional training (i.e national vocational qualifications NVO 5)		
11 bachelor's degree (university, ETH/EPF, university of applied sciences, university of teacher education)		
12 master's degree, licentiate, diploma, state examination, post-graduate degree (university, ETH/EPF, university of applied sciences, university of teacher education)		
13 doctorate (Ph.D, Research), post-doctoral qualification (habilitation)		

28. What is the highest educational level completed by the person? (only one answer)	Only to be completed for persons <u>aged 15 and older</u> .	
1 none or up to 7 years of compulsory education	1 <input type="checkbox"/>	1 <input type="checkbox"/>
2 compulsory education (lower secondary school, middle education, special needs school), preliminary course or other bridging courses	2 <input type="checkbox"/>	2 <input type="checkbox"/>
3 basic vocational education and training (2-4 year apprenticeship or full-time vocational school, commercial diploma, technical college, national vocational qualifications NVO 1-3 or similar)	3 <input type="checkbox"/>	3 <input type="checkbox"/>
4 higher secondary general education (high school-leaving certificate giving access to universities, general or vocational baccalaureate, higher secondary general education not giving access to universities or similar)	4 <input type="checkbox"/>	4 <input type="checkbox"/>
5 advanced technical and professional training (vocational associate's degree, national vocational qualifications (NVO 4 or higher)), higher technical, commercial college or similar	5 <input type="checkbox"/>	5 <input type="checkbox"/>
6 university, ETH/EPF, university of teacher education, university of applied sciences, doctorate (Ph.D., Research)	6 <input type="checkbox"/>	6 <input type="checkbox"/>



Current and completed educations of respondents, highest educational qualification of household members

METHODS

EDUCATIONAL QUALIFICATIONS IN ADMINISTRATIVE DATA

Different education registers (data collected by schools, checked, combined and harmonized by SFSO/OFS)

- Statistics on pupils and students (SdL, «Statistik der Lernenden»)
- Statistics of educational qualifications (SBA, «Statistik der Bildungsabschlüsse»)
- Statistics of vocational basic education (SBG, «Statistik der beruflichen Grundbildung»)
- Swiss university information system (SHIS, «schweizerisches Hochschulinformationssystem»)



Completed educations on upper secondary and tertiary level

METHODS

SAMPLE & ANALYTIC APPROACH

- All respondents of the structural surveys of 2012, 2013 and 2014 with a personal entry in one of the education registers (as student or graduate) in the same year (for 2014, also previous years).
- All respondents of the structural surveys of 2012, 2013 and 2014 living in the same households as somebody with an entry in the registers (as student or graduate) in the same year for whom the respondents were required to provide proxy reports.
 - **RQ1 and RQ2: Comparison of self- and proxy reports to administrative data on aggregate and personal level**
 - **RQ3: Comparisons across sub-groups to test hypotheses relating to mechanisms underlying reporting errors (bivariate and multivariate)**

METHODS

SAMPLE COMPOSITION

Year	Type of report				Sex				Age	
	Self-report		Proxy-report		Male		Female		Range	Mean
2012	22433	29.3%	54225	70.7%	39302	51.3%	37356	48.7%	14-74	20.5
2013	22679	30.1%	52738	69.9%	38354	50.9%	37063	49.1%	14-76	20.6
2014	22699	29.5%	54203	70.5%	38929	50.6%	37973	49.4%	14-80	20.7
Total	67811	29.6%	161166	70.4%	116585	50.9%	112392	49.1%	14-80	20.6

METHODS

DEFINITION OF MEASUREMENT ERROR

- **Measurement error** defined as any discrepancy between comparable categories in the two data sources
 - Highest educational qualifications of respondents and household members either **under-** and **over-reported**
 - Beware of “false positives” and “false negatives”
- **Mechanisms** underlying reporting errors investigated by comparing groups:
 - Education
 - Country of birth
 - Language

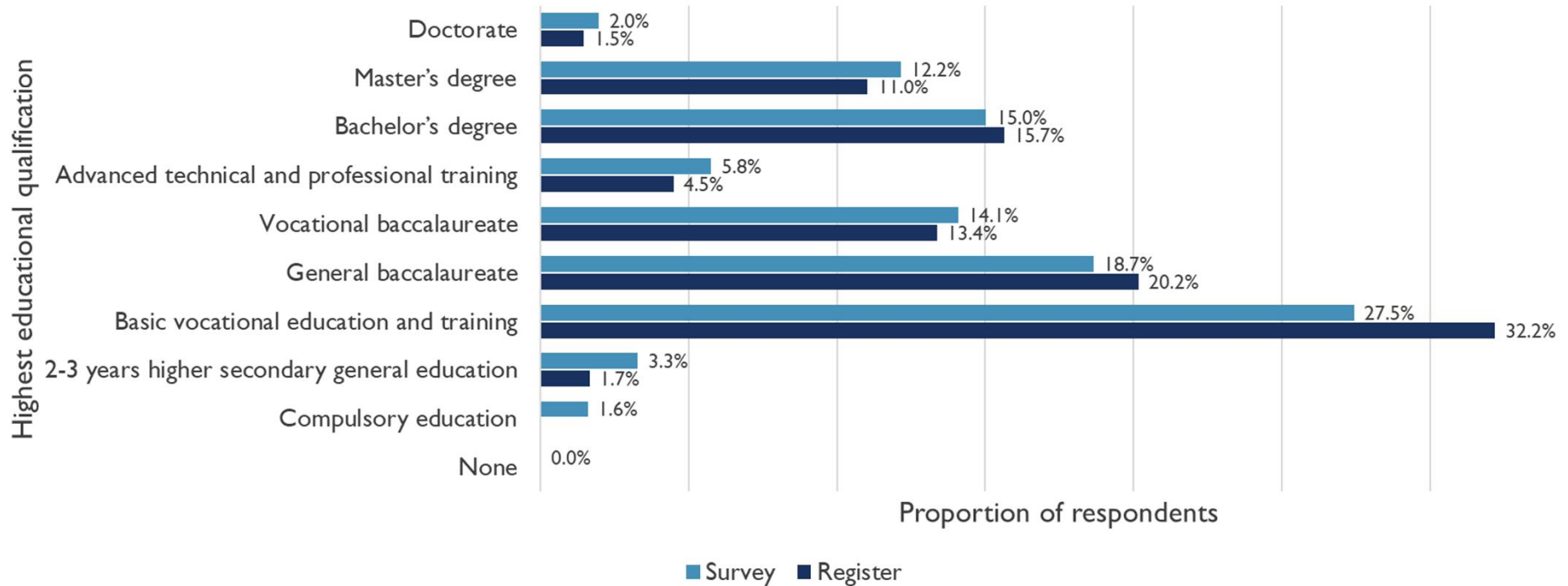


RESULTS

RESULTS

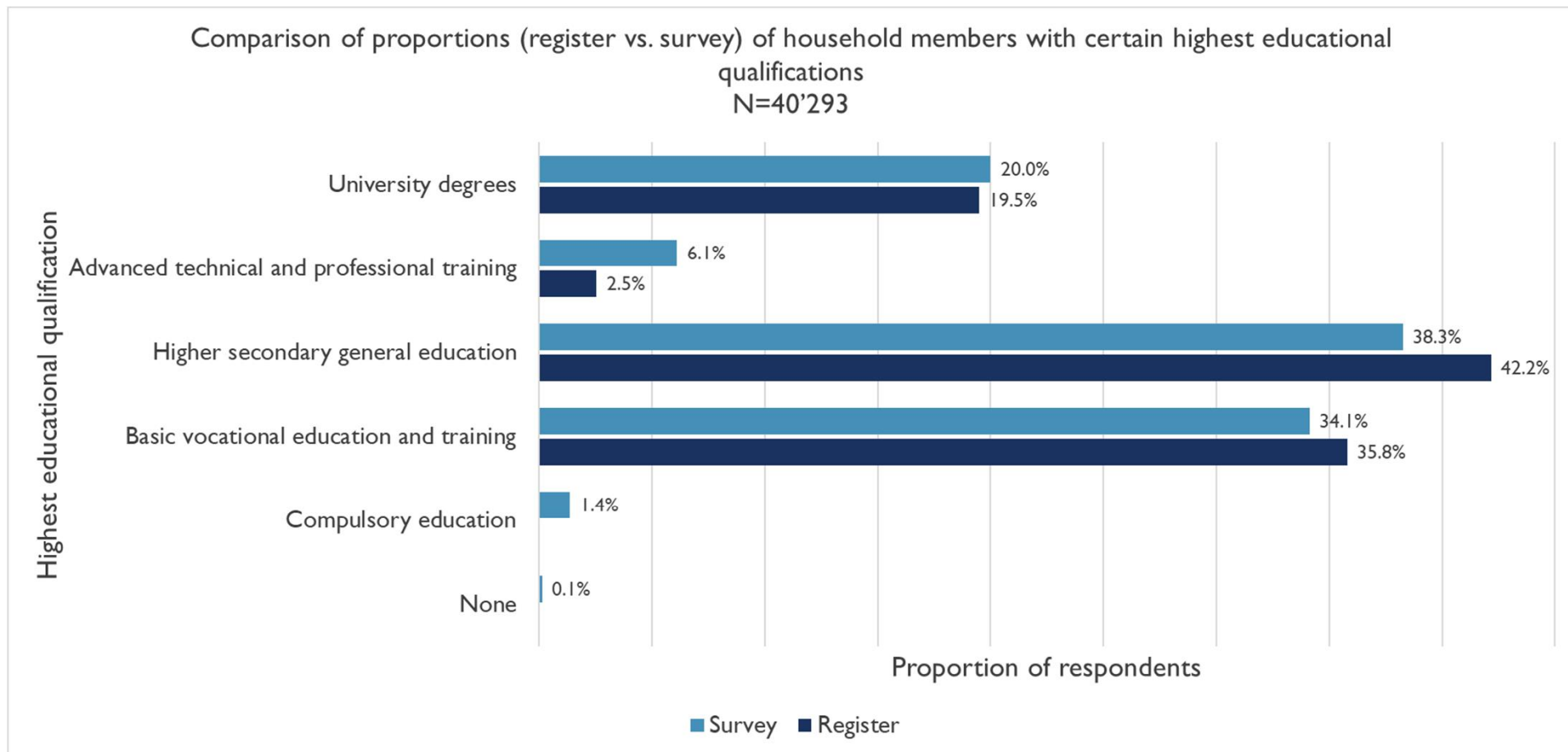
AGGREGATE LEVEL COMPARISONS: HIGHEST EDUCATIONS (SELF)

Comparison of proportions (register vs. survey) of respondents with certain highest educational qualifications
N=18'054



RESULTS

AGGREGATE LEVEL COMPARISONS: HIGHEST EDUCATIONS (PROXY)



RESULTS

PREVALENCE OF MEASUREMENT ERROR

- 11.8% of the respondents misreport their **highest educational qualification**:
 - 7.3% of the respondents **overreport** their highest educational qualification.
 - 4.7% **underreport** their educational attainment.
- 16.1% of the proxy reports concerning the **highest educational qualification of household members** are misreported:
 - 10.7% of the respondents **overreport** the educational level of their household members.
 - 7.6% **underreport** the highest educational qualification completed by their household members.

RESULTS

UNDERLYING MECHANISMS OF MEASUREMENT ERROR

- Bivariate and multivariate analyses lead to similar findings across most types of errors concerning **characteristics of respondents**.
- Positively associated with misreporting are:
 - Lower levels of education
 - Foreign country of birth
 - Foreign language (also within Switzerland, Italian and French vs. German)
- No clear trends are found for **characteristics of the task** (response categories, household size, household type).

RESULTS

FORWARD TELESCOPING?

- Respondents might forward telescope educational qualifications, i.e. report programmes which are not yet ongoing / completed according to registers.
- Potential forward telescoping rates for different error types:
 - 50.8% of overreports of highest educational qualifications of respondents (N=788)
 - 70.3% of overreports of highest educational qualifications of household members (N=2119)

RESULTS SUMMARY

- Point estimates on the aggregate level are quite accurate (max. difference 4.8%).
- Personal-level comparisons result in higher error rates (4.4%-16.1% misreports).
- Multivariate analyses suggest certain underlying mechanisms of measurement error linked to characteristics of the respondents.
- Forward telescoping could possibly explain parts of the observed overreports.



DISCUSSION

DISCUSSION

- Prevalence: Sample restrictions
 - Conservative estimates of measurement error?

- Underlying mechanisms: Interaction of characteristics of the respondents and the task
 - Task difficulty seems linked to respondents' characteristics (education, country of birth, language), but explanations for misreporting (social desirability bias, cognitive abilities, language barrier, unfamiliarity with educational system) cannot be tested.

- “ Conclusions rest on the assumption that the administrative data are free from error
 - À discuter!

MERCI BEAUCOUP POUR VOTRE ATTENTION!

Questions? Remarks?

Contact:

caroline.roberts@unil.ch

Fiona.Mueller@bfs.admin.ch



ADDITIONAL MATERIAL

METHODS

CORRESPONDENCE OF CATEGORIES OF EDUCATIONAL QUALIFICATIONS

- Respondents: upper secondary level

LABB variable	Survey variable
Federal VET certificate (2 years)	
Federal VET diploma (3 years)	
Federal VET diploma (4 years)	6 basic vocational education and training
Other further training	
Not regulated programme	
Upper secondary specialised schools	5 2-3 years higher secondary general education
Baccalaureate schools	7 high school-leaving certificate
Vocational baccalaureate	
Specialised baccalaureate certificate	8 vocational baccalaureate

METHODS

CORRESPONDENCE OF CATEGORIES OF EDUCATIONAL QUALIFICATIONS

- Respondents: tertiary level

LABB variable	Survey variable
Professional education and training colleges diplomas	9 advanced technical and professional training 10 advanced technical and professional training
Bachelor's degree	11 bachelor's degree
Licentiate / diploma (equivalent to the today's Master's degree)	12 master's degree, licentiate, diploma, state examination, post-graduate degree
Master's degree	
Doctorate	13 doctorate (Ph.D, Research), post-doctoral qualification (habilitation)

METHODS

CORRESPONDENCE OF CATEGORIES OF EDUCATIONAL QUALIFICATIONS

■ Household members

LABB variable	Survey variable
Federal VET certificate (2 years)	
Federal VET diploma (3 years)	
Federal VET diploma (4 years)	3 basic vocational education and training
Other further training	
Not regulated programme	
Upper secondary specialised schools	
Baccalaureate schools	4 higher secondary general education
Vocational baccalaureate	
Specialised baccalaureate certificate	
Professional education and training colleges diplomas	5 advanced technical and professional training
Bachelor's degree	
Licentiate / diploma (equivalent to the today's	6 university, ETH/EPF, university of teacher
Master's degree)	education, university of applied sciences,
Master's degree	doctorate (Ph.D., Research)
Doctorate	

RESULTS

PREVALENCE OF MEASUREMENT ERROR

- 2132 of the 18054 (11.8%) respondents misreport their highest educational qualification:
 - 1279 out of 17500 (7.3%) respondents overreport their highest educational qualification.
 - 853 out of 18054 (4.7%) underreport their educational attainment.
- 4626 of the 40293 (16.1%) proxy reports concerning the highest educational qualification of household members are misreported:
 - 853 out of 18054 (10.7%) respondents overreport the educational level of their household members.
 - 3070 out of 40293 (7.6%) underreport the highest educational qualification completed by their household members.

RESULTS

MISREPORTS OF HIGHEST EDUCATIONAL QUALIFICATIONS

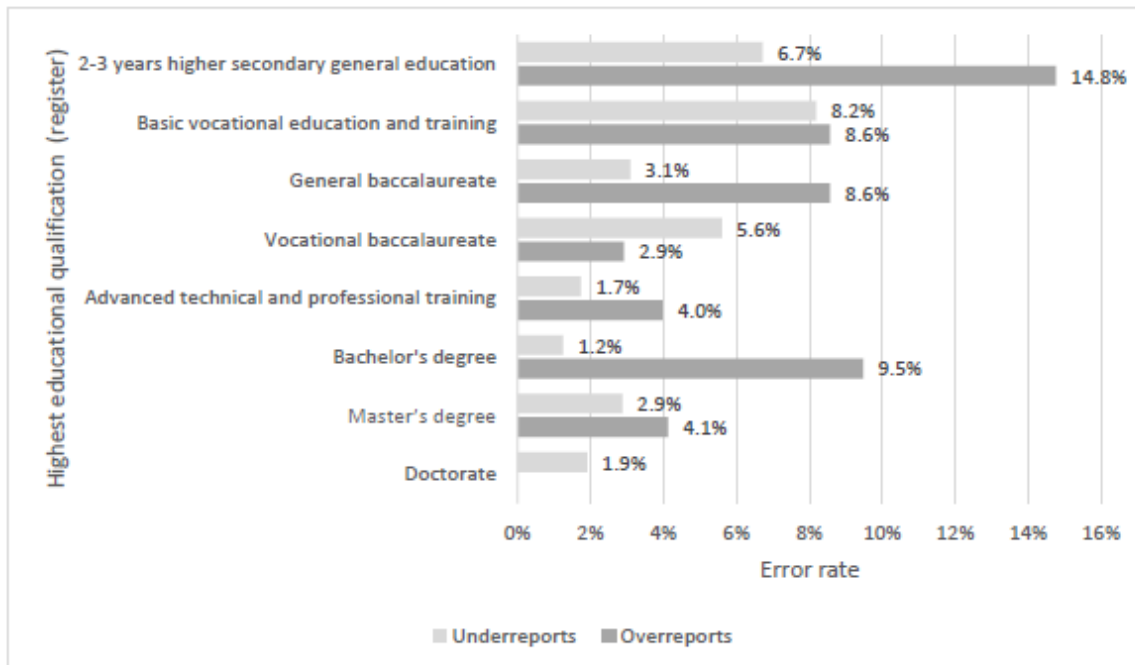


Figure 9: Error rates (misreports of highest educational qualifications of respondents) across highest educational qualifications of respondents (over- and underreports), $N=18054$ (underreports) / 17500 (overreports)

RESULTS

MISREPORTS OF HIGHEST EDUCATIONAL QUALIFICATIONS

Table 17: Error rates (misreports of highest educational qualifications of respondents) across respondent characteristics

	Respondent characteristics	Error rate
Country of birth	Switzerland	9.8%
	Other country	23.2%
Main language	German	9.2%
	French	13.1%
	Italian	23.7%
	Other language	32.5%

N=18054 / 18021

Table 18: Error rates (misreports of highest educational qualifications of household members) across respondent characteristics

	Respondent characteristics	Error rate
Country of birth	Switzerland	13.5%
	Other country	17.7%
Main language	German	12.1%
	French	15.8%
	Italian	22.3%
	Other language	17.2%

N=9605 / 9572

Table 22: Predictors of over- and underreports of highest educational qualifications of respondents

Predictor	Overreports		Underreports	
	B	OR	B	OR
Highest educational qualification				
Basic vocational education and training	-1.012*** (0.1827)	0.363*** (0.0664)	0.319 (0.2405)	1.376 (0.3309)
General baccalaureate	-0.894*** (0.1845)	0.409*** (0.0755)	-0.849*** (0.2528)	0.428*** (0.1082)
Vocational baccalaureate	-1.941*** (0.2124)	0.144*** (0.0305)	-0.085 (0.2506)	0.919 (0.2302)
Advanced technical and professional training	-2.529*** (0.2631)	0.080*** (0.0210)	-1.281*** (0.3632)	0.278*** (0.1008)
Bachelor's degree	-1.412*** (0.1901)	0.244*** (0.0463)	-1.762*** (0.2927)	0.172*** (0.0503)
Master's degree	-2.911*** (0.2197)	0.054*** (0.0120)	-0.980*** (0.2783)	0.375*** (0.1045)
Doctorate			-1.516*** (0.5181)	0.220*** (0.1138)
Country of birth				
Other country	0.730*** (0.0751)	2.075*** (0.1558)	0.569*** (0.0939)	1.766*** (0.1657)
Main language				
French	0.337*** (0.0707)	1.400*** (0.0990)	0.498*** (0.0800)	1.645*** (0.1316)
Italian	1.241*** (0.1006)	3.458*** (0.3478)	0.947*** (0.1239)	2.578*** (0.3194)
Other language	1.104*** (0.1470)	3.015*** (0.4431)	0.744*** (0.1984)	2.103*** (0.4174)
Sex				
Female	-0.013 (0.0624)	0.987 (0.0615)	0.069 (0.0719)	1.071 (0.0770)
Age in years				
	0.096*** (0.0042)	1.101*** (0.0046)	0.001 (0.0066)	1.001 (0.0066)
Constant				
	-3.990*** (0.2016)	0.0185*** (0.0037)	-3.119*** (0.2755)	0.044*** (0.0122)
N	17468		18021	
Log Likelihood	1199.120		445.250	
df	12		13	
Pseudo-R ² (McFadden)	0.131		0.065	

Reference categories: 'Highest educational qualification': 2-3 years higher secondary general education, 'Country of birth': Switzerland, 'Main language: German, 'Sex': Male
*p<0.1 ** p<0.05 *** p<0.01; standard errors in parantheses

Table 23: Predictors of over- and underreports of highest educational qualifications of household members

Predictor	Overreports		Underreports	
	B	OR	B	OR
Highest educational qualification of respondent				
Tertiary level	0.082 (0.1815)	1.086 (0.1971)	-0.442** (0.1863)	0.643** (0.1198)
Household size				
3 people	0.607* (0.3642)	1.835* (0.6682)	0.223 (0.3775)	1.250 (0.4717)
4 people	0.623* (0.3680)	1.864* (0.6861)	0.292 (0.3751)	1.339 (0.5023)
5 people	0.664* (0.379)	1.942* (0.7365)	0.520 (0.3840)	1.682 (0.6458)
6 or more people	0.669* (0.4032)	1.952* (0.7868)	0.742* (0.4025)	2.101* (0.8456)
Reporting for				
parent	-0.013 (0.8690)	0.987 (0.8576)	0.077 (1.0930)	1.081 (1.1810)
child	-1.141*** (0.3338)	0.319*** (0.1066)	0.236 (0.3696)	1.267 (0.4682)
otherwise related	-0.483 (0.4801)	0.617 (0.2960)	0.323 (0.5538)	1.382 (0.7653)
not related	-0.246 (0.3677)	0.782 (0.2874)	0.541 (0.3811)	1.718 (0.6548)
Country of birth				
Other country	0.521** (0.2151)	1.684** (0.3622)	0.003 (0.2288)	1.003 (0.2294)
Main language				
French	0.872*** (0.1650)	2.392*** (0.3947)	0.044 (0.1515)	1.045 (0.1583)
Italian	0.808*** (0.3071)	2.244*** (0.6891)	0.477* (0.2705)	1.611* (0.4358)
Other language	1.501*** (0.4111)	4.487*** (1.8446)	-0.906 (0.7475)	0.404 (0.3021)
Sex				
Female	-0.147 (0.1531)	0.863 (0.1322)	-0.300** (0.1408)	0.741** (0.1044)
Age in years				
	0.009 (0.0156)	1.009 (0.0157)	0.001 (0.0187)	1.000 (0.0187)
Constant				
	-2.519*** (0.4734)	0.081*** (0.0381)	-2.917*** (0.5556)	0.054*** (0.0301)
N	2014		3007	
Log Likelihood	-615.956		-780.885	
df	15		15	
Pseudo-R ² (McFadden)	0.067		0.020	

Reference categories: 'Highest educational qualification of respondent': upper secondary level, 'Household size': 2 people, 'Reporting for': partner, 'Country of birth': Switzerland, 'Main language: German, 'Sex': Male
*p<0.1 ** p<0.05 *** p<0.01; standard errors in parantheses