



HAL
open science

Who Do They Think You Are? Inconsistencies in Self- and Proxy-Reports of Education Within Families

Chloé Lavest, Mathieu Ferry, Mathieu Ichou, Patrick Präg

► **To cite this version:**

Chloé Lavest, Mathieu Ferry, Mathieu Ichou, Patrick Präg. Who Do They Think You Are? Inconsistencies in Self- and Proxy-Reports of Education Within Families. 2024. hal-04790051

HAL Id: hal-04790051

<https://hal.science/hal-04790051v1>

Preprint submitted on 19 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Who Do They Think You Are? Inconsistencies in Self- and Proxy-Reports of Education Within Families*

Chloé Lavest¹, Mathieu Ferry², Mathieu Ichou³, and Patrick
Präg⁴

¹ICS/Utrecht University, Netherlands

²Printemps, UVSQ, Université Paris-Saclay, France

³French Institute for Demographic Studies INED, France

⁴CREST, ENSAE, Institut Polytechnique de Paris, France

November 15, 2024

Abstract

Estimates of intergenerational educational mobility are generally computed using a combination of self- and proxy-reports of one's and one's parents' education. Such reports are easily collected, offering a cost-effective alternative to collecting multiple self-reports or register data. However, the bias that proxy-reports could introduce in the measurement of intergenerational educational mobility is rarely assessed. Our study fills this gap and assesses how reliable people are when they report their parents' or their child's educational attainment. We find

*A previous version of this study was presented at the 2023 ECSR conference in Prague. We thank the discussants at this occasion as well as Ivaylo Petev and Mathis Sansu for helpful comments on previous versions of this draft. Patrick Präg's work is supported by a grant of the French National Research Agency ANR, 'Investissements d'Avenir' (LabEx Ecodec/ANR-11-LABX-0047).

that both parents and children tend to underestimate the educational distance between themselves and their family members, thus inflating estimates of educational reproduction. This trend is larger when children act as proxy-reporters. Another limitation of using children's proxy-reported information is the number of missing answers, which is lower when parents are asked to proxy-report their child's education. In a simulation exercise, we establish that the bias introduced by proxy reports is not negligible, with self-reported intergenerational regression coefficients being 16% higher when a proxy-report is used.

Introduction

Many surveys collect data on respondents' parents' educational attainment and occupation. Although this socioeconomic background information forms the basis of many sociological analyses, the quality of these routinely collected proxy-reports is rarely examined. Indeed, any analysis that includes parental socioeconomic status (SES) could be seriously misleading if low-quality or biased information was used. This is especially the case for studies of inter-generational social mobility, where parents' educational attainment or class is the crucial independent variable.

Few studies analyze how proxy-reporting of parental background drives substantive research findings. Studies that do so demonstrate considerable variation in various aspects of socioeconomic status when proxies are used in place of self-reports (Engzell and Jonsson, 2015; Lee and Lee, 2012; Reynolds and Wenger, 2012; Tamborini and Kim, 2013; de Vries, 2006; de Vries and de Graaf, 2008; Jaspers *et al.*, 2008). Houseworth and Fisher (2020) find that measurement error in occupation and, in turn, class is significantly higher when spouses are used as proxy-reporters. They suggest that consequences on occupational mobility should be assessed, although they have no data to address the issue empirically. Indeed, empirical research focusing on the consequences of proxy-reporting on biases in intergenerational mobility is, to the best of our knowledge, nonexistent.

We examine the extent to which self- and proxy-reports of educational attainment are substitutable, and how differences between the two affect estimates of intergenerational educational mobility. By exploiting a survey design specificity of the recent French 'Trajectoires et Origines 2' (TeO2, Beauchemin *et al.*, 2023) survey, we study dyads consisting of a parent and a child who both give information about themselves and about the other member of the dyad. We focus on educational attainment as our focal variable, as it is widely assumed to be easily proxy-reportable and is used in many analyses in the field of social stratification and mobility. Using self-reports as a baseline to evaluate proxy-reports, we first measure how much self- and proxy-reports of education diverge and for whom in particular. We then assess the conse-

quences of biased proxy-reports on measures of intergenerational educational mobility, both within our data and by using simulations. Doing so, we provide the first account of proxy-reporting quality from both children and parents and the effect of possible biases on common measures of intergenerational mobility.

Multiple factors explain the lack of research on proxy-reports of socioeconomic attainment. First, data are often not available to assess the validity of proxy-reports. Therefore, most research in this field relies on comparing independent populations and is thus unable to assess the actual magnitude of mis-reporting and its consequences (Houseworth and Fisher, 2020; Massagli and Hauser, 1983; Reynolds and Wenger, 2012). Comparisons of proxy-reports to register data can be considered a gold standard. However, only a limited set of countries offers this possibility (Engzell and Jonsson, 2015). Second, educational attainment is often considered a stable and salient characteristic that is reliably reported (Hout and Hastings, 2016). Therefore, proxy-reports of education are often considered unproblematic. Third, the lack of alternatives to proxy-reported data may deter researchers from investigating this issue in depth (Reynolds and Wenger, 2012). However, we agree with Wagmiller’s (2009) argument that it is crucial to have an understanding of proxy-reports’ shortcomings to potentially adapt our analytical strategies.

Our study makes two main contributions to the existing literature. First, thanks to the unique features of our data, we directly compare reports within dyads of parents and children, assessing both nonresponse and misreporting biases. We investigate the quality of information from children reporting on their parents’ educational attainment and we present the first piece of evidence on parents’ proxy-reporting behaviors about their children. The rise of aging surveys such as the Health and Retirement Study (HRS, Sonnega *et al.*, 2014) and its sister surveys all over the world (Lee *et al.*, 2021) collect data from 50+ year-old respondents who then, among other things, report on their (adult) children’s living conditions. Proxy-reports from parents about their adult children are becoming more common. Still, we have little information on their proxy-reporting behaviors. Further, substantive trends within mobility research such as the increasing attention to multigenerational

(e.g. Anderson *et al.*, 2018), prospective (e.g. Skopek and Leopold, 2020), or sibling similarity-based (e.g. Anderson *et al.*, 2024) research make the use of proxy-reported data more common. Our study represents the first comprehensive assessment of potential biases from both perspectives.

Our second contribution is that our study goes beyond assessing biases in nonresponse and (mis)reports: We investigate their consequences on inter-generational educational mobility. We compare social mobility estimates relying on different sources (self- or proxy-reports) and find a consistent overestimation of the link between educational origin and destination when a proxy-report is used.

Background: What do we know about others?

Proxy reports are often used to assess respondents' socioeconomic background. Researchers rely on proxy-reports because retrieving first-hand information is costly and not always possible (e.g. a respondent's parents could be deceased or live in a different country). Most surveys rely on children's reports about their parents, assuming that such a close relationship should enable the child to report their parent's SES accurately. Therefore, most of the literature surrounding the validity of proxy-reports of SES stems from studying parent-child dyads. Researchers conducted such studies with the primary goal of assessing whether very young children were capable of giving valid information about their parents (West *et al.*, 2001). Indeed, issues of understanding and knowledge are assumed to be more crucial among young children than among adults who have completed their education (West *et al.*, 2001). The main takeaways of Looker's early review (Looker, 1989) still summarize most of our knowledge of the situation: age and belonging to a racial minority (most typically Blacks in the US) consistently influence the quality of proxy-reports. Children under 11 years of age give less valid proxy-reports than older children and Black children have less information about their parents' education (especially fathers). Since then, only scattered studies have assessed inconsistencies in self- and proxy- reports in parent-child dyads.

In Table 1, we synthesize relevant information on studies published after

Looker’s review detailing differences between self- and proxy-reports of SES in parent–child dyads.

Table 1: Current knowledge about parent–child SES reports discrepancies. Studies published since [Looker \(1989\)](#).

Study	Data Country	Age children	Variable	Result
Hovestadt and Schneider (2021)	NEPS (2008–13) <i>DEU</i>	15–6 y.	Father’s edu.	% agreement = 63 Cohen’s κ = 0.50 % nonresponse = 25
			Mother’s edu.	% agreement = 66 Cohen’s κ = 0.52 % nonresponse = 21
Wittrock et al. (2017)	2012 Qatar Education Study <i>QAT</i>	17–8 y.	Parental edu.	% agreement = 68–87
Engzell and Jonsson (2015)	CILS (2010–1) <i>GBR DEU NLD SWE</i>	14–5 y.	Parental edu.	Cohen’s κ = 0.58–0.70 % nonresponse = 1–13 (chi.), 20–78 (par.)
			Parental occ.	Pearson’s r = 0.37–0.61 % nonresponse = 3–16 (chi.), 24–81 (par.)
Jerrim and Micklewright (2014)	PISA (2006, 2009) <i>DEU, KOR, NZL, ITA, ISL, DNK, LUX, PRT, POL, TUR</i>	15 y.	Father’s occ.	Pearson’s r = 0.37–0.61 % agreement = 58–85
			Parental edu.	Cohen’s κ >0.60 % nonresponse = 7–40 % agreement = 51–85 Cohen’s κ >0.40 and <0.60 % nonresponse = 5–45
Ridolfo and Maitland (2011)	Add Health <i>USA</i>	19 y.	Mother’s edu.	% agreement = 66–71
Kreuter et al. (2010)	PISA (2000) <i>DEU</i>	15 y.	Father’s edu.	Cohen’s κ = 0.60 % agreement = 55.8–69.9
			Mother’s edu.	Cohen’s κ = 0.42–0.59 (0.48–0.62, weighted) % agreement = 50.9–74.3

Table 1: continued

Study	Data Country	Age children	Variable	Result
				Cohen's $\kappa = 0.35$ – 0.63 (0.40 – 0.66 , weighted) $R^2 = 0.8$
Wagmiller (2009)	PSID (1968–93) USA	≈ 22 y.	Father's edu.	
			Mother's edu.	$R^2 = 0.6$
De Vries and de Graaf (2008)	FSDP (1992, 1998, 2000) NLD	25–54 y.	Father's edu.	Cronbach's $\alpha = 0.93$
			Father's occ.	Pearson's $r = 0.80$ – 0.85 Cronbach's $\alpha = 0.93$ Pearson's $r = 0.78$ – 0.86
Vereecken and Vandegheuchte (2003)	Authors' data BEL	11–2 y.	Father's Activity	% agreement = 66–78
			Mother's Activity	Cohen's $\kappa = 0.58$ – 0.72 % agreement = 79–84
Lien <i>et al.</i> (2001)	NLHBS (1990) NOR	13 y.	Father's edu.	Cohen's $\kappa = 0.71$ – 0.77 % agreement = 57
			Mother's edu.	Cohen's $\kappa = 0.38$ % nonresponse = 14 (chi.), 0 (par.) % agreement = 51 Cohen's $\kappa = 0.3$ % nonresponse = 13 (chi.), 0 (par.)
West <i>et al.</i> (2001)	West of Scotland 11 to 16 Study (1994–6) GBR	11 y.	Father's activity	Cohen's $\kappa = 0.57$ – 0.59 (0.69 – 0.76 weighted)
			Mother's activity	% nonresponse = 1.8 (chi.), 5.4 (par.) Cohen's $\kappa = 0.65$ – 0.72 (0.82 weighted) % nonresponse = 0.6 (chi.), 4.6 (par.)
Ensminger <i>et al.</i> (2000)	CHIP-AE (1991–3) USA	11–8 y.	Father's edu.	% agreement = 58.1–88.2
			Father's emp.	Cohen's $\kappa = 0.43$ – 0.82 % agreement = 66.7–93.3
			Mother's edu.	Cohen's $\kappa = 0.35$ – 0.77 % agreement = 70–82.3
			Mother's emp.	Cohen's $\kappa = 0.47$ – 0.70 % agreement = 76.5–80.7

Table 1: continued

Study	Data Country	Age children	Variable	Result
Cohen's $\kappa = 0.58-0.69$				
<i>Notes:</i> '% nonresponse' and '% agreement' are percentages. Unless otherwise specified, nonresponse concerns <i>children</i> who select 'I don't know.' When parents are also included, we display both figures. '% agreement' indicates the percentage of dyads giving similar answers. Cohen's κ is an index of inter-rater reliability, varying between 0 (no agreement beyond chance) and 1 (perfect agreement). See Cohen (1960) for more information about this index.				

Most studies focus on parental occupation and education and include children below the age of 16. We only found four studies that investigate individuals aged 18 or older as proxy-respondents of their parents' SES ([de Vries and de Graaf, 2008](#); [Ridolfo and Maitland, 2011](#); [Wagmiller, 2009](#); [Wittrock *et al.*, 2017](#)). The lack of research on adult proxy respondents likely reflects the belief that stable SES indicators, such as education, are easily reported by proxies, alongside the limited availability of cost-effective alternatives to proxy-reporting.

We argue that such premises should be challenged. [Wagmiller \(2009\)](#) for example finds that children's proxy-reports on mother's education only accounts for 60 per cent of the variance in self-reported education of mothers. This is in line with the result of [Ridolfo and Maitland \(2011\)](#), who find a relatively modest agreement (Cohen's $\kappa = 0.60$) between children and mothers when asked about mothers' education. [De Vries and de Graaf \(2008\)](#) find stronger agreement using a different operationalization, but they do not rely on direct comparisons within parent-child dyads. Considerable differences in reports do arise when adult children are asked about their parents' education, a major indicator of SES background. Moreover, parental proxy-reports about their children are never investigated by these studies. The recent development of older populations' surveys and the considerable number of surveys asking about one's children should urge us to do otherwise. Thus, we present below our framework to analyze proxy-reports of adult children and parents on one another's SES.

Theoretical framework and hypotheses

Access to information

In order to be reported in a survey, information on one's SES must be known to the respondent. In that regard, self-reporters are assumed to always have high-quality information on their own degrees, as they have direct experience of their own education and have material evidence of it. This is not automatically the case for proxy-reporters. In general, a proxy-reporter does not necessarily know the educational level of the target¹. This knowledge can either be similar to the self-report, or dissimilar to it.

Part of the inconsistencies between self- and proxy-reports are due to random errors. However, we argue that they also follow from systematic biases. Proxy-reporters may make systematic mistakes based on inference strategies they use when they try to report their parent/child's education. These mistakes may be triggered by the environment of the survey, which encourages individuals to provide answers (Tourangeau *et al.*, 2000). However, in certain situations, respondents can choose to report that they do not know about the information at hand. This may be because they believe they do not know the correct information, or are unsure about it. This strategy might affect children more than parents.

Whereas children are dependent on their parents' to obtain information about their educational attainment, parents usually experience their child's education: they enroll them in school in earlier years, may participate in school-related activities, and discuss school-related matters, especially mothers (Fingerman *et al.*, 2020; Kayser and Summers, 1973; Wagmiller, 2009). Parents thus hold direct information on their child's education. Additionally, because children are younger than their parents, information about their highest degree is likely to be more recent and salient than that of parents. Children only have distant information about their parents' education: their parents

¹In the remainder of this manuscript, we use the word target when referring to the individual whose characteristic we want to know about. We use the word respondent to designate the person who answers survey questions. Respondents can provide self-reports, if they give information on themselves, or proxy-reports if they give information on another target individual.

usually finished their studies before their children were born, and therefore the children do they have first-hand knowledge of the educational system that their parents experienced. Difficulties in transposing their parents' education in the current educational system further complicates proxy-reporting for children. Information on their parents' education is thus both subject to stronger memory decay and depends more on contact opportunities. Overall, children should be less likely to know their parents' 'true' education, and thus more likely not to report it. We expect:

H1: Parents' nonresponse on their children's educational attainment is lower than children's nonresponse on their parents' educational attainment.

The French educational system is intricate, and horizontal differentiation has become more prevalent within higher education (Ichou and Vallet, 2013). This growing complexity of paths available to students makes the experience of the educational system an asset to navigate available options. Compared to countries where students are stratified very early into paths leading to predefined types of academic or vocational training (such as the Netherlands or Germany, Forster and van de Werfhorst, 2019), France has a much more open educational system, where important choices can be made later on in the educational career. Thus, higher educational attainment means greater time spent in education, and often attending more (diverse) schools and education institutions, leading to a better understanding of the complex educational landscape (Erikson and Jonsson, 1996).

Moreover, individuals with higher educational credentials have on average undergone more choices of educational pathways and options. For children, this means they had more opportunities to talk about these steps with their parents, discuss their parents' own educational experiences, and gather advice from them. When parents are highly educated, they usually are able to provide more school-related support, and feel more legitimate in interactions with teachers (Crosnoe, 2001; Smith, 2006; West *et al.*, 1998; Zhang *et al.*, 2011), which can foster conversations and information sharing. We then expect that:

H2: People with higher education display less nonresponse in proxy-reports.

What then happens when people with ‘wrong’ information still report it? The proxy-reporter may believe that they know the ‘true’ information. Alternatively, they can believe to have sufficient information to make a guess and infer this information based on other relevant knowledge. For example, [Wagmiller \(2009\)](#) shows that children tend to report their parents’ educational levels as being more similar to each other than they actually are. This is because they use their knowledge about one parent to recall the educational attainment of the other parent. Such strategies to recover missing information are more likely among children because they have less information about their parents. Most results presented by [Looker \(1989\)](#) show that fathers’ education tend to be reported more accurately than mother’s education. This tendency is not as clear in more recent results as can be seen in Table 1. Assuming that fathers’ education is recognized more easily than mothers’ education, this should drive up reports of mothers’ education—as they are, for the period studied, less educated than fathers overall. Children moreover face a more difficult task than their parents in trying to recover the true information, for it is older and thus less accessible. Finally, structural changes influence occupation–education linkages. In European countries, average levels of educational attainment have greatly increased over time ([Breen and Müller, 2020](#)). Credential inflation changed the typical level of education an occupation requires ([Araki and Kariya, 2022](#); [Collins, 2019](#)). Hence, current information on the occupation–education linkage does not reflect the opportunity structure of previous generations. This can lead children to overestimate their parents’ education, considering that occupation is a more salient feature than education ([Looker, 1989](#), and see Table 1). Because parents are knowledgeable about the previous and current educational landscape, their guesses are less likely to present such a bias. We hypothesize:

H3: Children proxy-reporters overestimate the level of education of their parents.

Intergenerational change and over or underestimation

Gauging someone's education is a relative task. Parents and children could serve as a reciprocal reference group for comparison when reporting education. Someone else's education can be (un)consciously compared to one's own education especially when information about the former is uncertain. Then, intergenerational changes may well influence proxy-reports of SES. Some studies suggest that individuals who have experienced upward social mobility may downplay their own advancement because they have a broader awareness of the educational landscape (Merllié, 2008; Gugushvili, 2021). By comparing their personal journey to a larger social context, they perceive their mobility as less significant than it truly is. Conversely, those who have not experienced social mobility tend to magnify modest intergenerational differences, as their knowledge is limited to a much smaller social sphere (Merllié, 2008). Experiences of significant social mobility may also create tensions between social origin and destination, prompting individuals to downplay the distance between the two in their narratives. The resolution of these conflicts can take various forms, from staying 'truthful' to one's social origins while still embracing one's social destination, to fully 'make it' Naudet and George (2018), i.e. belong to one's social class of destination. If the emphasis is on the former, mobile individuals may be more likely to reduce the distance between themselves and their parents when asked about educational attainment to maintain coherence and increase similarity with their background. This would lead to an artificial decrease in estimated mobility across generations. For example, De Vries and de Graaf (2008), who worked with dyads of parents and children, found that the latter underestimated the educational distance between them and their parents. Additionally, in the face of missing information, respondents rely on available information to answer. As mentioned in the previous section, Wagmiller (2009) shows that an increased covariance between the father's and mother's educational attainment arises in children's reports to compensate for their lack of knowledge. Relying on one's own education as a clue to gauge the other's educational attainment would also entail an underestimation of the distance between self and the target, also biasing mobility estimates downward.

Conversely, if the emphasis is on the latter strategy (i.e. fully ‘making it’), the social distance may increase when using proxy-reports, leading to an increase in estimates of social mobility. Irwin (2015) underlines this pattern where those who ‘make it’ tend to emphasize their own achievements over chance or privileged upbringing. In line with these findings, Friedman *et al.* (2021) found that many respondents, when asked about their class origins, tend to identify as coming from a working-class background, despite objective measures placing them in middle- or upper-class backgrounds. However, no quantitative evidence of such behavior has been found yet, the former focusing on the discourses of individuals, and not their answers in surveys.

The impact of proxy-reports on the distance between parents and children varies depending on whose information is used (Broom *et al.*, 1978). Parents tend to reduce the distance between themselves and their children, while children overestimate this distance, suggesting a different effect of educational distance for parents and children as proxy-reporters. The incentives to under- or overestimate the distance between self and the target may vary depending on the respondent and how desirable it is for them to enhance their achievements relative to the target.

Building on the available research on the link between social mobility and agreement, we put forward two alternative hypotheses:

H4a: Proxy-reports minimize the educational difference between the respondent and the target of information.

H4b: Proxy-reports emphasize the educational difference between the respondent and the target of information.

Data and methods

Data and variables

Our data stem from ‘Trajectoires et Origines 2’ (TeO2, Beauchemin *et al.*, 2023), the second wave of a French survey conducted by INED and INSEE in 2019 and 2020 ($N = 27,181$). The survey focuses on immigrants and their

descendants, while including individuals without any immigrant ancestry as a comparison group. TeO2 is composed of a nationally representative sample randomly drawn from the French population census. Immigrants and children of immigrants are overrepresented. The data used for this study comprise parent–child dyads in which both individuals were administered the (same) questionnaire during separate face-to-face interviews.

Parent–child dyads are constructed based on two approaches. 241 dyads were purposely included at the data collection stage. TeO2 sets out to survey the immigrant-origin population, including third generation individuals (i.e. grandchildren of at least one immigrant), specifically from non-European origin. Because a strictly random sample of the general population would have included very few such individuals, a specific indirect sampling strategy was carried out to reach them. Interviewers were directed to ask second-generation interviewees from non-European origins to provide the contact information of their adult children (aged over 18) and interviewed those children. We thus have independent interviews for parents and children. An additional 180 dyads were included by chance. Multiple individuals from the same household were sometimes independently drawn to be part of the survey. If two or more people in one household completed the survey, we were able to connect potential parents and children. We added them to our analytical sample totaling 421 parent–child dyads.

Our data comprise children’s reports of parental information, as well as parents’ reports of child-related information. Indeed, respondents of the survey were asked about their educational attainment, their parents’ educational attainment, and—if they had any—their adult children’s educational attainment.

Importantly, we label the dyads under study as ‘parent–child dyads’ to highlight the family bond that unites the respondents, but our sample comprises only adult children aged 18 and over². Further, some parents have up to four children in our final sample, which explains the lower total number of parents versus children’s observations (Table 2). Table S2 in the Supplementary

²The survey was limited to people living in France between the ages of 18 and 60. This means that the parents in our dyads are all 60 or younger.

Materials reports this and other dyad-level characteristics.

Table 2: Descriptive statistics. Percentages (number of cases) except for age

	Child, $N = 421$	Parent, $N = 331$
<i>Gender (% Female)</i>	51 (215)	59 (194)
<i>Age, M (SD)</i>	23 (4)	51 (5)
<i>Migration Status</i>		
Immigrant	0 (0)	39 (130)
Child of 2 immigrants	27 (115)	32 (105)
Child of 1 immigrant	19 (78)	22 (72)
Other	54 (228)	7 (24)
<i>Educational attainment</i>		
None	5 (23)	24 (79)
Primary school	0 (1)	1 (3)
Middle school	6 (25)	8 (25)
Vocational Diploma	8 (35)	30 (98)
Vocational or Technical High School Diploma	26 (111)	8 (26)
Academic High School Diploma	20 (83)	6 (19)
Short university degree	9 (38)	8 (27)
Bachelor or equivalent	14 (58)	8 (25)
Master's degree or higher	11 (46)	8 (26)
(Missing)	1	3
<i>Country of Highest Diploma (% from foreign country)¹</i>		10 (33)
<i>Number of children¹</i>		
1		77 (255)
2		19 (64)
3		3 (10)
4		1 (2)
<i>Coresidence with parent (% Yes)¹</i>	72 (302)	
<i>Student (% Yes)¹</i>	37 (155)	
(Missing)	0	331

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Note: weighted percentages (N) unless otherwise specified.

¹Omitted.

Outcome variables

We test hypotheses both on nonresponse in self- and proxy-reports of education and disagreement between the two. Both parents and children report their own educational attainment following the same protocol: multiple questions regarding the type of education, the number of years spent in education in France and abroad, and degrees completed are used to assign one educational level that we recode to match the educational categories available in proxy-reports. Parents are also asked to report the highest diploma of each of their children among a predefined list of eleven educational levels. Children are similarly asked to report both of their parents' highest diploma using nine predefined educational levels. More information on these variables can be found in the Supplementary Materials (Table S1).

Non-response comprises individuals who spontaneously reported 'I don't know' to questions about the education of the target of information. Refusal to answer was also allowed but never chosen by our respondents. For self-reports, there are in fact no 'I don't know' responses. However, four observations were disregarded due to non-codable levels of education.

Measures of *disagreement* are of two types. The first one is a dummy variable contrasting exact agreement (1) between the self-report and the proxy-report, and disagreement (0). The second one distinguishes between overestimation, agreement, and underestimation (as a categorical variable). Both disagreement and (under) overestimation are defined based on the self-report: a proxy-report is an overestimation (respectively underestimation) if the proxy-reporter gave the respondent a higher (respectively lower) education than the one that was self-reported. We consider higher degrees those that necessitate more years of study. This means that two degrees that necessitate the same number of years will be considered similar in this categorization, but not when studying exact agreement. Exact agreement means that the proxy-reporter selects the

same degree as the self-reporter. Practically, compared to the binary exact agreement measure, this categorical measure groups together two diplomas that require the same number of years of study (academic high school diploma and vocational or technical high school diploma). Therefore, agreement in the categorical measure is slightly different from agreement measured in the exact agreement (dummy) measure.

Independent variables

Educational attainment is defined by the self-reported highest diploma obtained by one individual. In some analyses, we use coarsened educational variables to maximize power. We thus distinguish only between individuals with a degree lower than a secondary degree, a secondary degree, or a post-secondary degree. However, unless otherwise specified, we define education as a continuous variable, based on the number of years needed to obtain a degree (from 0, when one has no education, to 17 when one has a master's degree or more).

To investigate the impact of *intergenerational social mobility*, we need a measure of intergenerational educational change. Because we have both self- and proxy-reports of both dyad members' education we compute three different measures of educational difference between generations. First, 'objective difference' is the difference between a child's educational attainment and their parent's in years of education based on self-reports only. We thus compare one report from the child to one report by the parent. Because children are mostly more educated than their parents, it is mostly positive. In addition, we compute two 'subjective differences.' The idea behind the variable is the same, but it is based either on parents' reports or on children's reports only. It represents the subjective difference between the dyad members, based on either member's perceptions. Both the objective and subjective measures are standardized. Results with unstandardized variables are similar and are reported in the Supplementary Materials, Figure S1.

Control variables

We control for co-residence between parent and child because parents and children in those pairs have more contact and are thus less likely to either avoid answering or to make mistakes. Some parents are first-generation immigrants and may have completed their education in their country of origin, which make it more difficult for children to translate their diploma in the French educational categories. Such biases are beyond the scope of the present study. We thus control for completion of education abroad (only for parents). When studying disagreement, we also control for students who are still in education with a dummy variable and for the time lag between parents' and children's interviews (less than six months between the two, between six and twelve months, and more than twelve months). We do so to avoid inflating errors due to the time difference in parents and children's answers, as children can complete an additional degree.

We finally control for three socio-demographic variables in all analyses, age (in quartiles), gender, and migration status (immigrant, child of two immigrants, child of one immigrant, or native with two native parents). We add these controls for both children and parents.

Analytical plan

Our analysis proceeds in five steps. In a first step, we compare nonresponse rates across children and parent proxy-reporters to test Hypothesis H1. We use chi-squared tests with [Rao and Scott's \(1984\)](#) second-order correction to account for the fact that some parents have more than one child in the sample. In a second step, we model nonresponse behavior by children proxy-reporters as a function of their education to test Hypothesis H2. For this, we use a logistic regression model with not reporting one's parent's educational attainment as the outcome, children's education as the key independent variable, and the previously mentioned control variables. We derive predicted probabilities from the model for illustration of results ([Arel-Bundock *et al.*, forthcoming](#)). Standard errors are obtained using the delta method for the difference between two coefficients.

In a third step, we assess the agreement between self- and proxy-reports of educational attainment. Hypothesis H3 is tested by comparing rates of underestimation and overestimation between parents and children, employing the same method as for Hypothesis H1. By way of information, we also report Cohen’s (1960) κ values for the agreement between children’s and parent’s proxy-reports at this point. Cohen’s κ measures inter-rater agreement beyond chance. Its value range from 0 to 1, 0 indicating no agreement beyond chance, 1 indicating perfect agreement. We report this measure to allow the reader to compare our estimates to previous findings summarized in Table 1.

In a fourth step, to adjudicate between the competing Hypotheses H4a and H4b, we fit a series of logistic regression models with a variable indicating over- or underestimation by proxy-reporters as the outcome. The key independent variables are the actual educational attainment of the target on whom the proxy is reporting, and the difference in educational attainment between the proxy-reporter and the target. Again, we derive predicted probabilities and test for differences using the abovementioned method.

In a final step, we assess the consequences of our findings on estimates of intergenerational mobility. First, we calculate the intergenerational educational regression coefficient by regressing children’s education in years on parent’s education in years. The resulting coefficient is a measure of educational immobility: the greater the coefficient, the larger the importance of parent’s education for children’s education. For our sample of parent–child dyads, we calculate the intergenerational regression coefficient in several variants, comparing the coefficient obtained from all-self-reported education to proxy-reported variants. We then proceed to extrapolate quasi-self-reported educational distributions based on our analysis for the full TeO2 sample and calculate the intergenerational education regression coefficient for this sample. We conclude the analyses by illustrating our key finding in the context of the large-scale Global Database on Intergenerational Mobility (GDIM, [van der Weide *et al.*, 2024](#)).

In regression models, we cluster standard errors at the parent level, as some parents report information on more than one child. In the Figures, confidence intervals are displayed at the 90 per cent level due to the small sample size.

We provide a replication package with all code necessary for data preparation and analysis to reproduce all results shown at https://osf.io/nvp96/?view_only=b38b885f3b974f199991ae2dcd788a24.

Results

Nonresponse: Children are worse proxy-reporters than parents

Panel A of Figure 1 confirms our Hypothesis H1 that children have a higher likelihood of nonresponse than their parents. Parents have a very low nonresponse rate of 1.4 per cent when proxy-reporting their children’s education. By contrast, 13.5 per cent of children do not report their parents’ education. This means that typical studies leveraging children’s reports of education may lose more than ten per cent of their effective sample size. This effect does not appear to be due to the missingness of the ‘true’ information, as in all cases where a proxy-reporter fails to report the education of the second dyad member, the information is self-reported by them (see Table S9 in the Supplementary Materials). In general, questions about educational attainment generate low nonresponse rates in self-reports, as every participant reported their own education.

Panel B of Figure 1 reveals some support for Hypothesis H2, which posited that better-educated respondents are more likely to provide proxy-reports. Predicted probabilities from a logistic model that adjusted non-response for child and parent characteristics show that non-response rates are lower for children with educational attainment above secondary level (eight per cent), which is different from the secondary-educated children (17 per cent). There are relatively few individuals who report having a level of education below high school degree, which could explain the insignificant difference between this level and people with higher than secondary education. Qualitatively, the probability to report ‘I don’t know’ is very similar between people with education below or at high-school level (upper secondary). When pooling both these levels together, the comparison becomes statistically significant

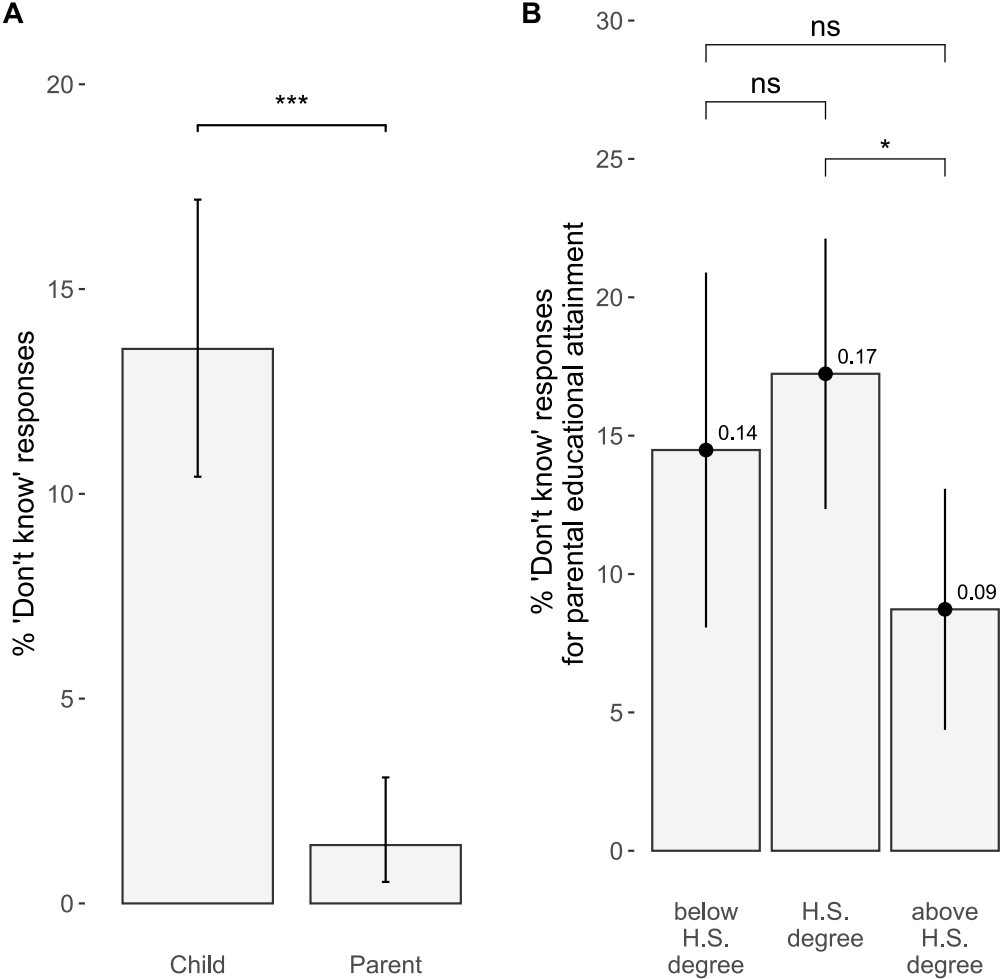


Figure 1: *Panel A:* Percentage of ‘Don’t know’ answer among parents and children as proxy-reporters *Panel B:* Probability to select ‘Don’t know’ when proxy-reporting parent education, by educational level.

Notes: $N_{\text{parents}} = 420$, $N_{\text{children}} = 421$. Error bars denote 90% confidence intervals. Panel A: Chi-square tests with Rao and Scott’s (1984) second-order correction. Panel B: Predicted probabilities based on a logistic regression model also controlling for age, sex, education and migration background of both child and parent, plus co-residence status and if the parent holds a foreign diploma. Standard errors are clustered at the family level. The complete regression model is shown in the Supplementary Materials, Table S3.

Source: TeO2 (Beauchemin *et al.*, 2023) and authors’ calculations.

again (see Table S3 in the Supplementary Materials). Overall, we thus provide some support for H2.

Over- and underestimation of education by proxy-reporters: Children are more likely to overestimate their parents' education

Panel A of Figure 2 provides support for Hypothesis H3, which stipulated that children are more likely to overestimate their parents' education. Indeed, children are more likely than parents to overestimate the education of their targets ($p < .05$). 22% of children overestimate their parent's education, while only 15% of parents do so. When it comes to levels of disagreement or underestimation, the observed differences appear small (in the area of one to two percentage points) between children and parent proxy-reporters, and are not different from zero at conventional levels of statistical significance.

Panels B and C of Figure 2 provide overwhelming support for Hypothesis H4a, which posited that parents and children both tend to reduce the differences in education between them and the targets (and conversely no support for the competing Hypothesis H4b). Predictions from models shown in Panels B and C account for the target's education and display the impact of the z -standardized difference in years of education between targets and respondents. Panel B shows the results for children's education when parents act as proxy-reporters. When a child is more educated than their parent, parents are more likely to underestimate the child's education. On average, children are three years more educated than their parents and are underestimated with a probability of 14%. When they become relatively more educated (plus one standard deviation in the educational distance), children are underestimated 17% of the time, an increase of three percentage points. Further, parents are more likely to overestimate their child's education when the distance between them becomes smaller or reverses (i.e. the parent becomes more educated than the child). The probability to overestimate one's child education goes from 13% to 25% for a decrease of one standard deviation in educational distance. Panel C shows the reverse pattern for parents' education: the more

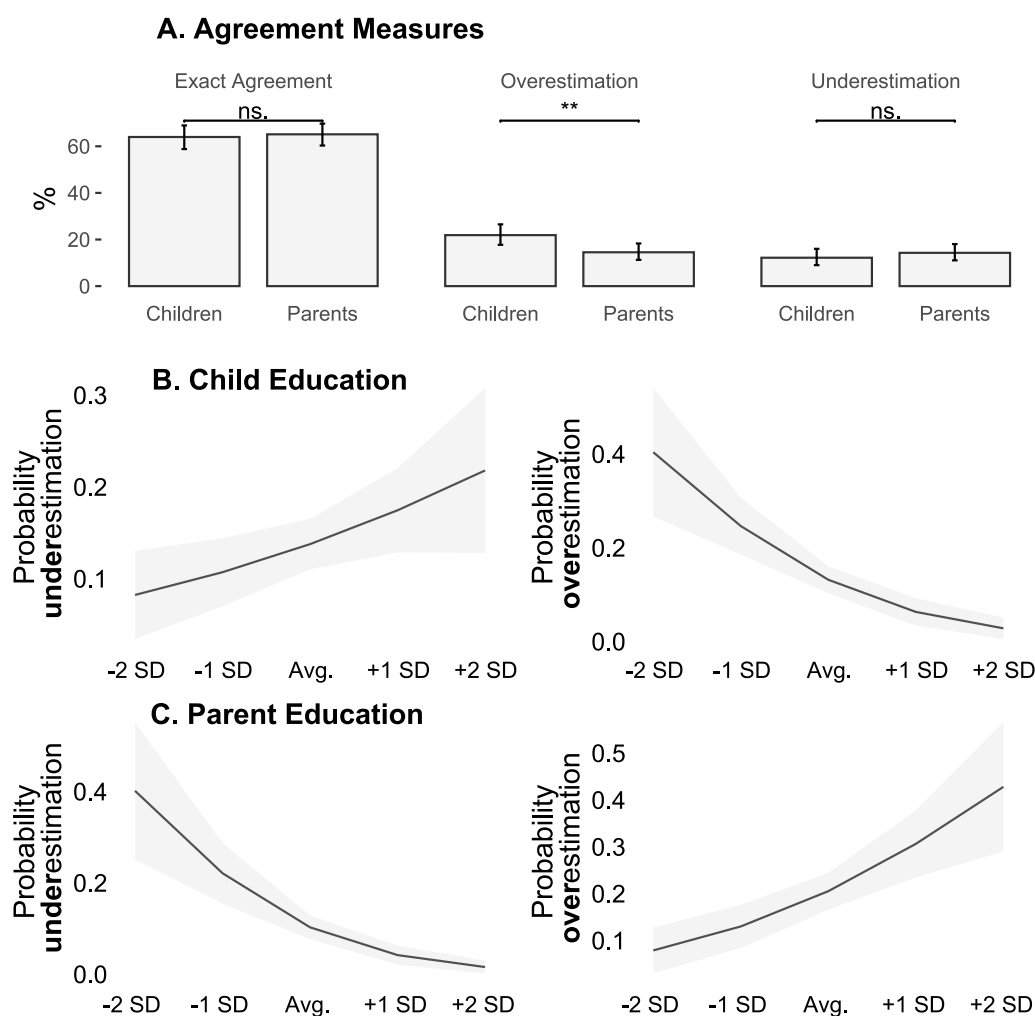


Figure 2: *Panel A*: Percentage agreement, underestimation and overestimation in proxy-reporters' answers. *Panel B*: Predicted probability to over- and under-estimate child education by educational distance. *Panel C*: Predicted probability to over- and under-estimate parent education by educational distance.

Notes: $N_{\text{parents}} = 410$, $N_{\text{children}} = 360$. *Panel A*: Chi-square tests with Rao and Scott's (1984) second-order correction. Panels B and C: Predicted probabilities based on a logistic regression model also controlling for age, sex, education and migration background of both child and parent, adding coresidence status, if the parent holds a foreign diploma, if the child is a student and time between parent and child interviews. We add the education of the proxy-reporter (in Panel B, child's education, in Panel C, parent's education) 90% confidence intervals based on standard errors clustered at the family level. The complete regression models are shown in the Supplementary Materials, Tables S4 and S5.

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

they are educated compared to their child, the more likely their education is to be underestimated (from 10% to 22%, for a one standard deviation decrease in the distance), but the less they are educated, the more likely it is overestimated (from 20% to 31%, for a one standard deviation increase in the distance).

Misreporting inflates educational reproduction estimates, but not by much

Figure 3 explores the consequences of proxy-reporting of educational attainment for educational mobility. Panel A shows the intergenerational education regression coefficient, resulting from regressing the number of years of education of children on the years of education of parents in several perspectives. First, in a classical perspective where children serve as proxy-reporters for their parents (i.e. retrospectively when seen from a generational standpoint); second, where parents serve as proxy-reporters for their children (i.e. prospectively); and finally, when we base our estimate exclusively on self-reports from both generations. In the sample of dyads from the TeO2 survey, we find a coefficient of .15 for the retrospective measure, i.e. every additional year of parental education goes along with additional two months of education for the child. For the prospective measure, the coefficient is .17. Both of these are higher than the .12 coefficient only based on self-reports, yet the difference between the coefficients is statistically significance at conventional levels.

In order to gauge the extent to which this lack of statistical significance is due to small sample size, we extrapolated quasi-self-reported educational distributions based on our analysis for the full TeO2 sample and calculated the intergenerational education regression coefficient for the full sample (full details on the simulation procedure are available in the Supplementary Materials). Using this method, we confirm that proxy-reports tend to underestimate educational mobility, as the proxy-report based coefficients are larger than the quasi-self-reported regression coefficient. Our simulated self-reports yield different intergenerational regression coefficients because they rely on different populations: children reporting on their parents (retrospective) are younger

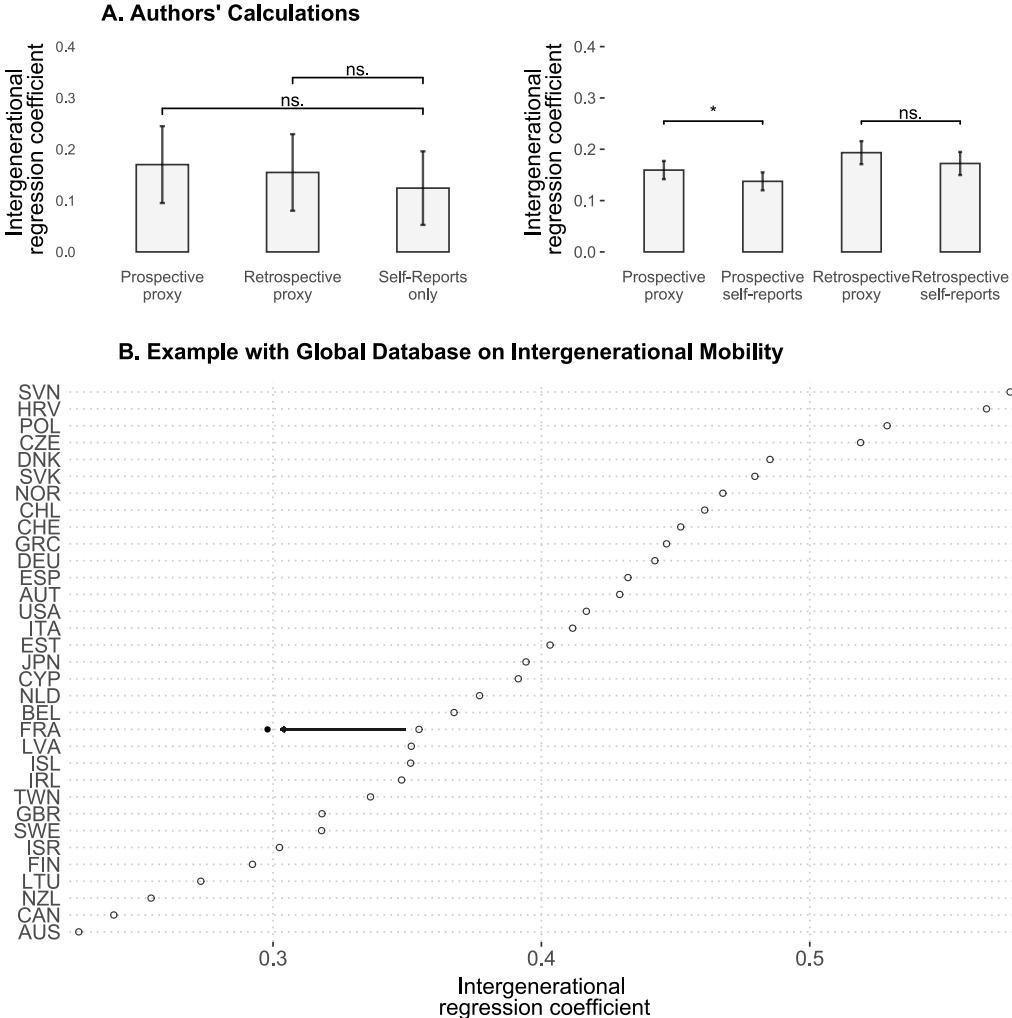


Figure 3: *Panel A*: Educational mobility is higher when using only self-reported information. Bar chart of intergenerational regression coefficients observed via children’s proxy-reports (retrospective) and parent’s proxy-reports (prospective perspective) as well as the intergenerational regression coefficient based on self-reports only, for the dyads-only sample ($N_{prospective} = 363$, $N_{retrospective} = 382$) as well as predicted for the full TeO2 sample ($N_{prospective} = 8,400$, $N_{retrospective} = 31,387$). *Panel B*: Consequence of proxy-reports of education for France in international comparison. Black arrow indicates the extent of the shift for the France. Dot plot of the intergenerational regression coefficient for the 1970s birth cohort of children across selected high-income countries.

Notes: *Panel A*: Two-sample z -test, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The self-report only coefficients in the full TeO2 sample are simulated by predicting errors that individuals in the full TeO2 survey make when reporting education for their parent/child and is thus different in both perspectives. Methodological details and full results reported in the Supplementary Materials. *Panel B*: Distribution of intergenerational regression coefficients based on GDIM data. We apply a 14% decrease to the French estimate, reflecting the estimated bias in proxy-reports *Sources*: TeO2 (Beauchemin *et al.*, 2023), GDIM (van der Weide *et al.*, 2024), and authors’ calculations.

than parents reporting about their children (prospective). Intergenerational mobility measured later in the life course tends to be higher, which we see on our simulated coefficients. For both the prospective and retrospective perspectives, we observe the same qualitative results: proxy-reports yield higher intergenerational regression coefficients than (simulated) self-reports. This difference is statistically significant for the classical retrospective perspective only, i.e. when children act as proxy-reporters, the association between generations is overestimated. When parents act as proxy-reporters, in the alternative prospective perspective, the difference is not statistically significant. In our simulations, we find that the intergenerational coefficients moves from .14 when self-reports are used, to .16 when children act as proxy-reporter. This is similar in magnitude to the 16% increase in the dyads-only sample.

Panel B of Figure 3 illustrates our finding using the Global Database on Intergenerational Mobility (GDIM, [van der Weide *et al.*, 2024](#)) as a backdrop. Proxy-reports in our sample lead to an overestimation of intergenerational mobility of 16%. We thus inflate the French coefficient by 16% to illustrate our findings. We see that there is remarkable variability among estimates, even within the group of ‘High-income countries’ that we chose to plot our results against³. Considering this, the change in the intergenerational mobility estimates appears limited. The magnitude of this change should also be compared to error-adjusted estimates for all other countries, as the GDIM data relies exclusively on retrospective reports.

Robustness checks

We reproduce our analyses on non-standardized variables of education (using the number of years needed to obtain a degree) to check that our findings do not depend on the relative education within child and parent samples, and find similar results (see Figure S1). We also estimated multinomial regression models when studying agreement, overestimation and underestimation, and find similar results (see Tables S10 and S11), and thus report binary logistic models to ease interpretation. Lastly, we provide additional analyses that distinguish between different migration backgrounds (see Figure S2). We find

³See [van der Weide *et al.* \(2024, p. 12\)](#) for more information on the groupings.

no significant differences between subgroups. However, our analysis relies on small groups, a conclusive test of differences would require a larger sample.

Discussion and conclusion

Summary of findings and implications

In this study, we set out to analyze the validity of proxy-reports on educational attainment within family dyads, shedding light on an often overlooked aspect of intergenerational mobility research. Our investigation provided a comprehensive examination of the quality of self- and proxy-reports, examining biases in both children and parents reporting on each other.

Our first contribution lies in the assessment of biases and nonresponse in proxy-reports from both children and parents, an analytical approach made possible by a unique feature of the ‘Trajectoires et Origines 2’ (Beauchemin *et al.*, 2023) survey. Beyond the traditional focus on children reporting on their parents, we studied the little-explored issue of parents reporting on their adult children—an increasingly relevant dimension in contemporary aging surveys. Most of our hypotheses were confirmed. Regarding non-response, we found clear evidence that children know less about their parents’ education than the other way around, even when they are adults, and still live with their parents. We found some evidence that children with lower education are more likely to answer ‘I don’t know’ to questions regarding their parents’ educational attainment, a concerning pattern when studying intergenerational mobility.

Importantly, our findings consistently confirm that proxy-reports tend to minimize the educational distance between themselves and the target of information, whether it involves overestimation when the reporter is more educated than the target or underestimation when the reporter is less educated than the target. This pattern challenges conventional wisdom and the results of Broom *et al.* (1978): one would expect parents to overestimate their children more than the other way around. We see our finding that any proxy-reporter tends to reduce the distance between themselves and the target as

additional evidence for ‘anchoring strategies’ also documented by [Wagmiller \(2009\)](#). Respondents might (uncsciously) use their own education more than they should when trying to retrieve their parents/children education, thus inflating intergenerational reproduction. Another potential explanation is a ‘dissonance-reducing’ one: both parents and children overestimate their similarity, placing themselves in a larger intergenerational picture where origin and destination shape their understanding of each other’s social position ([Irwin, 2015](#); [Friedman *et al.*, 2021](#)).

Our second main contribution is to assess the consequences of these biases on measures of intergenerational educational mobility. Based on our sample, our findings suggest a consistent overestimation of the link between educational origin and destination when proxy-reports are employed, which is the case in most mainstream mobility studies. This bias is not negligible, with intergenerational regression coefficients being 16% higher when a proxy-report is included.

Our research brings to the fore the importance of significant others and social mobility to answer survey questions, even when these pertain to salient characteristics such as education. Both one’s own characteristics and others’ can be reinterpreted in the light of life experiences, and differently reported depending on the link between the reporter and the target of the information. [Caron *et al.* \(2023\)](#) investigated the impact of demographic events on self-description as ‘born’ French rather than ‘naturalized’ or ‘became French’ among individuals who were not French by birth. They find that intermarriage, naturalization, and residential mobility trigger reclassification into the ‘born French’ category. Our study shows that interindividual differences can lead to similar systematic reclassification patterns among pairs of individuals. By collecting data from the parents’ and children’s sides, we were able to assess that it is indeed the distance between educational levels rather than absolute education that matters in reclassifying someone as more or less educated than they are. We thus confirmed the intuition of [Broom *et al.* \(1978\)](#) to some extent: the evaluation of the target depends on one’s own experience. However, we find that both parents and children underestimate the distance between them, unlike Broom and coauthors who uncovered opposing trends

among parents (prone to underestimating the distance) and children (prone to overestimating the distance).

Our findings also resonate with insights from social psychology, where overestimation of similarity among close peers is well-established on several outcomes (e.g. political value (Collisson and Howell, 2014), personality traits (Collisson and Howell, 2014; Murray *et al.*, 2002; Kouros and Papp, 2019) or job satisfaction (Kenny and Acitelli, 2001)). We show that similar processes may be at play within families, even on more ‘objective’ traits than beliefs or attitudes.

Limitations

Several aspects of our data lead us to expect that our estimates of the potential biases in children’s (and parents’) proxy-reports are conservative. All child–parent dyads in our sample are in contact with one another: the parents know how their children can be reached, or parents and children live together. Most children, although adults, are relatively young, and information on their education is recent. Similar results in a broader and older population may bring about stronger doubts of proxy-report validity.

Our sample was limited in size, preventing investigation of effects heterogeneity. Some proxy-reporters likely have only limited access to information about their parents’ education. Most of the parents in our sample are non-European immigrants or immigrants’ descendants. Immigrants from these countries are, on average, more likely to not hold any degree than natives (INSEE, 2023) and may have more trouble getting their degrees recognized (Li and Lu, 2023; Weber *et al.*, 2024). However, our analyses that distinguish between different migration backgrounds show no statistically significant differences.

Our data rely on self- and proxy-report differences rather than on differences between self-reports and register data. Attewell and Domina (2011) criticize this approach, arguing that self-reports may be as invalid as proxy-reports. We believe that this bias is mitigated in the TeO2 data considering the procedure to collect individuals’ education information. The survey has very low levels of non-response on respondents’ own education, suggesting that no bias arises from excluding individuals who do not report information.

Additionally, respondents' own education is computed by using information drawn from multiple questions, tailored to accommodate both individuals educated in France or abroad, with specific questions about degree attainment. All in all, we believe that our analyses provide a conservative estimate of bias in proxy-reports and intergenerational social mobility.

Outlook

While our study offers valuable insights into the validity of self- and proxy-reports on educational attainment within family dyads, we acknowledge the limitations of our study, which, in turn, chart pathways for future research. Future research should try to expand the scope by using larger samples when available, to assess the external validity of our conclusions. Further, our focus on educational attainment leaves other dimensions of socioeconomic status unexplored. Subsequent research should extend the investigation to encompass other crucial dimensions, such as occupation, to offer a more comprehensive understanding of the complexities in measuring intergenerational mobility.

Our research moves the field forward by not only identifying biases in the quality of proxy-reports of educational attainment but also highlighting their potential impact on social mobility estimates. These insights could contribute to methodological advancements in survey research but also carry implications for the broader field of intergenerational social mobility studies. By offering a clearer perspective on the complexities of proxy-reporting of educational attainment, we hope to contribute valuable insights that resonate beyond the scope of our specific study, inviting social mobility researchers to critically engage with the quality of the data underpinning their analyses.

References

Anderson, Lewis R., Patrick Präg, Evelina T. Akimova, and Christiaan Monden, 2024. 'The Total Effect of Social Origins on Educational Attainment: Meta-analysis of Sibling Correlations From 18 Countries.' *Demography* 61(5): 1637–1666. doi: [10.1215/00703370-11579806](https://doi.org/10.1215/00703370-11579806).

- Anderson, Lewis R., Paula Sheppard, and Christiaan Monden, 2018. ‘Grandparent Effects on Educational Outcomes. A Systematic Review.’ *Sociological Science* 5(6): 114–142. doi: [10.15195/v5.a6](https://doi.org/10.15195/v5.a6).
- Araki, Satoshi and Takehiko Kariya, 2022. ‘Credential Inflation and Decredentialization. Re-Examining the Mechanism of the Devaluation of Degrees.’ *European Sociological Review* 38(6): 904–919. doi: [10.1093/esr/jcac004](https://doi.org/10.1093/esr/jcac004).
- Arel-Bundock, Vincent, Noah Greifer, and Andrew Heiss, forthcoming. ‘How to Interpret Statistical Models Using ‘marginaleffects’ in R and Python.’ *Journal of Statistical Software* .
- Attewell, Paul and Thurston Domina, 2011. ‘Educational Imposters and Fake Degrees.’ *Research in Social Stratification and Mobility* 29(1): 57–69. doi: [10.1016/j.rssm.2010.12.004](https://doi.org/10.1016/j.rssm.2010.12.004).
- Beauchemin, Cris, Mathieu Ichou, Patrick Simon, and the TeO2 survey design team, 2023. ‘Trajectories and Origins 2 (2019–20). A Survey on Population Diversity in France.’ *Population* 78E(1): 11–28. doi: [10.3917/popu.2301.0011](https://doi.org/10.3917/popu.2301.0011).
- Breen, Richard and Walter Müller, 2020. *Education and Intergenerational Social Mobility in Europe and the United States*. Stanford, CA: Stanford University Press. doi: [10.11126/stanford/9781503610163.001.0001](https://doi.org/10.11126/stanford/9781503610163.001.0001).
- Broom, Leonard, F. Lancaster Jones, Patrick McDonnell, and Paul Duncan-Jones, 1978. ‘Is It True What They Say About Daddy?’ *American Journal of Sociology* 84(2): 417–426. doi: [10.1086/226791](https://doi.org/10.1086/226791).
- Caron, Louise, Haley McAvay, and Mirna Safi, 2023. ‘Born Again French. Explaining Inconsistency in Citizenship Declarations in French Longitudinal Data.’ *American Sociological Review* 88(6): 1066–1103. doi: [10.1177/00031224231207392](https://doi.org/10.1177/00031224231207392).
- Cohen, Jacob, 1960. ‘A Coefficient of Agreement for Nominal Scales.’ *Educational and Psychological Measurement* 20(1): 37–46. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- Collins, Randall, 2019. *The Credential Society. An Historical Sociology of*

- Education and Stratification*. New York: Columbia University Press, legacy edition. doi: [10.7312/coll19234](https://doi.org/10.7312/coll19234).
- Collisson, Brian and Jennifer L. Howell, 2014. ‘The Liking-Similarity Effect: Perceptions of Similarity as a Function of Liking.’ *The Journal of Social Psychology* 154(5): 384–400. doi: [10.1080/00224545.2014.914882](https://doi.org/10.1080/00224545.2014.914882).
- Crosnoe, Robert, 2001. ‘Parental Involvement in Education: The Influence of School and Neighborhood.’ *Sociological Focus* 34(4): 417–434. doi: [10.1080/00380237.2001.10571211](https://doi.org/10.1080/00380237.2001.10571211).
- de Vries, Jannes, 2006. *Measurement Error in Family Background Variables. The Bias in the Intergenerational Transmission of Status, Cultural Consumption, Party Preference, and Religiosity*. ICS/Radboud University.
- de Vries, Jannes and Paul M. de Graaf, 2008. ‘The Reliability of Family Background Effects on Status Attainment. Multiple Informant Models.’ *Quality & Quantity* 42(2): 203–234. doi: [10.1007/s11135-006-9042-6](https://doi.org/10.1007/s11135-006-9042-6).
- Engzell, Per and Jan O. Jonsson, 2015. ‘Estimating Social and Ethnic Inequality in School Surveys. Biases from Child Misreporting and Parent Nonresponse.’ *European Sociological Review* 31(3): 312–325. doi: [10.1093/esr/jcv005](https://doi.org/10.1093/esr/jcv005).
- Ensminger, Margaret E., Christopher B. Forrest, Anne W. Riley, Myungsa Kang, Bert F. Green, Barbara Starfield, and Sheryl A. Ryan, 2000. ‘The Validity of Measures of Socioeconomic Status of Adolescents.’ *Journal of Adolescent Research* 15(3): 392–419. doi: [10.1177/0743558400153005](https://doi.org/10.1177/0743558400153005).
- Erikson, Robert and Jan O. Jonsson, 1996. *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, CO: Westview.
- Fingerman, Karen L., Meng Huo, and Kira S. Birditt, 2020. ‘Mothers, Fathers, Daughters, and Sons. Gender Differences in Adults’ Intergenerational Ties.’ *Journal of Family Issues* 41(9): 1597–1625. doi: [10.1177/0192513X19894369](https://doi.org/10.1177/0192513X19894369).
- Forster, Andrea G. and Herman G. van de Werfhorst, 2019. ‘Navigating Institutions. Parents’ Knowledge of the Educational System and Students’

- Success in Education.’ *European Sociological Review* 36(1): 48–64. doi: [10.1093/esr/jcz049](https://doi.org/10.1093/esr/jcz049).
- Friedman, Sam, Dave O’Brien, and Ian McDonald, 2021. ‘Deflecting Privilege. Class Identity and the Intergenerational Self.’ *Sociology* 55(4): 716–733. doi: [10.1177/0038038520982225](https://doi.org/10.1177/0038038520982225).
- Gugushvili, Alexi, 2021. ‘Why Do People Perceive Themselves as Being Downwardly or Upwardly Mobile?’ *Acta Sociologica* 64(1): 3–23. doi: [10.1177/0001699320929742](https://doi.org/10.1177/0001699320929742).
- Houseworth, Christina and Jonathan Fisher, 2020. ‘Measurement Error in Occupation and the Impact on Intergenerational Mobility.’ *Research in Social Stratification and Mobility* 70(100469): 1–8. doi: [10.1016/j.rssm.2019.100469](https://doi.org/10.1016/j.rssm.2019.100469).
- Hout, Michael and Orestes P. Hastings, 2016. ‘Reliability of the Core Items in the General Social Survey. Estimates from the Three-Wave Panels, 2006–14.’ *Sociological Science* 3(43): 971–1002. doi: [10.15195/v3.a43](https://doi.org/10.15195/v3.a43).
- Hovestadt, Till and Thorsten Schneider, 2021. ‘Liefern Jugendliche valide Informationen zum Bildungsstand ihrer Eltern in standardisierten Erhebungen? Befunde zu Schülerinnen und Schülern der 9. Jahrgangsstufe in Deutschland [Can adolescents provide valid information on parents’ educational attainment in standardized surveys? Results from a survey conducted in grade 9 in Germany].’ *Zeitschrift für Erziehungswissenschaft* 24(3): 715–742. doi: [10.1007/s11618-021-01016-5](https://doi.org/10.1007/s11618-021-01016-5).
- Ichou, Mathieu and Louis-André Vallet, 2013. ‘Academic Achievement, Tracking Decisions, and Their Relative Contribution to Educational Inequalities: Change over Four Decades in France.’ In Jackson, Michelle, ed., *Determined to Succeed?: Performance versus Choice in Educational Attainment*, pp. 116–148. Stanford University Press. doi: [10.2307/j.ctvqsdrrj](https://doi.org/10.2307/j.ctvqsdrrj).
- INSEE, 2023. *Immigrés et descendants d’immigrés en France*. Paris: Institut national de la statistique et des études économiques. IMM23.
- Irwin, Sarah, 2015. ‘Class and Comparison. Subjective Social Location and

- Lay Experiences of Constraint and Mobility.’ *British Journal of Sociology* 66(2): 259–281. doi: [10.1111/1468-4446.12121](https://doi.org/10.1111/1468-4446.12121).
- Jaspers, Eva, Marcel Lubbers, and Jannes de Vries, 2008. ‘Parents, Children, and the Distance between Them. Long-Term Socialization Effects in the Netherlands.’ *Journal of Comparative Family Studies* 39(1): 39–58. doi: [10.3138/jcfs.39.1.39](https://doi.org/10.3138/jcfs.39.1.39).
- Jerrim, John and John Micklewright, 2014. ‘Socio-economic Gradients in Children’s Cognitive Skills. Are Cross-Country Comparisons Robust to Who Reports Family Background?’ *European Sociological Review* 30(6): 766–781. doi: [10.1093/esr/jcu072](https://doi.org/10.1093/esr/jcu072).
- Kayser, Brian D. and Gene F. Summers, 1973. ‘The Adequacy of Student Reports of Parental SES Characteristics.’ *Sociological Methods & Research* 1(3): 303–315. doi: [10.1177/004912417300100302](https://doi.org/10.1177/004912417300100302).
- Kenny, David A. and Linda K. Acitelli, 2001. ‘Accuracy and Bias in the Perception of the Partner in a Close Relationship.’ *Journal of Personality and Social Psychology* 80(3): 439–448. doi: [10.1037/0022-3514.80.3.439](https://doi.org/10.1037/0022-3514.80.3.439).
- Kouros, Chrystyna D. and Lauren M. Papp, 2019. ‘Couples’ Perceptions of Each Other’s Daily Affect: Empathic Accuracy, Assumed Similarity, and Indirect Accuracy.’ *Family Process* 58(1): 179–196. doi: [10.1111/famp.12344](https://doi.org/10.1111/famp.12344).
- Kreuter, Frauke, Stephanie Eckmann, Kai Maaz, and Rainer Watermann, 2010. ‘Children’s Reports of Parents’ Education Level. Does It Matter Whom You Ask and What You Ask About?’ *Survey Research Methods* 4(3): 127–138. doi: [10.18148/srm/2010.v4i3.4283](https://doi.org/10.18148/srm/2010.v4i3.4283).
- Lee, Jinkook, Drystan Phillips, Jenny Wilkens, and Gateway to Global Aging Data Team, 2021. ‘Gateway to Global Aging Data. Resources for Cross-National Comparisons of Family, Social Environment, and Healthy Aging.’ *Journals of Gerontology B* 76(S1): S5–S16. doi: [10.1093/geronb/gbab050](https://doi.org/10.1093/geronb/gbab050).
- Lee, Jungmin and Sokbae Lee, 2012. ‘Does it Matter *Who* Responded to the Survey? Trends in the US Gender Earnings Gap Revisited.’ *ILR Review* 65(1): 148–160. doi: [10.1177/001979391206500108](https://doi.org/10.1177/001979391206500108).

- Li, Xiaoguang and Yao Lu, 2023. ‘Education–Occupation Mismatch and Nativity Inequality Among Highly Educated US Workers.’ *Demography* 60(1): 201–226. doi: [10.1215/00703370-10404849](https://doi.org/10.1215/00703370-10404849).
- Lien, Nanna, Christine Friestad, and Knut-Inge Klepp, 2001. ‘Adolescents’ Proxy Reports of Parents’ Socioeconomic Status. How Valid Are They?’ *Journal of Epidemiology & Community Health* 55(10): 731–737. doi: [10.1136/jech.55.10.731](https://doi.org/10.1136/jech.55.10.731).
- Looker, E. Dianne, 1989. ‘Accuracy of Proxy Reports of Parental Status Characteristics.’ *Sociology of Education* 62(4): 257–276. doi: [10.2307/2112830](https://doi.org/10.2307/2112830).
- Massagli, Michael P. and Robert M. Hauser, 1983. ‘Response Variability in Self- and Proxy Reports of Paternal and Filial Socioeconomic Characteristics.’ *American Journal of Sociology* 89(2): 420–431. doi: [10.1086/227873](https://doi.org/10.1086/227873).
- Merllié, Dominique, 2008. ‘How to Compare “Subjective” and “Objective” Mobility.’ *Sociologie du Travail* 50: e50–e65. doi: [10.1016/j.soctra.2008.07.004](https://doi.org/10.1016/j.soctra.2008.07.004).
- Murray, Sandra L., John G. Holmes, Gina Bellavia, Dale W. Griffin, and Dan Dolderman, 2002. ‘Kindred Spirits? The Benefits of Egocentrism in Close Relationships.’ *Journal of Personality and Social Psychology* 82(4): 563–581. doi: [10.1037/0022-3514.82.4.563](https://doi.org/10.1037/0022-3514.82.4.563).
- Naudet, Jules and Renuka George, 2018. *Stepping into the Elite: Trajectories of Social Achievement in India, France, and the United States*. New Delhi, India: Oxford University Press.
- Rao, Jon N. K. and Alastair Scott, 1984. ‘On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data.’ *Annals of Statistics* 12: 46–60. doi: [10.1214/aos/1176346391](https://doi.org/10.1214/aos/1176346391).
- Reynolds, Jeremy and Jeffrey B. Wenger, 2012. ‘He Said, She Said. The Gender Wage Gap According to Self and Proxy Reports in the Current Population Survey.’ *Social Science Research* 41(2): 392–411. doi: [10.1016/j.ssresearch.2011.10.005](https://doi.org/10.1016/j.ssresearch.2011.10.005).
- Ridolfo, Heather and Aaron Maitland, 2011. ‘Factors That Influence the Accuracy of Adolescent Proxy Reporting of Parental Characteris-

- tics. A Research Note.’ *Journal of Adolescence* 34(1): 95–103. doi: [10.1016/j.adolescence.2010.01.008](https://doi.org/10.1016/j.adolescence.2010.01.008).
- Skopek, Jan and Thomas Leopold, 2020. ‘Educational Reproduction in Germany. A Prospective Study Based on Retrospective Data.’ *Demography* 57(4): 1241–1270. doi: [10.1007/s13524-020-00896-2](https://doi.org/10.1007/s13524-020-00896-2).
- Smith, Jane Graves, 2006. ‘Parental Involvement in Education among Low-Income Families. A Case Study.’ *School Community Journal* 16(1): 43–56.
- Sonnega, Amanda, Jessica D. Faul, Mary Beth Ofstedal, Kenneth M. Langa, John W. R. Phillips, and David R. Weir, 2014. ‘Cohort Profile. The Health and Retirement Study (HRS).’ *International Journal of Epidemiology* 43(2): 576–585. doi: [10.1093/ije/dyu067](https://doi.org/10.1093/ije/dyu067).
- Tamborini, Christopher R. and ChangHwan Kim, 2013. ‘Are Proxy Interviews Associated With Biased Earnings Reports? Marital Status and Gender Effects of Proxy.’ *Social Science Research* 42(2): 499–512. doi: [10.1016/j.ssresearch.2012.11.004](https://doi.org/10.1016/j.ssresearch.2012.11.004).
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski, 2000. *The Psychology of Survey Response*. Cambridge University Press. doi: [10.1017/CBO9780511819322](https://doi.org/10.1017/CBO9780511819322).
- van der Weide, Roy, Christoph Lakner, Daniel Gerszon Mahler, Ambar Narayan, and Rakesh Gupta, 2024. ‘Intergenerational Mobility Around the World. A New Database.’ *Journal of Development Economics* 166(103167): 1–32. doi: [10.1016/j.jdeveco.2023.103167](https://doi.org/10.1016/j.jdeveco.2023.103167).
- Vereecken, Carine Anna and Ann Vandeghechuchte, 2003. ‘Measurement of Parental Occupation. Agreement Between Parents and Their Children.’ *Archives of Public Health* 61: 141–149.
- Wagmiller, Jr, Robert L., 2009. ‘A Fixed Effects Approach to Assessing Bias in Proxy Reports.’ *International Journal of Public Opinion Research* 21(4): 477–505. doi: [10.1093/ijpor/edp035](https://doi.org/10.1093/ijpor/edp035).
- Weber, Rosa, Mathieu Ferry, and Mathieu Ichou, 2024. ‘Which Degree for

- Which Occupation? Vertical and Horizontal Mismatch Among Immigrants, Their Children, and Grandchildren in France.' *Demography* (Forthcoming).
- West, Anne, Philip Noden, Ann Edge, and Miriam David, 1998. 'Parental Involvement in Education in and out of School.' *British Educational Research Journal* 24(4): 461–484. doi: [10.1080/0141192980240407](https://doi.org/10.1080/0141192980240407).
- West, Patrick, Helen Sweeting, and Ewen Speed, 2001. 'We Really Do Know What You Do. A Comparison of Reports from 11-Year Olds and their Parents in Respect of Parental Economic Activity and Occupation.' *Sociology* 35(2): 539–559. doi: [10.1177/S0038038501000268](https://doi.org/10.1177/S0038038501000268).
- Wittrock, Jill, Linda Kimmel, Brian Hunscher, and Kien Trung Le, 2017. 'Proxy Reporting in Education Surveys. Factors Influencing Accurate Reporting in the 2012 Qatar Education Study.' *International Journal of Social Research Methodology* 20(6): 737–748. doi: [10.1080/13645579.2017.1301078](https://doi.org/10.1080/13645579.2017.1301078).
- Zhang, Dalun, Hsien-Yuan Hsu, Oi-man Kwok, Michael Benz, and Lisa Bowman-Perrott, 2011. 'The Impact of Basic-Level Parent Engagements on Student Achievement. Patterns Associated with Race/Ethnicity and Socioeconomic Status.' *Journal of Disability Policy Studies* 22(1): 28–39. doi: [10.1177/1044207310394447](https://doi.org/10.1177/1044207310394447).

Supplementary Materials

Contents

Additional information	38
Table S1: Coding of the educational variables	39
Table S2: Dyad-level characteristics	42
Table S3: Regression model from Figure 1	43
Table S4: Regression models from Figure 2, panels B	44
Table S5: Regression models from Figure 2, panels C.	46
Table S6: Regression models from Figure 3, Panel A - Dyads-only .	47
Table S7: Regression models from Figure 3, Panel A - Simulations (Part 1)	47
Table S8: Regression models from Figure 3, Panel A - Simulations (Part 2)	48
Description of the simulation procedure to obtain Figure 3	48
Robustness checks	49
Table S9: Availability of Information: Are Missing Proxy-Reports due to ‘Don’t know’ Self-Reports?	49
Table S10: Multinomial models: Predicting Agreement, Underesti- mation and Overestimation in Child Education	49
Table S11: Multinomial models: Predicting Agreement, Underesti- mation and Overestimation in Parent Education	51
Figure S1: Probability to over- or underestimate education, non- standardized educational distance	53
Figure S2: Probability to over- or underestimate education stratified by migrational status of child	54

Additional information

Table S1: Coding of the educational variables

Variable	Self-report	Proxy-report
Parental Education	diploder	t_diplom or t_diplom
	Multiple questions asked	Quel est le plus haut diplôme détenu par votre [père/mère] ?
	0. Non concernés (Etudiants + ceux qui ne sont pas allés à l'école)	1. aucun diplôme
	11. Doctorats sauf santé	2. CEP (certificat d'études primaires) ou diplôme étranger de même niveau
	12. Doctorats de santé	3. Brevet des collèges, BEPC, brevet élémentaire ou diplôme étranger de même niveau
	13. Ecoles de commerce	4. CAP, BEP ou diplôme étranger de même niveau
	14. Ecoles d'ingénieur	5. Baccalauréat technologique ou professionnel ou diplôme étranger de même niveau
	15. DESS, masters professionnels	6. Baccalauréat général (Séries A, B, C, D, E, ES, L, S), brevet supérieur, capacité en droit, DAEU ou diplôme étranger de même niveau

Table S1: continued

Variable	Self-report	Proxy-report
	16. DEA, Magistères, masters recherche	7. Diplôme de niveau BAC +2 (DEUG, BTS ou équivalent)
	17. Master non différencié	8. Diplôme de niveau BAC +3 ou 4 (licence, maîtrise, master 1 ou équivalent)
	21. Maîtrise, MST, Miage, maîtrise IUP	9. Diplôme de niveau BAC +5 et plus (DEA, DESS, master 2, diplôme d'une grande école, doctorat)
	22. Licence	
	23. Licence professionnelle	
	24. Autres diplômes supérieurs (niveau bac +3 et plus)	
	31. DUT	
	32. BTS	
	33. Deust, DTS, DNTS, DPECF	
	34. Diplômes paramédicaux et sociaux (niveau bac +2)	
	35. Deug	
	36. Autres diplômes niveau technicien supérieur (niveau bac +2)	
	41. Capacité en droit, DAEU, ESEU	
	42. Bac général	
	43. Bac technologique	
	44. Bac professionnel	

Table S1: continued

Variable	Self-report	Proxy-report
	45. Brevet de technicien, brevet professionnel 51. BEP 52. CAP 53. Autres diplômes de niveau BEP-CAP 60. Brevet des collèges 70. Certificat d'études pri- maires 71. Aucun diplôme 99. Indéterminé	
Child Edu- cation	Same as Parental Education	e_diplo Quel est le diplôme le plus élevé de [PRÉNOM] ? 1. Aucun diplôme 2. CEP (certificat d'études primaires) ou diplôme étranger de même niveau / 3. Brevet des collèges ou diplôme étranger de même niveau 4. CAP, BEP ou diplôme étranger de même niveau 5. Baccalauréat profession- nel ou diplôme étranger de même niveau 6. Baccalauréat tech- nologique ou diplôme étranger de même niveau

Table S1: continued

Variable	Self-report	Proxy-report
		7. Baccalauréat général ou diplôme étranger de même niveau
		8. Diplôme de niveau BAC +2 (DEUG, BTS ou équivalent)
		9. Diplôme de niveau BAC +3 (licence ou équivalent)
		10. Diplôme de niveau BAC +4 (maîtrise, master 1 ou équivalent)
		11. Diplôme de niveau Bac +5 et plus (DEA, DESS, master 2, diplôme d'une grande école, doctorat)

Source: TeO2 (Beauchemin *et al.*, 2023) questionnaire.

Table S2: Dyad-level characteristics

Pair-level characteristics	N = 421
<i>Agreement on Child Edu. (% Similar)</i>	
Similar	65 (269)
(Missing)	8
<i>Agreement on Parent Edu. (% Similar)</i>	
Similar	64 (231)
(Missing)	60
<i>Over/Underestimation of Child Edu.</i>	
Agreement	71 (294)
Underestimation	14 (59)
Overestimation	15 (60)

(Missing)	8
<i>Over/Underestimation of Parent Edu.</i>	
Agreement	66 (238)
Underestimation	12 (44)
Overestimation	22 (79)
(Missing)	60
<i>Educational Distance (years), M (SD)</i>	3.0 (6.0)
(Missing)	4
<i>Time between Interviews</i>	
6 months or less	50 (211)
From 6 months to a year	42 (175)
Over a year	8 (35)
<i>Gender Composition</i>	
Father-Daughter	21 (87)
Father-Son	21 (87)
Mother-Daughter	30 (128)
Mother-Son	28 (119)

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Note: weighted percentages (N) unless otherwise specified.

Table S3: Regression model from Figure 1

	M1			M2		
	$\log(\text{OR})^{1,2}$	SE^2	p-value	$\log(\text{OR})^{1,2}$	SE^2	p-value
<i>Education Parent</i>						
below H.S. degree	—	—		—	—	
H.S. degree	0.06	0.491	0.91	0.07	0.488	0.89
above H.S. degree	-0.15	0.379	0.70	-0.13	0.376	0.73
<i>Education Child</i>						
				-0.72*	0.397	0.071
below H.S. degree	—	—				
H.S. degree	0.22	0.412	0.60			
above H.S. degree	-0.59	0.461	0.20			
<i>Migration Status Parent</i>						
Immigrant	—	—		—	—	
Child of 2 immigrants	-0.35	0.669	0.60	-0.31	0.673	0.64

Child of 1 immigrant	0.06	0.731	0.93	0.08	0.717	0.91
Other	0.47	0.738	0.53	0.46	0.668	0.49
<i>Migration Status Child</i>						
Child of 2 immigrants	—	—		—	—	
Child of 1 immigrant	0.24	0.552	0.66	0.20	0.575	0.73
Other	0.08	0.732	0.91	0.02	0.741	0.97
<i>Age Parent</i>						
Less than 48	—	—		—	—	
From 48 to 51	-0.66	0.511	0.19	-0.63	0.554	0.26
From 52 to 55	0.19	0.433	0.66	0.22	0.411	0.59
Over 56	0.50	0.475	0.29	0.51	0.482	0.29
<i>Age Child</i>						
Less than 20	—	—		—	—	
From 20 to 22	-0.39	0.420	0.35	-0.40	0.447	0.37
From 23 to 25	-0.07	0.491	0.88	-0.14	0.458	0.77
Over 26	-0.89	0.590	0.13	-0.95*	0.574	0.10
<i>Mother</i>	-0.30	0.310	0.33	-0.31	0.329	0.35
<i>Daughter</i>	-0.01	0.308	0.97	0.01	0.314	0.99
<i>Coresidence</i>	-0.63	0.411	0.12	-0.64*	0.382	0.10
<i>Foreign Diploma</i>	0.25	0.627	0.69	0.24	0.664	0.71
Deviance	315			315		
No. Obs.	417			417		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Note: M1 presents the full regression model used to compute probabilities in 1, Panel B.

M2 presents the same model, with a 2-level educational variable for the children, as discussed in the text. The level 'Below H. S. Degree' pools together the categories 'Below H.S. Degree' and 'H.S. Degree' from M1.

Table S4: Regression models from Figure 2, panel B

	<i>Child Education</i>					
	Underestimation			Overestimation		
	log(OR) ^{1,2}	SE ²	p-value	log(OR) ^{1,2}	SE ²	p-value
<i>Coresidence</i>	0.82*	0.448	0.070	-0.37	0.441	0.40
<i>Mother</i>	0.24	0.326	0.45	-0.42	0.315	0.18

<i>Daughter</i>	-0.01	0.328	0.98	0.27	0.313	0.39
<i>Age Parent</i>						
Less than 48	—	—		—	—	
From 48 to 51	-0.42	0.493	0.40	0.84	0.510	0.10
From 52 to 55	-0.24	0.419	0.57	1.2**	0.485	0.016
Over 56	-0.35	0.547	0.52	1.7***	0.508	0.001
<i>Age Child</i>						
Less than 20	—	—		—	—	
From 20 to 22	-0.30	0.439	0.49	0.26	0.453	0.57
From 23 to 25	0.11	0.576	0.85	0.47	0.596	0.43
Over 26	-0.63	0.764	0.41	0.87	0.651	0.18
<i>Migration Status Child</i>						
Child of 2 immigrants	—	—		—	—	
Child of 1 immigrant	0.24	0.500	0.63	-1.4***	0.526	0.009
Other	0.38	0.462	0.41	-1.0**	0.463	0.027
<i>Education Child</i>						
below H.S. degree	—	—		—	—	
H.S. degree	-0.15	0.517	0.78	-0.06	0.444	0.89
above H.S. degree	1.3**	0.557	0.021	-0.84*	0.457	0.067
<i>Educational Distance</i>	0.31*	0.178	0.079	-0.88***	0.223	<0.001
<i>Student</i>	0.23	0.350	0.51	0.30	0.423	0.48
<i>Foreign Diploma</i>	0.70	0.700	0.32	-0.39	0.513	0.45
<i>Time between interviews</i>						
6 months or less	—	—		—	—	
From 6 months to a year	1.1***	0.352	0.001	-0.02	0.343	0.94
Over a year	0.36	0.643	0.57	-0.49	0.662	0.46
Deviance	292			276		
No. Obs.	410			410		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Table S5: Regression models from Figure 2, panel C

	<i>Parent Education</i>					
	Underestimation			Overestimation		
	log(OR)^{1,2}	SE²	p-value	log(OR)^{1,2}	SE²	p-value
<i>Coresidence</i>	0.32	0.550	0.57	0.39	0.389	0.32
<i>Mother</i>	0.36	0.407	0.37	-0.21	0.334	0.54
<i>Daughter</i>	0.44	0.387	0.26	0.09	0.291	0.76
<i>Age Parent</i>						
Less than 48	—	—		—	—	
From 48 to 51	0.50	0.542	0.36	-0.28	0.441	0.53
From 52 to 55	0.46	0.564	0.42	-0.22	0.436	0.62
Over 56	-0.35	0.808	0.40	-0.43	0.496	0.39
<i>Age Child</i>						
Less than 20	—	—		—	—	
From 20 to 22	-0.98**	0.461	0.035	-0.18	0.458	0.69
From 23 to 25	-0.24	0.574	0.68	0.43	0.446	0.33
Over 26	-0.42	0.727	0.57	1.0*	0.531	0.054
<i>Migration Status Child</i>						
Child of 2 immigrants	—	—		—	—	
Child of 1 immigrant	-1.3**	0.621	0.038	0.75	0.491	0.13
Other	0.17	0.573	0.76	1.3***	0.459	0.005
<i>Education Child</i>						
below H.S. degree	—	—		—	—	
H.S. degree	0.48	0.614	0.43	0.55	0.483	0.26
above H.S. degree	0.74	0.704	0.29	0.42	0.499	0.41
<i>Educational Distance</i>	-1.0***	0.257	<0.001	0.59***	0.186	0.002
<i>Student</i>	-0.27	0.424	0.53	-0.03	0.331	0.94
<i>Foreign Diploma</i>	-1.8***	0.566	0.001	0.19	0.615	0.76
<i>Time between interviews</i>						
6 months or less	—	—		—	—	
From 6 months to a year	0.61	0.396	0.12	0.71**	0.321	0.029
Over a year	0.18	0.737	0.80	0.66	0.569	0.24
Deviance	218			334		

No. Obs. 360 360

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Table S6: Regression models from Figure 3, Panel A - Dyads-only

	Self-Reports Only			Retrospective Proxy			Prospective Proxy		
	β^1	SE ²	p-value	β^1	SE ²	p-value	β^1	SE ²	p-value
<i>Education Parent (Self-Report)</i>	0.12***	0.037	< 0.001				0.17***	0.038	< 0.001
<i>Education Parent (Proxy-Report)</i>				0.15***	0.038	< 0.001			
Deviance	5,320			4,175			4,783		
No. Obs.	417			363			411		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Table S7: Regression models from Figure 3, Panel A - Simulations of Retrospective Reports

	Proxy Reports			Self-Reports		
	β^1	SE ²	p	β^1	SE ²	p
<i>Education Parent (Proxy-Report)</i>	0.16***	0.009	<0.001			
<i>Education Parent (Simulated)</i>				0.14***	0.009	<0.001
<i>Education Parent (Self-Report)</i>						
Deviance	108,825			110,451		
No. Obs.	8,400			8,400		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Table S8: Regression models from Figure 3, Panel A - Simulations of Prospective Reports

	Proxy Reports			Self-Reports Simulated		
	β^1	SE ²	p	β^1	SE ²	p
<i>Education Parent (Proxy-Report)</i>						
<i>Education Parent (Simulated)</i>						
<i>Education Parent (Self-Report)</i>	0.19***	0.011	<0.001	0.17***	0.012	<0.001
Deviance	465,473			468,579		
No. Obs.	31,387			31,387		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Description of the simulation procedure to obtain Figure 3

We simulate self-reports of the target's education in the full TeO2 sample by predicting errors in such reports in the dyad sample. We predict errors using a linear regression, where self-reports for both members of the dyads are available. We predict mistakes using self-reported variables only. We use age (child and parent), sex, coresidence status, migration status, whether the child is a student and education. Based on the regression models, we predict the probability for each proxy-reporter to make a mistake. We model mistakes linearly. Using the probability that we predicted, we run binomial trial to decide whether the individual will or not make a mistake, and how large. When the individual makes a mistake, we construct a corrected report, which is then considered to be the self-report (thus, simulated). This way, we reconstruct an "error-free" measure of education for the absent dyad member. Thus, we compare social mobility estimates based on two proxy-reported measures: one with and one without error. In order to present plausible estimations, we restrict the sample so that common support is verified on

all the variables used in the preliminary regressions. Coding material is available at https://osf.io/nvp96/?view_only=b38b885f3b974f199991ae2dcd788a24.

Robustness checks

Table S9: Availability of Information: Are Missing Proxy-Reports due to ‘Don’t know’ Self-Reports?

	<i>Self-Reports</i>			
	Parent Education		Child Education	
	Reported	Don’t know	Reported	Don’t know
Proxy-Reports				
Reported	86 (361)	NA (0)	99 (413)	NA (0)
Don’t know	14 (57)	NA (0)	1.4 (6)	NA (0)

Source: TeO2 (Beauchemin *et al.*, 2023) and authors’ calculations.

Note: Percentages (N). There is 1 missing answer in children’s self-reports, 3 in parents’ self-reports. There is 1 missing parents’ proxy-reports.

Table S10: Multinomial models: Predicting Agreement, Underestimation and Overestimation in Child Education

	Underestimation			Overestimation		
	$\log(\text{OR})^{1,2}$	SE^2	p-value	$\log(\text{OR})^{1,2}$	SE^2	p-value
<i>Coresidence</i>	0.67	0.445	0.13	-0.28	0.455	0.54
<i>Mother</i>	0.19	0.323	0.55	-0.41	0.341	0.23
<i>Daughter</i>	0.15	0.309	0.63	0.24	0.329	0.46
<i>Age Parent</i>						
Less than 48						
From 48 to 51	-0.23	0.473	0.62	0.96	0.594	0.10
From 52 to 55	0.02	0.425	0.97	1.3**	0.587	0.026
Over 56	0.06	0.489	0.91	1.9***	0.621	0.003
<i>Age Child</i>						

Less than 20						
From 20 to 22	0.17	0.446	0.70	0.13	0.484	0.79
From 23 to 25	0.92*	0.499	0.065	0.40	0.553	0.47
Over 26	0.27	0.621	0.67	0.69	0.619	0.26
<i>Migration Status</i>						
<i>Child</i>						
Child of 2 immi- grants						
Child of 1 immi- grant	0.26	0.504	0.61	-1.5***	0.543	0.007
Other	0.48	0.451	0.29	-1.1**	0.461	0.015
<i>Education Parent</i>						
below H.S. de- gree						
H.S. degree	-0.27	0.590	0.65	-0.04	0.505	0.93
above H.S. de- gree	-0.14	0.490	0.78	-0.87*	0.454	0.055
<i>Educational Dis- tance</i>	0.29	0.221	0.18	-1.1***	0.227	<0.001
<i>Student</i>	0.44	0.364	0.23	0.47	0.415	0.26
<i>Foreign Diploma</i>	0.37	0.646	0.57	-0.08	0.548	0.89
<i>Time between in- terviews</i>						
6 months or less						
From 6 months to a year	0.99***	0.331	0.003	0.17	0.347	0.63
Over a year	0.42	0.626	0.50	-0.30	0.622	0.63
Deviance	567			567		
No. Obs.	410			410		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

Table S11: Multinomial models: Predicting Agreement, Underestimation and Overestimation in Parent Education

	Underestimation			Overestimation		
	$\log(\text{OR})^{1,2}$	SE^2	p-value	$\log(\text{OR})^{1,2}$	SE^2	p-value
<i>Coresidence</i>	0.39	0.522	0.45	0.39	0.391	0.32
<i>Mother</i>	0.28	0.409	0.50	-0.20	0.300	0.50
<i>Daughter</i>	0.54	0.388	0.17	0.19	0.294	0.53
<i>Age Parent</i>						
Less than 48						
From 48 to 51	0.53	0.519	0.30	-0.15	0.448	0.73
From 52 to 55	0.51	0.530	0.33	-0.11	0.424	0.80
Over 56	-0.78	0.704	0.27	-0.53	0.478	0.27
<i>Age Child</i>						
Less than 20						
From 20 to 22	-0.90*	0.502	0.073	-0.29	0.441	0.51
From 23 to 25	-0.16	0.559	0.77	0.25	0.491	0.61
Over 26	-0.17	0.702	0.81	0.85	0.562	0.13
<i>Migration Status</i>						
<i>Child</i>						
Child of 2 immigrants						
Child of 1 immigrant	-1.2*	0.678	0.072	0.56	0.504	0.27
Other	0.48	0.451	0.42	1.3***	0.451	0.003
<i>Education Parent</i>						
below H.S. degree						
H.S. degree	0.65	0.559	0.24	0.61	0.491	0.21
above H.S. degree	0.60	0.511	0.24	0.22	0.448	0.62
<i>Educational Dis-</i>	-0.60**	0.280	0.032	0.62***	0.210	0.003
<i>tance</i>						
<i>Student</i>	-0.25	0.427	0.55	-0.03	0.357	0.94
<i>Foreign Diploma</i>	-1.8***	0.594	0.003	-0.04	0.645	0.96
<i>Time between in-</i>						
<i>terviews</i>						

6 months or less						
From 6 months to	0.79**	0.392	0.045	0.80***	0.305	0.008
a year						
Over a year	0.39	0.766	0.61	0.73	0.533	0.17
Deviance	531			531		
No. Obs.	360			360		

¹*p<0.1; **p<0.05; ***p<0.01

²OR = Odds Ratio, SE = Standard Error

Source: TeO2 (Beauchemin *et al.*, 2023) and authors' calculations.

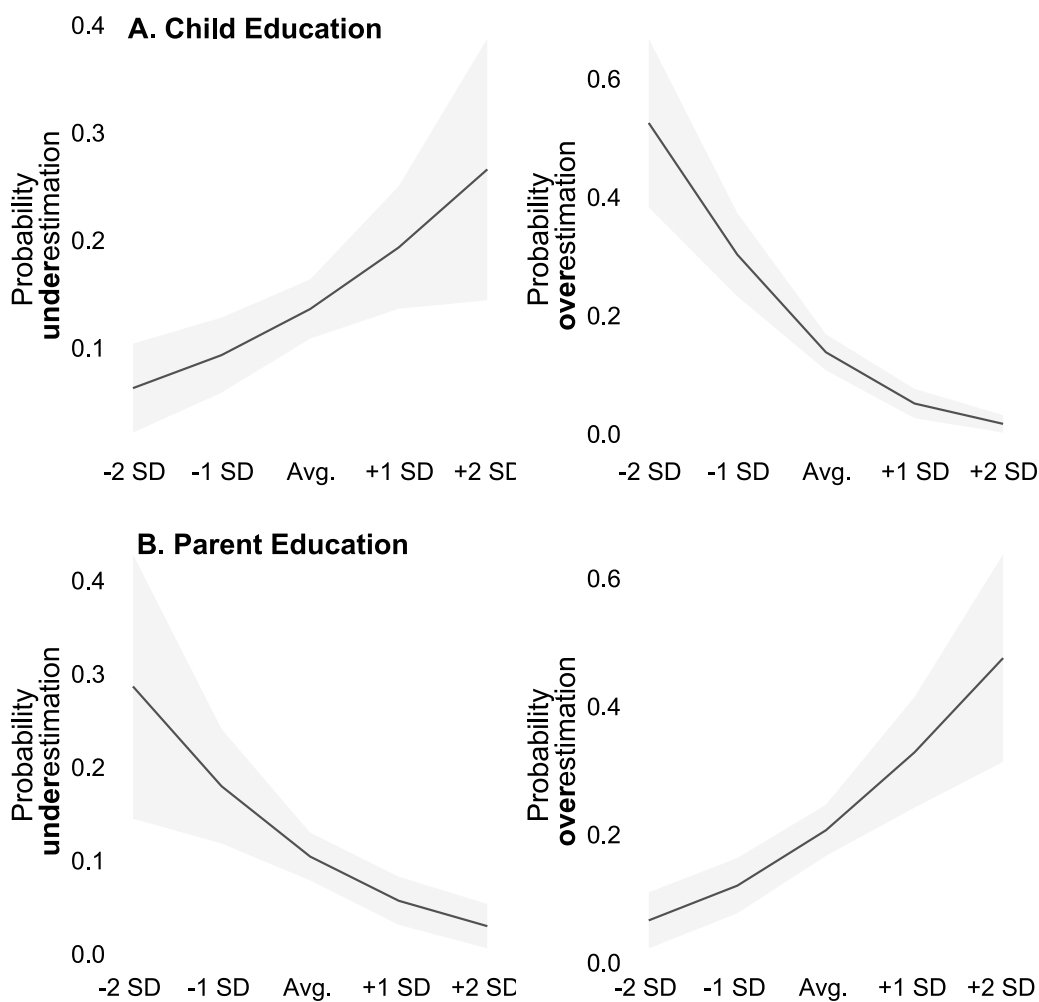


Figure S1: Probability to over- or underestimate education, non-standardized educational distance

Notes: Predicted probabilities based on a logistic regression model also controlling for age, sex, education and migration background of child, adding cohabitation status, if the parent holds a foreign diploma, if the child is a student and time between parent and child interviews. We add the education of the proxy-reporter (in Panel B, child's education, in Panel C, parent's education) 90% confidence intervals based on standard errors clustered at the family level.

Source: TeO2 and authors' calculations.

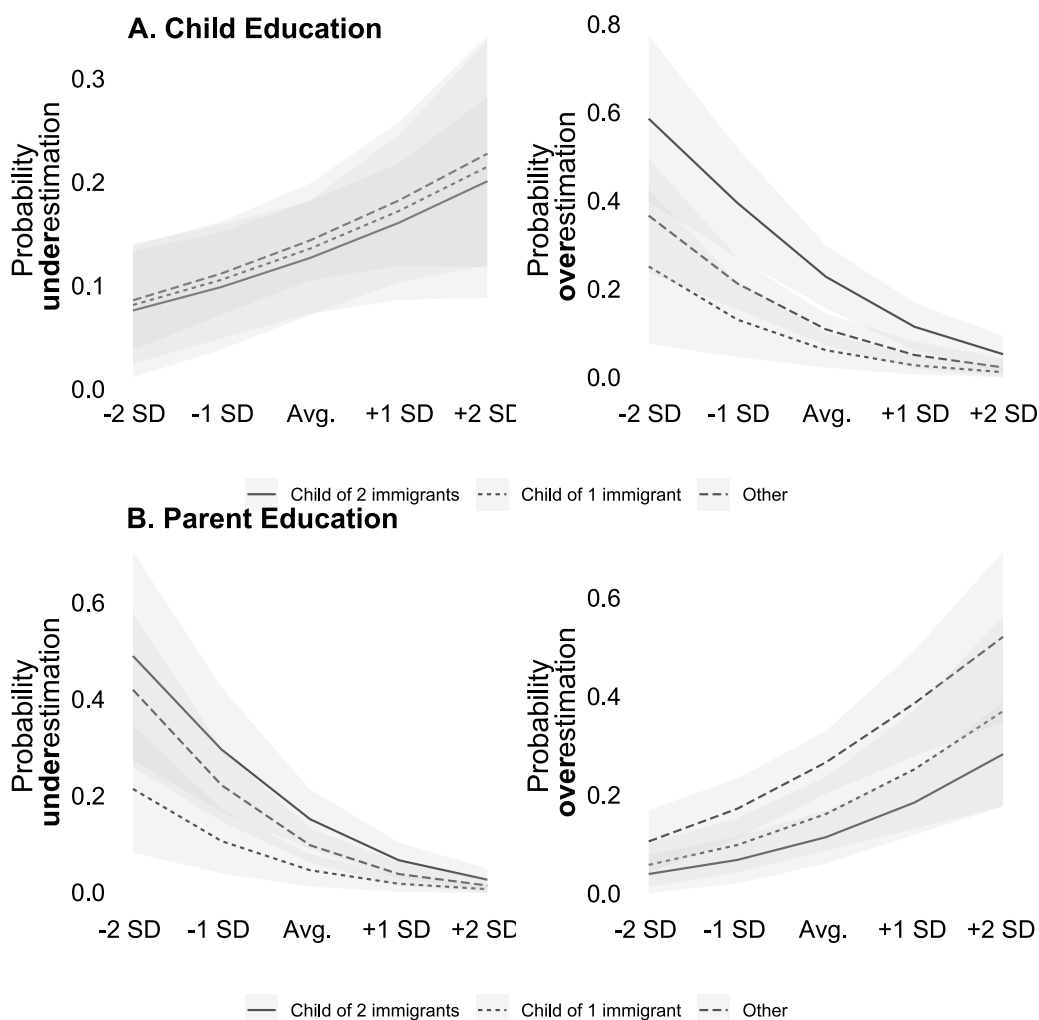


Figure S2: Probability to over- or underestimate education stratified by migrational status of child

Notes: Predicted probabilities based on a logistic regression model also controlling for age, sex, education and migration background of child, adding cohabitation status, if the parent holds a foreign diploma, if the child is a student and time between parent and child interviews. We add the education of the proxy-reporter (in Panel B, child's education, in Panel C, parent's education) 90% confidence intervals based on standard errors clustered at the family level.

Source: TeO2 and authors' calculations.