

I think my father was a project manager: Reliability and validity of parental occupation and schooling measurement in the HILDA survey

Introduction

The socio-economic status (SES) of an individual is a significant factor that influences a variety of aspects of their life, including family formation patterns and health outcomes. While SES is a multidimensional construct whose definition is the subject of ongoing debate (Oakes & Andrade, 2017), it is most commonly operationalised and measured using characteristics such as income, occupation, and education level (Long & Renbarger, 2023). While SES is usually measured at the individual level, there are however a number of contexts in which the SES level of their mother or father is of primary interest.

It is common for research to use parental SES as an independent variable in studies where it is not possible or suitable to measure an individual's own occupation or education level. This is the case for studies of the academic, behavioural, and health outcomes of children and adolescents, as these population do not yet have their own occupational or educational characteristics. Parental SES measures are also used in research investigating the effect of childhood circumstances on health or behavioural outcomes in later life (Andersson & Upenieks, 2023; Huurre et al., 2003), and it is central to studies of intergenerational mobility where the occupation or education level of parents and children are compared. Parental education is also commonly used as an instrumental variable for individual education, because it is highly correlated with a person's own schooling but may be plausibly exogenous to certain outcomes of interest (Black et al., 2005).

The widespread use of these variables is reflected in their inclusion in many major surveys internationally, including the Demographic and Health Surveys, the Programme for International Student Assessment (PISA), as well as large-scale household socio-economic surveys such as the National Longitudinal Survey of Youth (NLSY) in the United States, and in Australia, the Household, Income and Labour Dynamics in Australia (HILDA) Survey. As most surveys sample individuals rather than whole households or families, information about parental occupation and education is typically derived through proxy reports where people report this information on behalf of their mother and father. As with any measure, it is important for the measure to understand and assess the quality of the data. For parental SES questions about the degree to which respondents are able to provide valid and reliable proxy reports regarding their parents occupation and education have persisted since they first came into widespread use. A key concern is whether reporting accuracy is systematically related to SES itself.

In this study, we examine two key dimensions of data quality with respect to the parental occupation and education reports in the HILDA Survey. Because HILDA is a household panel survey, parents and children can be identified and their responses compared. We use this feature of the survey design to assess validity by comparing children's proxy reports of parental occupation and education against the reports provided by parents themselves. We assume that the parent's own reports are accurate and assess the degree to which the children's reports agree. Second, we assess reliability by comparing the reports provided by different children of the same parent. Where multiple children within a household each report on their father's or mother's occupation, the degree of agreement between siblings provides an independent measure of reporting consistency that does not rely on either report being treated as a gold standard. Together, these two approaches allow us to evaluate the quality of these widely used measures and to assess whether the patterns of measurement error documented in

international studies, including differential accuracy by respondent socioeconomic background, are also present in the Australian context.

Background

The measurement of socio-economic status, is prone to a number of systematic and random measurement errors. In particular, occupation has been described as '*one of the most challenging variables social scientists regularly encounter*' (Martin-Caughey, 2026). In many surveys, including HILDA, occupation information is collected in open text format which is then subsequently coded into an occupational classification system. The quality of the occupation information provided by respondents can vary significantly, and two respondents with the same occupation may provide very different descriptions about their jobs (Kim et al., 2020). While brief and vague answers to occupation questions are one common problem, lengthy responses which contain conflicting or superficial details, can equally cause issues when it comes to coding the answers to an occupation (Conrad et al., 2016). However, even with clear information from respondents some occupations are inherently more difficult to code than others. Some occupations such as pharmacist, or postal service worker are more likely to be described in terms of their titles or associated tasks with a high level of similarity across respondents, while for other occupations, titles as well as task descriptors vary significantly (Martin-Caughey, 2026).

Research on the coding of occupations, across multiple international classifications systems and national frameworks, has demonstrated that occupational coding from the raw text data is a highly complex exercise that can involve a relatively high degree of subjective interpretation (Massing et al., 2019), and subsequently high levels of error. This can be seen in studies looking at inter-coder reliability, or the extent to which different coders assign a particular record the same occupation code. Generally, inter-coder reliability is relatively high with broad 1-digit level groupings, with agreement rates then declining as the levels become more detailed and granular (Belloni et al., 2016). For example, in a recent evaluation of coding in the Korean General Social Survey, found an average 69% match rate for 1-digit major groupings of occupation, falling to 49% at the 4-digit detailed unit group level (Kim et al., 2020). Much higher and lower agreement rates have been found across other studies (Conrad et al., 2016; Ge et al., 2026; Massing et al., 2019). An early assessment of the quality of occupation coding in HILDA found inter-coder agreement levels at broadest level ranging 85-94% across three selected samples of occupations coded in Wave 6 (Watson & Summerfield, 2009).

While the collection of education information is more straightforward compared to occupation, one issue that can impact on the quality of education information is that respondents may face 'transposition difficulties' in translating older or foreign educational credentials into contemporary national classification schemes (Lavest et al., 2025). While the issues explored above affect reporting of the quality of data of one's *own* occupation and education, additional errors may be introduced when occupation and education data is reported by someone else, as in the case of proxy reporting by children (Houseworth & Fisher, 2020).

For proxy reporting of parental schooling and occupation, the respondent's knowledge and familiarity with the information is thought to be a key determinant of the overall quality of the data provided. A parent's occupation is something respondents are expected to have a relatively good knowledge of because parental occupation is tied to observable daily routines and is a frequent topic of conversation in most families, which means it is information that is accessible to children from an early age (Jerrim & Micklewright, 2014). In contrast, a parent's

level of education may be something respondents are less familiar with, or that is a less salient topic to them, because it would be something that, in most cases, had been achieved before the respondent was born and would not have been part of day-to-day information in the same way as occupation.

These considerations suggest that the data quality of proxy reports may differ systematically across the types of parental characteristics being reported. A limited number of previous studies have investigated this possibility using a range of methodological approaches, including parent-child comparisons, sibling concordance checks, linkage to administrative records, and test-retest designs. Each of these approaches captures a different dimension of data quality.

Missing data

A key indicator of data quality in survey responses, is the level of item non-response or the proportion of respondents who are unable, or unwilling, to provide an answer to the question. For parental SES measures the level of non-response has been found to vary according to the variable being studied, the characteristics of the respondents, as well as aspects of the survey delivery. While some of these variations are relatively well documented across numerous studies, others show mixed evidence across different studies.

Non-response rates have been found to vary according to whether the question relates to parental occupation or schooling, though the direction of this difference is not consistent across studies. Some studies report a higher rate of missing answers for parental education compared to parental occupation (Lavest et al., 2025; Lien et al., 2001; Pu et al., 2011), while others find the opposite (Engzell & Jonsson, 2015; Pueyo et al., 2007).

Non-response may also vary by how questions are formatted and asked (Avvisati & Wuyts, 2024), and whether a '*don't know*' option is included can also make a large difference to response patterns. Survey mode has also been shown to be associated with the completeness of information provided. In a study of the adult respondents in the South Australian Health Omnibus, the proportion of respondents with missing answers on their father's and mother's highest education level was significantly lower in the face-to-face interview (14% and 12%, respectively) than in telephone interviews (20%) (Chittleborough et al., 2008).

Non-response to parental SES questions may also be associated with respondent characteristics and family background. A consistent finding across a variety of studies is that growing up in non-standard family structures, including single-parent homes, is associated with significantly higher rates of missing answers particularly with respect to information about a parent who is absent or did not live with the child (Pueyo et al., 2007; Svedberg et al., 2016).

Validity

Among respondents who do provide an answer on the parental SES measures, there is still a question about whether the proxy report is valid, i.e. whether it reflects what the parent actual occupation or education. The most common approach to assessing this aspect of data quality, is criterion validity, in which proxy reports from children are compared with direct information about the parents. The level of congruency can be quantified using measures such as Cohen's kappa, weighted kappa, percentage agreement, and correlation coefficients (Looker, 1989). In most validation studies, the required data is obtained by interviewing parents themselves and comparing the answers to those provided by their children (Avvisati & Wuyts, 2024; Ensminger et al., 2000; Lien et al., 2001; West et al., 2001), though some studies have also used

administrative or register data to obtain the comparison parent information (Engzell & Jonsson, 2015). Most studies suggest a reasonable level of agreement for parental occupation, though there can be significant variation across different samples. These studies tend to be based on contemporaneous reports of parent occupation, and therefore not necessarily generalisable to situations where respondent asked to recall historical parental occupation (Houseworth & Fisher, 2020), as in the case of HILDA where aged 14 as in the case in HILDA.

The validity of parental education reports is more mixed, although a relatively consistent finding is that agreement tends to be lower for education than for occupation (Avvisati & Wuyts, 2024; Lien et al., 2001). However, there are some exceptions where education is reported with greater congruence than occupation (Pu et al., 2011). In terms of respondent characteristics, there are few clear patterns across demographic characteristics. However, as with non-response, validity is significantly poorer when the measure regards a parent who was/is not living in the household (Ensminger et al., 2000), and a positive association has also been found between higher levels of parent-child agreement in households with higher levels of SES (Ensminger et al., 2000; Pu et al., 2011).

Reliability

Another dimension of data quality, is the reliability or the consistency of results. To date, reliability of parental SES measures has been assessed in two ways: by comparing reports provided by different siblings within the same family (Ward, 2011), and by examining the consistency of an individual's reports across survey waves (Aarø et al., 2009).

Ward (2011) used data from the National Survey of Midlife Development in the United States (MIDUS), examined agreement between 1,280 adult sibling pairs, including 804 twin pairs on different measures of childhood SES. As with validity, the number of categories used in the classifications of occupation and education have has a substantial effect on the level of agreement observed. For father's occupation, concordance was 0.76 when occupation was coded into nine categories. For parental education, concordance was lower at 0.73 for father's education and 0.69 for mother's education, with weighted kappas of 0.78 and 0.73.

Concordance for father's occupation did not vary by age, sex, income level, or whether the sibling pair were twins or not. However, concordance for parental education was significantly lower among sibling pairs where at least one member had a high school education or less, while college-educated pairs showed markedly better agreement.

The existing literature suggests that the overall quality of proxy reports of parental SES, particularly as assessed by parent-child agreement, varies considerably across different studies. An important result, that is relatively consistent across the different measures of data quality is that greater socio-economic disadvantage associated with poorer reporting quality and higher rates of missing data (Ensminger et al., 2000; Lavest et al., 2025). This pattern indicates that measurement error is potentially correlated with both the independent and dependent variables in many empirical designs. Consequently this could have implications for the estimation of socioeconomic gradients and intergenerational associations (Caro & Cortés, 2012).

However, several gaps remain in our understanding of these dynamics. First, much of the empirical evidence regarding the quality of about proxy reporting of parental SES measures comes from European and North American contexts (Domarchi et al., 2026; Engzell & Jonsson, 2015; Lien et al., 2001). Second, the majority of studies have focused on child or young

adolescent samples (Ensminger et al., 2000; Lien et al., 2001), leaving the proxy reporting of older adolescents, young adults, and older adult offspring understudied (Lavest et al., 2025). Related to existing validation studies typically compare contemporaneous reports, failing to account for the historical, retrospective nature of the information that respondents are asked to recall in many major social surveys. Finally, historically many studies have examined validity (parent-child agreement), reliability (sibling concordance), and missingness (item non-response) in isolation rather than as interconnected dimensions of data quality.

In this study, we address these gaps by examining the quality of parental occupation and education reports in the HILDA Survey in terms of their validity, and reliability. The household structure of HILDA, combined with its longitudinal design and rich demographic information, means we can assess each dimension while accounting for a wide range of respondent, parent, and survey characteristics that may be associated with reporting quality. In doing so, we provide the first comprehensive assessment of these widely used measures in the Australian context.

Data

Data come from Waves 1 to 23 of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. We examine two key measures of parental SES:

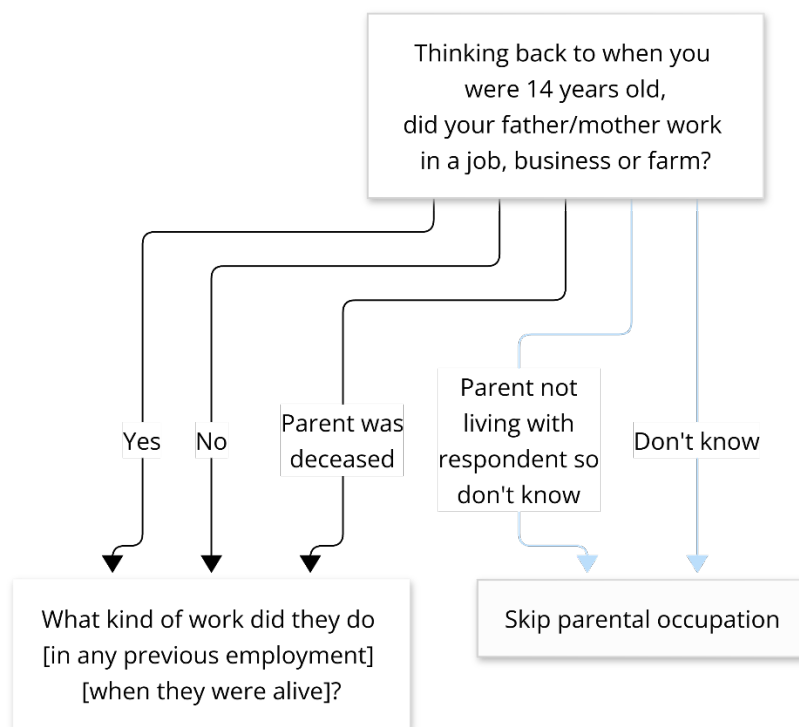
- 1) Mothers and father's occupation, as categorised into the 8 major groups of the Australian and New Zealand Standard Classification of Occupations (ANZCO), First Edition/2006 standard.
- 2) Mother's and father's highest level of education
 1. University degree
 - 2.

For each of these measures we investigate validity by examining the degree of parent-child congruence, and the reliability as measured by the congruence of proxy reports across children of the same parent.

Proxy reports of parental occupation and education across the waves.

In HILDA, respondents are asked about their *own* occupation and education information at every wave. When a respondent is 'new' and is participating for the first time, they are also asked a series of one-off questions about their family background. These family background questions include parental SES measured by mother's and father's occupation around the time the respondent themselves was aged 14, and their highest level of education.

Specifically, for parental occupation, they are asked:



There are some differences in who was asked the occupation and education questions across the waves. These differences were taken into account when selecting the analytical sample for the study¹. As the sample selection and creation of the outcome measures was relatively complex, we outline the key points related to the data and methods for each component in Table 1, and explain it in further detail below. For each component we detail the process used for 1) the parent-child agreement analysis, and 2) the sibling-pair agreement analysis.

Sample selection

Parent-child agreement

For the analysis of parent-child agreement in parental occupation, the ‘child’ respondent was included if they participated for the first time in Wave 1, but were not living at home with both parents at the time, or if their first participation occurred anytime from Wave 2 onwards. The ‘parent’ respondent information was included from all waves with their occupation history reconstructed. For the parental education analysis, the ‘child’ respondent was included if they were a continuing respondent in Wave 5, or a new respondent at any wave from Wave 5 onwards, inclusive.

All respondents who were a biological parent to a child were identified and separated into a parent dataset. A dataset was created that linked the ‘child’ respondents to their biological fathers and mothers from the parent dataset using the cross-wave identifiers of parents (*hhbfxid*

¹ Parental education was only asked from Wave 5 onwards, however unlike other family background questions, at Wave 5 continuing respondents were also asked the question. Parental occupation was asked from the first wave, however in Wave 1 only, respondents were not asked the questions on parental occupation if they were still living at home with both their parents. In all other waves, new respondents were asked about their parent’s occupation, regardless of whether they lived with neither, one or both parents.

and *hhbmxd*). Respondents were excluded if their biological mother or father had never been a respondent themselves at any wave of the HILDA survey.

Cases where the responding child had grown up living with a step-parent of the same sex as the target parent were excluded. For example, when investigating father-child agreement in occupation reports, we excluded child respondents who indicated that at age 14 they had grown up with a mother and a stepfather. Similarly, for the analysis mother-child agreement we excluded child respondent that at age 14 had been living with their father and a stepmother. This exclusion was necessary because the survey questions regarding parental occupation and education instructs respondents to report on the person who “*mostly filled the role of mother/father*”. If they lived with a step-parent we cannot ascertain if they were reporting the occupation/education of their biological parent or their step-parent, and if it was a stepparent, the child’s report cannot be meaningfully compared with the biological parents’ own self-reported occupation.

For similar reasons we excluded respondents who had not been living with their own parents at age 14 because they had been fostered or adopted out², as we are not able to exclude that the child report was not in relation to their foster parent’s SES measures.

The final analytical sample for the occupation analysis was 1,901 father–child pairs and 3,201 mother–child pairs, while for the the education analysis there were 2,593 father–child pairs and 3,778 mother–child pairs.

Sibling-pair agreement

For the reliability analysis the aim was to compare how similar answers were between pairs of siblings. Siblings were identified from the household relationship grid, with any respondents who were recorded as ‘*Brother or sister – both parents same*’ in any wave from Wave 1-23 flagged. A union-find algorithm was then applied to the sibling data to assign a unique family identifier to each group of siblings. Siblings were matched even if they had never been recorded as living in the same household, as long as each one had lived with another full sibling at some point³. Where a respondent had multiple siblings, sibling pairs were created pairing siblings in sequence of their age. The final analytical sample is composed of 1,453 and 1,485 sibling pairs for the analysis of fathers and mothers’ occupation, respectively and 1,314 and 1,385 sibling pairs for paternal and maternal education measures.

Dependent and independent variables

Parent-child agreement

The proxy reports of parental occupation were compared with the parent’s own details about their occupation and education. For occupation the information is based on occupations measured according to the eight major group level of ANZCO 2006:

1. Managers
2. Professionals

² Identified by a code of 15 for variable *fmnprea* (History: Why were you not living with both your parents at age 14)

³ For example if Person A and Person B are identified as full-siblings living in the same household, and person A then moves out, while Person C moves in and is also identified as a full-sibling of Person B, then we assign Person C as a full-sibling also of Person A, even though Person A and Person C were never observed in the same household simultaneously.

3. Technicians and Trades Workers
4. Community and Personal Service Workers
5. Clerical and Administrative Workers
6. Sales Workers
7. Machinery Operators and Drivers
8. Labourers

The dependent variable was constructed as a three category variable to distinguish between agreement (1), substantive disagreement (category 2), non-response or lack of knowledge (category 3):

- (1) Agreement/match, where the child's report and the parent's self-report fall in the same ANZSCO major group;
- (2) Disagreement/mismatch, where both parties reported an occupation but it differs;
- (3) Child did not know, where the child indicates that they did not know their parent's occupation or parent was not living with them; child's report unable to be coded, where the child provided an occupation description that could not be assigned an ANZSCO code.

This three-category specification was also applied to the correspondence in parent-child reports of parent's education.

Creating these correspondence variables involved first creating variables for the child and for the parent that ensured the information was organised in the same categories. This involved some processing of the proxy reports of education provided by children to match the information provided by parents, and processing of the parental occupation variable to ensure it was reflective of the parent's occupation at the time the linked child was 14, or as close as possible to that time.

For highest level of education, since this variable tends to be fixed over time, we used the education level reported by the parent at the first time they participated. Education was coded into four substantive categories for both parent and child reports: university degree, other post-school qualification, Year 12 completion, and Year 11 or below. For the parent responses this information was taken the derived variable describing their highest level of education. For the child responses this categorisation was built up from three separate questions.

- The first one asked how much schooling their parent had (primary, secondary but no more than year 10, year 11 or equivalent, year 12 or equivalent, or don't know).
- The second asked if the parent had a post-school qualification.
- If they said yes, a third question asked what the highest level of qualification obtained was (university, teachers college, institute of technology, technical college/TAFE, Employer, or other, don't know).

If respondents answered don't know at any of the questions they were coded as a separate category.

For parental occupation, as noted earlier, respondents were asked to recall to when they were around 14 years old. Given that occupations changed over time, the occupation history of parents was reconstructed in order to identify the current or most recent occupation that they had, at the time the linked child was around 14 years old. A flag variable was used to identify if the parent's own report of their occupation when the target child was 14 was based on 1)

prospective information, or reconstructed retrospective information based on 2) previous occupations, or 3) historical occupation⁴.

Table 1 Outline of sample selection and model specification for validity and reliability analysis

	Validity (parent-child agreement)				Reliability (sibling-pair agreement)			
	<i>Occupation</i>		<i>Education</i>		<i>Occupation</i>		<i>Education</i>	
Sample selection								
<i>Child respondent sample</i>	- Wave 1 respondent, if not living at home with <i>both</i> parents - New respondents from Wave 2 onwards. Exclusion: - Respondents who lived with a stepparent of the same sex as target parent when aged 14. - Respondents who were not living with both parents at age 14 due to being fostered out.		- Continuing or new respondent in Wave 5. - New respondent from Wave 6 onwards.		Same selection as for validity study, but applicable to at least two children who are identified as full siblings.			
<i>Parent respondent sample</i>	All biological parents who participated in at least one wave of HILDA.				n/a		n/a	
Model specification								
<i>Dependent variable</i>	<i>Agreement between parent and child.</i>				<i>Agreement between two siblings</i>			
	(1) Child’s report is same as parent’s self-report (2) Child’s report differs from parent’s self-report (3) Child did not know.				(1) Sibling pair reports are same (2) Sibling pair reports differed, including cases where one or both siblings did not know.			
<i>Independent variables</i>	Child characteristics, parent characteristics, survey characteristics				Combined sibling characteristics, parent characteristics as reported by the youngest child in the sibling pair.			
<i>Model</i>	Multinomial logistic regression, clustered standard errors taking into correlation at parent level.				Logistic regression, clustered standard errors taking into correlation at family level.			
<i>Number of records</i>	Father 1,901 children	Mother 3,201 children	Father 2,593 children	Mother 3,778 children	Father 1,453 sibling pairs	Mother 1,232 sibling pairs	Father 1,314 sibling pairs	Mother 1,385 sibling pairs
<i>Number of clusters</i>	1,432 fathers	2,412 mothers	1,745 fathers	2,493 mothers	1,119 families	1,484 families	1,075 families	1,139 families

⁴ A number of assumptions were used in the reconstruction process. For example, consider a respondent who is a father with two children: one was aged 16 and one was aged 10, when the respondent participated in HILDA in Wave 1 (2001). In Wave 1 (2001), the respondent reported his occupation as Bus Driver, and noted that he had been with his current employer for the past 10 years i.e., since 1991. We assume from this that he was also a bus driver in 1998, when his first child turned 14. This reconstructed value is flagged as a "current job", since it is derived from the same job the respondent held at the time he reported it, using the years-with-employer variable to project backward. The respondent did not participate in HILDA again after Wave 1. Twenty years later, in 2021, his second child (now 30) takes part in HILDA and is asked what their father's occupation was when they were 14 (i.e., in 2005). Since we have no data on the respondent's occupation after 2001, we reconstruct his 2005 occupation by carrying forward his last known occupation (Bus Driver) and flagging it as a "known previous job". We then compare this reconstructed value against the child's report. Finally, a third flag, "historical", applies when the reconstructed occupation is based on information about a respondent's previous job- used for respondents who were not currently in the labour force at the time they answered the occupation question.

Independent variables were included relating to both the parent and the child. With respect to children we included:

- Sex
- Age at time of interview grouped into three categories (15–18, 19–24, and 25+)
- Whether was living at home at the time of proxy report
- Whether living with the target parent at age 14.
- survey mode (face-to-face with computer-assisted personal interviewing, telephone with computer-assisted interviewing, or prior to 2009 when mode was not separately recorded⁵).
- Presence of other adult at the time of interview.

For parents we included their country of birth, and presence of other adult at the time of interview. In addition, in the occupation model, we controlled for the parent's occupation category (as self-reported), and the occupation flag. For fathers, a binary variable was included describing if the parent had experienced a history of unemployment of 6 months or more. For the education model, we controlled for the parents self-reported level of education.

Sibling-pair agreement

For the reliability study, the dependent variable was operationalised as a binary one; equal to 1 if both children in a sibling pair matched on the parent's occupation or education and equal to zero if the two siblings gave different answers, or one of them did not know.

The independent variables included characteristics of the sibling pair, the parents, and, where relevant, each sibling's own report. Specifically, we included:

- Age difference between siblings, grouped into four categories (less than 2 years, 3 years, 4 years, and 5 or more years)
- Number of siblings in the family (2, 3, or 4 or more)
- Sibling gender composition (both male, both female, or mixed)
- Whether both siblings lived with the target parent at age 14, or at least one did not
- Whether both siblings lived at home at the time of interview, or at least one did not
- Parent's country of birth (Australia, other English-speaking country, or non-English-speaking country)

For the occupation model, we also included Sibling 1's report of the parent's occupation, coded to the one-digit ANZSCO (2006) major group level (Managers; Professionals; Technicians & Trades Workers; Community & Personal Service Workers; Clerical & Administrative Workers; Sales Workers; Machinery Operators & Drivers; Labourers), together with a "not able to be determined" and a "don't know" category. For the education model, we included Sibling 1's report of the parent's level of education (university, other post-school qualification, Year 12, Year 11 or below, or don't know). Sibling 1 was the younger of the siblings in each sibling pair.

As Sibling 1 was used as the reference report in both cases, as the covariate of interest is whether Sibling 2's report matched Sibling 1's, rather than the substantive occupation or education category itself.

⁵ Between Wave 1 and Wave 8, the field work was conducted by AC Nielsen, but from wave 9 the data this changed to Roy Morgan Research and this transition also involved a change in the method of data collection to computer-assisted personal interviewing (CAPI) (HILDA Survey Manual 2023).

Table 2 shows selected descriptive statistics and distribution of the dependent and independent variables used for the validity study. Overall, the level percentage of children who reported the same occupation for their parents, as the parents themselves did, at ANZCO level 1 was similar for mothers' and fathers' occupation at around 60%. For the parents' highest education level overall levels of disagreement are higher, and a higher percentage of children also did not know the parent's highest level of schooling.

The occupational distribution of parents differed substantially between fathers and mothers. The most common occupational categories for fathers were managers (23%), technicians and trades workers (20%), and professionals (18%), whereas for mothers the most common categories were clerical and administrative workers (22%), and professionals (19%). For parental education, the distributions are also different with the most significant difference being that mothers were more likely to have education up to Year 11 or below, whereas fathers were more likely to have an 'other' post-school qualification. This may be related to fathers being more likely to have a trade qualification.

The corresponding table of descriptive statistics for the sibling-pairs is shown in Table 3.

Method

Agreement between parent-child reports, and sibling-pair reports were first assessed using Cohen's kappa statistic and measurements of the percentage agreement.

To model the predictors of agreement, a multinomial logistic regression was used with the three-category agreement variable described earlier as the dependent variable. Multinomial logistic regression extends standard logistic regression to outcomes with more than two unordered categories, and estimates the log-odds of each outcome category relative to a reference category as a linear function of the independent variables. The model simultaneously estimates the effects of each predictor on the probability of mismatch versus match and on the probability of not knowing versus match, allowing the determinants of these two types of reporting failure to differ.

Standard errors were clustered at the parent level to account for the non-independence of observations where multiple children reported on the same parent. Predicted probabilities for each outcome category were computed from the estimated model and are reported with 95% confidence intervals.

For the sibling-pair analysis, the predictors of agreement were modelled using logistic regression. Standard errors were clustered at the family level, as some families contributed multiple sibling-pairs.

Table 2 Distribution of variables used for validity study, by parental SES measure, column percentages (%)

	Parental occupation		Parental education	
	Father-child	Mother-child	Father-child	Mother-child
	%	%	%	%
Dependent variable				
Parent-child match				
Match	63	56	68	61
Do not match	33	36	23	30
Child does not know/Unable to determine	4	8	9	9
Independent variables				
Child: Sex				
Son	52	50	52	50
Daughter	48	50	48	50
Child: Age at time of giving proxy report				
15-18	50	47	36	33
19-24	23	22	36	34
25+	27	31	28	33
Child: Living with parent(s) at time of proxy report				
No	41	45	42	47
Yes	59	55	58	53
Child: Lived with parent at age 14				
No	7	4	6	3
Yes	93	96	94	97
Parent: Country of birth				
Australia	69	71	70	71
English speaking country	12	12	12	11
Other	19	17	19	18
Parent: Own report of occupation				
Never worked		3		
Manager	23	9		
Professionals	18	19		
Technicians & Trades	20	4		
Community and Personal Service Work	4	15		
Clerical & Administrative	6	22		
Sales workers	4	9		
Machinery Operators & Drivers	14	3		
Labourers	13	17		
Parent: Occupation flag				
Current job	74	60		
Historical job	25	36		
Previous job	2	5		
Parent: Education				
University			25	19
Other post-school qualification			38	23
Year 12			8	13
Year 11 or below			29	45
Total %	100%	100%	100	100%
Total N parent-child pairs	1,901	3,201	2,593	3,778

Note: Additional controls were included for survey mode, and presence of others during the interview (not shown in table)

Table 3 Distribution of variables used for reliability study, by parental SES measure, column percentages (%)

	Parental occupation		Parental education	
	Father	Mother	Father	Mother
	%	%	%	%
Dependent variable				
Agreement between siblings in parental SES measure				
Yes	69	62	72	67
No	31	38	29	33
Independent variables				
Age difference between siblings				
<2 years	48	49	48	48
3 year	21	21	22	22
4 years	11	11	12	12
5+ years	19	20	19	19
Number of siblings in family				
2	67	68	67	68
3	24	24	24	24
4+	8	8	9	8
Sibling gender				
Both male	29	28	28	28
Both female	26	27	27	27
Mixed	45	45	45	45
Lived with target parent at age 14				
At least one sibling did not	20	8	17	6
Both siblings did	81	92	83	94
Lived at home at time of interview				
At least one sibling did not	63	64	61	62
Both siblings did	37	36	39	38
Parent country of birth				
Australia	69	70	68	70
Other English speaking country	10	10	9	10
Non-English speaking country	21	19	20	20
Sibling 1 report: Parent occupation				
Managers	18	6		
Professionals	19	20		
Technicians & Trades workers	20	4		
Community & Personal Service	4	10		
Clerical & Administrative	4	16		
Sales workers	4	7		
Machinery Operators & Drivers	10	2		
Labourers	10	11		
Not able to be determined	6	10		
Do not know	6	4		
Sibling 1 report: Parent level of education				
University			24	20
Other post-school qual			36	28
Year 12			7	11
Year 11 or below			23	31
Don't know			10	10
Total %				
Total N sibling-pairs	1,453	1,485	1,314	1,385

Results

Validity- Occupation

The validity of children's reports of parental occupation was assessed by comparing child responses to parent self-reports, among 1,901 father-child pairs and 3,201 mother-child pairs. Overall agreement was moderate for both parents. For fathers' occupation, 63% of child reports matched the father's own report ($\kappa = 0.56$), rising to 66% agreement ($\kappa = 0.59$) when cases where the child responded '*don't know*' or could not be coded were excluded. Agreement was slightly lower for mothers' occupation, at 56% ($\kappa = 0.49$), increasing to 61% ($\kappa = 0.54$) after excluding don't know/uncodeable responses.

A cross-tabulation of the parent's occupation, as reported by the parent themselves, compared to the report provided by their child(ren) is shown below in Table 4. It is apparent from this table that there are some similarities as well as differences in the agreement levels observed within different occupations when comparing mothers and fathers. For both parents, the 'Professional' category had the highest level of agreement between parents and children: In 75% of cases of fathers, and 78% of cases where mothers, reported an occupation coded as '*Professionals*', the responding child reported the same. For fathers, levels of agreement in their children's reports were also high for *Community & Personal Service workers*, *Managers*, and *Machinery operators and drivers*. In contrast for mothers, relatively high levels of mother-child agreement were evident where the mother's occupation was coded as '*Clerical & administrative worker*' (64%). If the mother had never worked, this also had a relatively high agreement (63%) regarding this fact.

We also see that some occupations were more likely to be classified differently by children, particularly for men, '*Machinery Operators & drivers*' and '*Labourers*'. For example, in 13% of cases where the father reported being a Machinery operator & driver, the child reported the father as being a 'labourer'. Conversely, in 11% of cases where the father reported being a 'labourer' the child reported the father as being a *Machinery Operator & Driver*.

There are some notable gender differences in the patterns of agreement between mother-child and father-child reports. Among mothers who were coded as being '*Managers*', only 35% of children gave a matching report, and instead there was a tendency for the child reports to be of lower level occupations including '*Professionals*', and '*Clerical & Administrative workers*'. For some occupations among fathers, in particular '*Clerical & Administrative workers*' and '*Sales*' children were more likely to report shift their father's reported occupation to a higher-level occupation such as *Manager*, although the number of cases of fathers with these occupations was small.

To further investigate the validity of parental occupation reports, multinomial logistic regression was used to examine factors associated with the likelihood of a child-parent match, mismatch, or the child not knowing/being unable to report. Predicted probabilities are reported in Appendix Table A1, and discussed below.

Table 4 Cross-tabulation of occupation (ANZCO 2006, 1 digit level) reported by parent, and occupation of parent as reported by the child, column percentages.

Occupation reported by father											
<i>Child's report</i>	Never worked	Managers	Professionals	Technicians & Trades	Community & Personal Serv.	Clerical & admin.	Sales	Machinery operators. & Drivers	Labourers	Total %	Total N
Managers		61	8	6	8	12	21	5	14	21	394
Professionals		11	75	6	3	14	9	0	2	19	354
Technicians & Trades		9	4	73	5	4	6	8	13	21	390
Community & Personal Service		2	2	1	69	2	2	1	0	4	77
Clerical & admin.		3	4	1	3	43	0	3	1	5	91
Sales		5	2	1	0	3	45	3	2	4	74
Machinery operators & Drivers		3	2	4	7	11	5	61	11	12	234
Labourers		3	1	4	5	8	9	13	51	11	204
Not able to be determined		3	2	3	0	2	2	4	4	3	53
Do not know		1	1	1	1	3	3	2	3	2	30
Total %		100	100	100	100	100	100	100	100	100	
Total N		438	341	374	77	117	67	250	237		1,901
Occupation reported by mother											
<i>Child's report</i>	Never worked	Managers	Professionals	Technicians & Trades	Community & Personal Serv.	Clerical & admin.	Sales	Machinery operators & Drivers	Labourers	Total %	Total N
Managers	3	35	3	5	5	3	6	5	4	7	224
Professionals	0	15	78	6	10	8	2	2	3	19.9	637
Technicians & Trades	4	4	1	47	3	1	3	4	4	4.2	133
Community & Personal Service	6	6	5	4	49	3	6	4	5	11.2	360
Clerical & admin.	2	14	3	9	6	64	8	6	4	18.7	600
Sales	1	9	2	7	4	5	49	6	6	8.5	273
Machinery operators & Drivers	1	1	0	3	0	1	1	38	4	2.2	72
Labourers	7	5	1	8	10	3	6	14	48	12.2	390
Not able to be determined	10	4	2	2	4	5	6	4	6	4.7	149
Do not know	3	2	3	3	3	3	5	8	4	3.3	107
Never worked	63	4	2	7	7	5	7	10	12	8	256
Total %	100	100	100	100	100	100	100	100	100	100	
Total N	99	294	593	121	477	699	274	101	543		3,201

Overall sex differences were modest, with little difference in the accuracy of the reports provided by sons versus daughters for either mothers or fathers. Respondents who were younger at the time of making the proxy report (aged 15-18) had a higher predicted probability of providing an occupation report for their parent that matched the parent's own report.

Living with the parent in question around age 14 was associated with higher parent-child agreement rates, while not living with the parent was associated with a higher probability of the child reporting they did not know their parent's occupation. Similarly, children still living in the parental home at the time of giving their proxy report were predicted to have higher agreement compared to those who were not living at home.

As seen earlier for the bivariate result, the type of parental occupation itself was a very strong predictor of agreement. The predicted probabilities, and the associated 95% confidence intervals are plotted in Figure 1. For fathers' occupation, the predicted probability of children's reports matching their fathers' reports, was highest when the father's reported occupation was *Professionals* at 73% [95% CI: 67-78] or *Technicians & Trades workers* 72% [95% CI 67,77]. For mothers' occupation, *Professionals* again had the highest predicted probability of mother-child agreement (76% [95% CI 72-80]), followed by *Clerical & Administrative workers* at 63% [95% CI 59-67] whereas *Managers* had a notably low match rate, with over half of children predicted to provide a different occupation. For mothers, unlike for fathers, we also see that lower-level occupations tended to have a higher predicted probability of the children reporting they did not know their parents' occupation.

For fathers we are also able to include a variable on unemployment history which shows that, if the father had a history of unemployment this was associated with a significantly higher probability of not knowing the father's occupation at 8% [5-11], compared to 3% [2-4] if there had been no history of unemployment.

There was also a significant association with whether the parent's occupation was derived from current, historical, or reconstructed reports. When the parent's self-report reflected their current occupation, this was associated with higher predicted agreement rates with their children and lower rates of children not knowing, compared to if the occupation in question was a historical one.

The survey experience was also a significant determinant with interviews conducted with children prior to 2009, having higher levels of agreement than ones conducted after 2009 if they were by telephone and CAPI. For responses to mother's occupation, the presence of other adults during the parent interview, or the child interview, was also associated with significantly lower rates of children indicating they did not know their parents occupation. This association was not evident for the father-child agreement where the presence of others was not associated with any significant difference in response patterns.

Sibling agreement in reporting parental occupation was also examined, using logistic regression to predict the probability that a pair of siblings gave matching reports of their father's or mother's occupation. Predicted probabilities are reported in Appendix Table 3.

Sibling reliability

Overall, sibling pairs showed reasonably high concordance. For fathers' occupation, 69% of sibling-pair reports matched ($\kappa = 0.63$), rising to 76% agreement ($\kappa = 0.71$) when cases where the either sibling did not know the parent's occupation or their answer could not be

coded were excluded. Agreement was slightly lower for mothers' occupation, at 62% agreement between sibling pairs ($\kappa = 0.57$), increasing to 71% ($\kappa = 0.66$) after excluding don't know/uncodeable responses.

There were some variations however in how reliable the data on parental occupation was across sibling pairs. The strongest effects were seen, for the variables describing living circumstances around age 14. If both children in the sibling pair had lived at with the target parent at age 14 there was a higher level of agreement between them.

As with the parent-child validity analyses, the parent's occupation type was again a strong predictor of sibling agreement as shown in Figure 2. For fathers' occupation, 69% of sibling-pairs reports matched ($\kappa = 0.63$), rising to 76% agreement ($\kappa = 0.71$) when cases where any of the siblings responding 'don't know' or where their answers not being able to be coded were excluded. Agreement was slightly lower for mothers' occupation, at 62% ($\kappa = 0.57$), increasing to 71% ($\kappa = 0.65$) after excluding don't know/uncodeable responses.

For both fathers and mothers occupation, Professionals had the highest sibling-pair concordance, with predicted probabilities of 78% and 80%, respectively. It is interesting to note that generally the gender gap in agreement levels between mothers and fathers are smaller within particular occupation categories was lower in sibling-pair reports compared to parent-child reports.

Figure 1. Predicted probability of parent and child reporting the same occupation for parent, by parent sex and occupation (as reported by parent)



Note: Model results controlling for other variables in multinomial models. See Appendix Table A1 for full results.

Figure 2. Predicted probability of sibling pair reporting the same occupation for parent, by parent sex and occupation (as reported by one of siblings)



Note: Model results controlling for other variables in multinomial models. See Appendix Table A2 for full results.

Education

Parent-child agreement

Turning to the parent's highest education level, that overall agreement between child and parent reports was 68% for fathers and 60% for mothers. When cases where the child reported '*don't know*' or the parent's education could not be determined were excluded, agreement rose to 75% and 66% respectively. The associated Cohen's kappa figures were 0.56 for fathers and 0.46 for mothers, indicating moderate agreement in both cases, improving to 0.64 and 0.53 when restricted to classifiable responses.

Overall parent-child agreement was highest for university educated parents, and for those with the lowest level of education, of Year 11 or below. For university educated parents, 86% of children reported that their fathers had a university degree, and 78% for their mothers. For fathers, and to a lesser extent for mothers, other post-school qualifications were generally correctly reported by children. However, Year 12 education was frequently misclassified with a relatively high proportion of children not knowing if their parents had a post-school qualification.

The predicted probabilities from the multinomial logistic regression models estimating the likelihood of three outcomes for parental education (child-parent match, child-parent mismatch, and child reporting '*don't know*') for fathers and mothers, are shown in Appendix Table A3, and discussed below.

Table 5 Cross-tabulation of highest education level reported by parent, and education level of parent as reported by the child, column percentages

Father's report						
Child's report	University	Other post-school qual	Year 12	Year 11 or below	Total %	Total N
University	86	7	12	1	25	658
Other post-school qual	9	70	31	23	38	984
Year 12	2	3	39	4	6	152
Year 11 or below	1	11	6	60	22	573
Don't know	3	9	12	13	9	226
Total %	100	100	100	100	100	
Total N	642	991	206	754		2,593
Mother's report						
Child's report	University	Other post-school qual	Year 12	Year 11 or below	Total %	Total N
University	78	12	10	1	19	723
Other post-school qual	17	59	26	22	30	1,147
Year 12	1	7	38	6	9	350
Year 11 or below	1	12	14	61	32	1,208
Don't know	3	9	13	11	9	350
Total %	100	100	100	100	100	
Total N	699	901	478	1,700		3,778

For fathers, the overall predicted probability of a match was approximately 62–69% across most factors, while the probability of mismatch ranged from 20–35% and the probability of not knowing from 3–13%. Several characteristics were associated differences in match probability. Children who were living their fathers at age 14 had the highest predicted match rate for fathers (69%, [95% CI : 67-71]). Similarly, children still living in the parental home at the time of the proxy report were more likely to match the education level reported by their parents, compared to those who had left home.

There is also an association with the parent's country of birth for fathers. If the father was born in a Non-English speaking country, the predicted probability of a mismatch was significantly higher than if they had been born in Australia.

However, as seen in the earlier cross-tabulation, the strongest predictor of parent-child agreement was the parent's own education level. In particular, when the parents had completed Year 12, or had Year 11 or below education this was associated with a higher predicted probability of children not knowing the parent's highest education level; due to the fact that they did not know if their parents had a post-school qualification.

Sibling-pair agreement

In the sibling-pair analysis, overall agreement on father's and mother's education level was relatively high, at 72% and 67% [kappa 0.62 and 0.57], respectively. Sibling agreement in reporting parental education was also examined using logistic regression, predicting the probability that a pair of siblings gave matching reports of their father's or mother's education. Predicted probabilities are shown in Appendix Table A4.

Overall, sibling concordance for education was moderately high. Living circumstances were again an important predictor with sibling pairs who were both living with the target parent when they were 14 more likely to agree regarding their parent's education level. As with occupation, education level itself was the strongest predictor of sibling agreement. If the parent was university educated (as reported by one sibling) this showed the highest concordance (86% for fathers, 83% for mothers), while if the parent was reported by one sibling as having a Year 12 education this had a lower predicted probability of associated agreement (59%; 55%). Year 11 or below education was also relatively reliably reported across siblings, with a predicted probability of 70% agreement for both fathers and mothers.

Discussion

Taking advantage of the design of HILDA as a household survey, collecting information from multiple members of the same family, this study examined the degree to which parents and children, and children from the same family, agreed on two key measures of parental socio-economic status: parental occupation and parental education. By comparing parents own reports of their occupation and education, with the proxy reports provided by their children our aim was to assess the validity of children's reports by calculating the level of agreement, and then to identify what factors, related to the child, the parent, or the survey context were associated with agreement. The purpose of the second part of the analysis, comparing sibling reports, was to investigate the reliability of children's reports.

We find that overall, the reports of parents and children were relatively consistent. In the selected sample, the percentage of children whose proxy report of their parent's occupation matched the parent's own report, at the 1-digit level of the ANZCO level, was 63% for fathers and 56% for mother's occupation respectively. There were some clear associations between certain child and parent characteristics and higher likelihood of having matching reports.

Significantly, the parent's own reported occupation was associated with the probability of agreement. In particular, if parents reported an occupation classified as '*Professionals*' this yielded the highest agreement between parents and children (75% for father-child pairs, 78% for mother-child pairs). As professional roles tend to possess clearer, culturally prominent boundaries and higher internal cohesion, this likely makes them more salient and easy for children to describe (Kim et al., 2020). In contrast, "Manager" occupations were significantly less likely to be reported in agreement, particularly for mothers (only 35% mother-child agreement). Children systematically downgraded their manager mothers to lower-level clerical or professional roles. Conversely, children of fathers in clerical or sales roles tended to inflate their fathers' occupations to '*Manager*.' This asymmetry may be driven by the blurry boundaries of management in occupation classifications rules and gendered understandings of workplaces.

We also find that more contemporaneous reports have higher levels of agreement. Younger proxy-reporters (15–18) who are closer to the age they are asked to recall back to, achieved much higher agreement. We also see that fathers with a history of unemployment were significantly more likely to have children who did not know their occupation (8% vs. 3%). It is likely that the father's job instability would disrupts the child's memory of the parent's work identity, and this missing data may disproportionately excludes reports from disadvantaged families.

Interestingly, having other adults present during the interview significantly reduced "*don't know*" responses, uniquely for mother-child dyads. This could point to third parties helping child respondents to remember their mother's work roles, if maternal occupation is less independently salient in the family's shared narrative.

A major contribution of this study is the analysis of sibling agreement in reporting parental education and occupation. We find that overall sibling concordance is moderately high. For example, sibling agreement on parental university education reached 86% for fathers and 83% for mothers. The high level of sibling agreement could point to the existing of a shared family narrative, or a shared understanding and knowledge regarding their parent's occupation and education levels, among siblings.

For education reports, a notable result was that in the multinomial models if a father was born in a Non-English speaking country (NESC), the probability of a child-parent educational mismatch is significantly higher than if the father was born in Australia. This finding highlights a challenge in surveys in countries, such as Australia, where a high proportion of the population, has been born overseas and where children may find it difficult to translate their parent's overseas qualification to an equivalent level (Lavest et al., 2025). If the parent completed their education abroad, particularly in a non-Anglo-Saxon systems, it may be difficult for the child respondents to translate these foreign credentials into standard Australian educational classifications (like Year 11, Year 12, or Certificate III/IV).

More generally, the analysis of parental educational reports revealed a distinct 'u-shaped' pattern of agreement, where children were more accurate at the extremes of the educational distribution but struggled with middle-range qualifications. In other words, children knew if their parents had a high level of education, or very little, but were less knowledgeable about parents whether or not their parent had a non-university degrees, or what that qualification was.

While this study leveraged the rich information provided by the household structure of the HILDA survey to provide a robust evaluation of proxy-reporting validity and reliability, there are some limitations. Although comparing child proxy-reports to parent self-reports is the standard validation method in sociology, in reality, particularly for occupation, there could be a number of other reasons for a mismatch between parent and child reports. Our results therefore likely represent a conservative estimate, underestimating the true validity of children's proxy reports.

While we assumed in the parent-child agreement analysis that the parent's self-reported occupation or education level, represented the 'true' value, given that both the child's and parent's proxy reports are originally open-ended text descriptions that are translated to a structured ANZCO code, both the parent's and children's reports could contain coding errors, which can be relatively common in occupation coding exercises.

Observed parent-child mismatched in our findings therefore do not necessarily imply the child respondent giving an erroneous answer, and instead could be the mathematical byproduct of:

1. The child's proxy occupation report was described 'incorrectly' (*study assumption*)
2. The parent's occupation was described 'incorrectly'
3. The parent's occupation was described 'correctly' and was miscoded.
4. The child's proxy occupation was described 'correctly' and was miscoded.

5. Both descriptions parent and child descriptions were correctly recalled, but independently miscoded into different categories.

Another limitation of the study is that to our sample consisted of parent-child and sibling-pair dyads with this sample selection relying on parents and their children participating as respondents in the survey, and for the sibling-pair study for at least two full siblings to participate. If characteristics such as parental country of birth, or parental occupation are themselves associated with survey participation, then the selectivity of the sample would limit the generalisability of our findings.

Finally, the findings relate to the specific measures used in the Australian context. For occupation the coding was based on the Australian and New Zealand Standard Classification of Occupations (ANZCO). While ANZSCO is conceptually rigorous and highly comparable to international standards like ISCO-08, any national classification framework possesses unique characteristics which may not be relevant in other national contexts. For example, the specific symmetric confusion seen between certain manual categories (such as 'Machinery Operators and Drivers' versus 'Labourers') may be shaped by local Australian industry structures and the way ANZSCO clusters manual tasks. Consequently, the findings relating to the pattern of agreement across the ANZCO categories might not be evident in other national frameworks (such as the US SOC).

Conclusion

Measuring parental socioeconomic status is a common practice of quantitative social science. Understanding the quality of parental socio-economic status is important as this has direct implications for the large literature on intergenerational mobility and for any study using parental occupation or education as an explanatory variable. By exploiting the household structure of the HILDA survey, this study has demonstrated that child proxy-reports of parental occupation and education are subject to measurement errors. However, the underlying cause of the errors is difficult to disentangle, as they may originate from the children's reports, the parent's reports, or particularly in the case of occupation from the coding process itself.

Overall, the results show that parents and children, and different children from the same family, have a high level of agreement for professional occupations and university education which were reported in agreement by the large majority of children and siblings alike, while mid-range and harder-to-classify categories (Clerical/Administrative and Managerial occupations, Year 12 education) produced substantially more mismatches.

There was some evidence of gender-specific influences. Typically, mothers' occupation and education were consistently harder for children to report accurately than fathers', both in raw agreement and in kappa statistics; this gap was small for occupation but more pronounced for education. There was also some evidence of gendered status-reporting biases where child reports downgraded maternal managerial occupations. Also, of important was that occupational instability (as measured by the father's history of unemployment) disrupted the child's ability to recall or map their parent's occupation when they were 14, driving up "*don't know*" rates and creating non-random missing data that systematically would exclude the most disadvantaged households.

A major longitudinal analysis of the U.S. Current Population Survey (CPS-ASEC) from 1991 to 2020 conducted by Kim and Kim (2026) discovered that occupational coding mismatches have risen significantly over time, with this rise likely driven by changes in agency-level coding

procedures, higher turnover among professional coders, or the increasing complexity of modern job duties (the New Economy) making open-ended text harder to classify consistently. These shifts in both the structure of the labour market and the technological tools used to code it highlight the importance of continuous evaluation of survey data quality.

As traditional manual office coding is increasingly replaced or supplemented by automated algorithms and administrative data linkages, it is important to remain conscious of new possible compounding sources of error, particularly in relation to occupation coding. The fundamental challenge remains bridging the gap between how individuals perceive their work and how institutions classify them (Domarchi et al., 2026), particularly as modern employment patterns evolve.

References

- Aarø, L. E., Flisher, A. J., Kaaya, S., Onya, H., Namisi, F. S., & Wubs, A. (2009). Parental education as an indicator of socioeconomic status: improving quality of data by requiring consistency across measurement occasions. *Scandinavian Journal of Public Health*, 37 Suppl 2(2_suppl), 16–27.
- Andersson, M. A., & Upenieks, L. (2023). How I was brought up? Parental SES and adult work conflict, stress, and satisfaction. *Sociological Spectrum: The Official Journal of the Mid-South Sociological Association*, 43(2–3), 53–71.
- Avvisati, F., & Wuyts, C. (2024). *The measurement of Socio-economic status in PISA* (OECD Education Working Papers). Organisation for Economic Co-Operation and Development (OECD). <https://doi.org/10.1787/0c5b793c-en>
- Belloni, M., Brugiavini, A., Meschi, E., & Tijdens, K. (2016). Measuring and detecting errors in occupational coding: An analysis of SHARE data. *Journal of Official Statistics*, 32(4), 917–945.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American Economic Review*, 95(1), 437–449.
- Caro, D., & Cortés, D. (2012). *Measuring family socioeconomic status : An illustration using data from PIRLS 2006*. https://ierinstitute.org/fileadmin/Documents/IERI_Monograph/Volume_5/IERI_Monograph_Volume_05_Chapter_1.pdf
- Chittleborough, C. R., Taylor, A. W., Baum, F. E., & Hiller, J. E. (2008). Non-response to a life course socioeconomic position indicator in surveillance: comparison of telephone and face-to-face modes. *BMC Medical Research Methodology*, 8(1), 54.
- Conrad, F. G., Couper, M. P., & Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes. *Journal of Official Statistics*, 32(1), 75–92.
- Domarchi, C., Brown, M., Maslovskaya, O., Calderwood, L., & Jessop, C. (2026). *Industry and occupation coding: A comparison of office-based coding and a closed-list approach*. University of Essex. <https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/survey-futures/sf-wp-24-comparison-of-office-based-coding-and-closed-list-approach.pdf>
- Engzell, P., & Jonsson, J. O. (2015). Estimating social and ethnic inequality in school surveys: Biases from child misreporting and parent nonresponse. *European Sociological Review*, 31(3), 312–325.
- Ensminger, M. E., Forrest, C. B., Riley, A. W., Kang, M., Green, B. F., Starfield, B., & Ryan, S. A. (2000). The validity of measures of socioeconomic status of adolescents. *Journal of Adolescent Research*, 15(3), 392–419.
- Ge, C., Friesen, M. C., Locke, S., Rothman, N., Lan, Q., & Vermeulen, R. (2026). Reliability of interrater occupation coding and potential impact on occupational exposure assessment. *Annals of Work Exposures and Health*, 70(5), wxag013.
- Houseworth, C., & Fisher, J. (2020). Measurement error in occupation and the impact on intergenerational mobility. *Research in Social Stratification and Mobility*, 70(100469), 100469.
- Huurre, T., Aro, H., & Rahkonen, O. (2003). Well-being and health behaviour by parental socioeconomic status: a follow-up study of adolescents aged 16 until age 32 years. *Social Psychiatry and Psychiatric Epidemiology*, 38(5), 249–255.
- Jerrim, J., & Micklewright, J. (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.

- Kim, C., Kim, J., & Ban, M. (2020). Do you know what you do for a living? Occupational coding mismatches between coders in the Korean General Social Survey. *Research in Social Stratification and Mobility*, 70(100467), 100467.
- Lavest, C., Ferry, M., Ichou, M., & Präg, P. (2025). Who do they think you are? Inconsistencies in self- and proxy-reports of education within families. *Social Science Research*, 132(103225), 103225.
- Lien, N., Friestad, C., & Klepp, K. I. (2001). Adolescents' proxy reports of parents' socioeconomic status: How valid are they? *Journal of Epidemiology & Community Health*, 55(10), 731–737.
- Long, K., & Renbarger, R. (2023). Persistence of poverty: How measures of socioeconomic status have changed over time. *Educational Researcher (Washington, D.C.: 1972)*, 52(3), 144–154.
- Looker, D. E. (1989). Accuracy of proxy reports of parental status characteristics. *Sociology of Education*, 62(4), 257.
- Martin-Caughey, A. (2026). A similarity index for all occupations. *Research in Social Stratification and Mobility*, 101(101113), 101113.
- Massing, N., Wasmer, M., Wolf, C., & Zuell, C. (2019). How standardized is occupational coding? A comparison of results from different coding agencies in Germany. *Journal of Official Statistics*, 35(1), 167–187.
- Oakes, M. J., & Andrade, K. E. (2017). The measurement of Socioeconomic status. In M. J. Oakes & J. S. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 23–42). Jossey-Bass.
- Pu, C., Huang, N., & Chou, Y.-J. (2011). Do agreements between adolescent and parent reports on family socioeconomic status vary with household financial stress? *BMC Medical Research Methodology*, 11(1), 50.
- Pueyo, M.-J., Serra-Sutton, V., Alonso, J., Starfield, B., & Rajmil, L. (2007). Self-reported social class in adolescents: validity and relationship with gradients in self-reported health. *BMC Health Services Research*, 7(1), 151.
- Svedberg, P., Nygren, J. M., Staland-Nyman, C., & Nyholm, M. (2016). The validity of socioeconomic status measures among adolescents based on self-reported information about parents occupations, FAS and perceived SES; implication for health related quality of life studies. *BMC Medical Research Methodology*, 16(1), 48.
- Ward, M. M. (2011). Concordance of sibling's recall of measures of childhood socioeconomic position. *BMC Medical Research Methodology*, 11(1), 147.
- Watson, N., & Summerfield, M. (2009). *Quality of the Occupation and Industry Coding in the HILDA Survey*. The University of Melbourne.
<https://melbourneinstitute.unimelb.edu.au/assets/documents/hilda-bibliography/hilda-discussion-papers/hdps309.pdf>
- West, P., Sweeting, H., & Speed, E. (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.

Appendix

Table A 1 Multinomial logistic regression of parent-child correspondence on parental occupation- predicted probabilities (P) and 95% Confidence intervals (CI)

		Father-child occupation report						Mother-child occupation report					
		Same		Different		Child does not know		Same		Different		Child does not know	
		P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI
Child gender	Male	61	[58, 64]	34	[31, 37]	5	[3, 6]	56	[53, 58]	35	[33, 38]	9	[7, 10]
	Female	65	[62, 68]	31	[27, 34]	4	[3, 6]	56	[54, 59]	36	[34, 39]	7	[6, 9]
Lived with target parent at age 14	No	58	[50, 66]	34	[25, 42]	8	[4, 12]	45 ^b	[36, 54]	34	[26, 42]	21 ^a	[14, 28]
	Yes	63	[61, 66]	32	[30, 35]	4	[3, 5]	57 ^a	[55, 58]	36	[34, 38]	7 ^b	[6, 8]
Living at home	No	58 ^b	[54, 63]	38 ^a	[33, 43]	4	[2, 5]	53 ^b	[50, 57]	39 ^a	[35, 42]	8	[6, 10]
	Yes	66 ^a	[62, 69]	28 ^b	[25, 32]	6	[4, 8]	59 ^a	[56, 61]	34 ^b	[31, 36]	8	[6, 10]
Age group	15-18	69 ^a	[65, 73]	29 ^b	[25, 33]	2 ^b	[1, 3]	59 ^a	[56, 63]	32 ^b	[29, 35]	8	[6, 10]
	19-24	59 ^b	[54, 64]	34 ^{ab}	[29, 39]	7 ^a	[4, 10]	54 ^b	[50, 58]	37 ^a	[33, 41]	9	[7, 11]
	25+	56 ^b	[51, 62]	37 ^a	[32, 43]	6 ^a	[3, 9]	53 ^b	[49, 57]	40 ^a	[36, 44]	7	[5, 9]
Parent: Country of birth	Australia	65	[62, 67]	31	[29, 34]	4	[3, 5]	56	[54, 58]	36	[34, 38]	8	[7, 10]
	English Speaking	62	[55, 69]	35	[28, 41]	3	[0, 6]	56	[51, 61]	37	[32, 42]	7	[4, 10]
	Other	58	[53, 64]	36	[30, 41]	6	[4, 9]	57	[52, 61]	36	[32, 41]	7	[5, 9]
Parent: Occupation flag	Current job	63	[61, 66]	34	[31, 37]	3 ^b	[2, 3]	60 ^a	[57, 62]	35 ^b	[33, 37]	6 ^b	[4, 7]
	Historical job	63	[58, 68]	29	[25, 34]	8 ^a	[5, 11]	52 ^b	[49, 56]	36 ^{ab}	[33, 39]	11 ^a	[9, 13]
	Previous job	58	[40, 76]	34	[16, 51]	8 ^{ab}	[0, 16]	43 ^b	[33, 52]	46 ^a	[36, 56]	11 ^a	[6, 16]
Parent: Occupation reported [ANZCO]	Never worked							71 ^{ab}	[60, 82]	18 ^{ab}	[9, 27]	12 ^{ab}	[5, 18]
	Managers	60 ^{bc}	[55, 65]	35 ^{bc}	[30, 40]	5	[2, 7]	34 ^e	[29, 40]	59 ^a	[53, 65]	7 ^{ab}	[4, 9]
	Professionals	73 ^a	[67, 78]	24 ^d	[19, 29]	3	[1, 6]	76 ^a	[72, 80]	18 ^d	[15, 22]	6 ^b	[4, 8]
	Technicians & Trades	72 ^a	[67, 77]	23 ^d	[19, 28]	4	[2, 6]	46 ^{cd}	[37, 56]	49 ^{ab}	[40, 58]	4 ^b	[1, 8]
	Community & Personal Serv.	68 ^{ab}	[57, 80]	30 ^{cd}	[18, 41]	2	[1, 4]	48 ^{cd}	[43, 53]	44 ^b	[39, 49]	8 ^{ab}	[5, 10]
	Clerical & Admin.	44 ^d	[34, 54]	51 ^a	[41, 61]	5	[1, 9]	63 ^b	[59, 67]	29 ^c	[25, 32]	8 ^{ab}	[6, 11]
	Sales Workers	47 ^{cd}	[34, 59]	49 ^{ab}	[36, 63]	4	[0, 8]	50 ^{cd}	[44, 56]	40 ^b	[34, 46]	10 ^a	[7, 14]
	Machinery Op. & Drivers	61 ^b	[55, 68]	33 ^c	[27, 40]	5	[2, 8]	39 ^{de}	[30, 49]	51 ^{ab}	[41, 61]	10 ^{ab}	[5, 15]
	Labourers	55 ^{bcd}	[48, 62]	41 ^{abc}	[34, 48]	5	[2, 7]	51 ^c	[46, 55]	41 ^b	[36, 45]	9 ^{ab}	[6, 11]

Table A1 (continued) Multinomial logistic regression of parent-child correspondence on parental occupation- predicted probabilities (P) and 95% Confidence intervals (CI)

		Father-child occupation report						Mother-child occupation report					
		Same		Different		Child does not know		Same		Different		Child does not know	
		P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI	P	95 % CI
Child: Mode of interview	Prior to 2009	62	[58, 66]	33	[29, 37]	5	[3, 6]	58 ^a	[55, 61]	35	[32, 39]	7	[5, 8]
	Face-to-face & CAPI	65	[59, 72]	31	[25, 37]	3	[1, 6]	53 ^a	[47, 58]	35	[30, 40]	12	[7, 17]
	Telephone & CAPI	52	[38, 67]	36	[21, 51]	12	[1, 22]	41 ^b	[29, 53]	42	[30, 54]	17	[5, 29]
Other adults present (child)	Yes	64	[60, 67]	32	[29, 35]	4	[3, 5]	58 ^a	[56, 61]	36	[34, 39]	6 ^b	[5, 7]
	No	63	[59, 66]	33	[30, 36]	5	[3, 6]	54 ^b	[52, 57]	36	[33, 38]	10 ^a	[9, 12]
Other adults present (parent)	Yes	64	[60, 67]	32	[28, 35]	5	[3, 6]	56	[53, 59]	35	[32, 38]	9	[8, 11]
	No	63	[59, 66]	33	[30, 37]	4	[3, 5]	56	[54, 59]	36	[34, 39]	7	[6, 8]
Unemployment history	No	65	[62, 67]	32	[29, 34]	3	[2, 4]						
	Yes	55	[49, 61]	37	[31, 43]	8	[5, 11]						
Number of observations (child-parent pairs)		1,901						3,201					
Number of clusters (parents)		1,432						2,412					
Log-likelihood		-1382.9						-2641.1					
Wald chi2 (44 df)		213.6 (p<0.001)						367.07 (p<0.001)					

Note: Superscript letters indicate results of pairwise comparisons of predicted probabilities within each variable, separately for each outcome (Same/Different/Child does not know) and separately for the father and mother models (Stata margins, pwcompare(pv), following the multinomial logistic regressions). Within a given variable and outcome column, categories that do not share a letter differ significantly at p<.05; categories sharing a letter do not differ significantly. No letters are shown where not pairwise comparison within that variable/outcome reached significance.

Table A 2 Logistic regression of sibling pair correspondence for fathers and mothers occupation - predicted probabilities and 95% confidence intervals

		Father's occupation		Mother's occupation	
		P	95 % CI	P	95 % CI
Age difference between siblings	1-2 years	71	[68, 74]	68	[64, 71]
	3 years	66	[61, 71]	61	[56, 66]
	4 years	71	[64, 78]	55	[48, 63]
	5+ years	65	[59, 70]	54	[48, 59]
Number of siblings in family	2	67	[65, 70]	62	[59, 65]
	3	74	[69, 78]	62	[57, 67]
	4+	65	[55, 76]	62	[52, 72]
Gender	Both male	70	[66, 74]	68	[64, 72]
	Both female	69	[65, 74]	62	[57, 66]
	Mixed	68	[64, 71]	59	[55, 62]
Living at home at time of interview	At least one sibling did not	67	[64, 70]	61	[58, 64]
	Both siblings	71	[68, 75]	65	[60, 69]
Lived with parent at age 14	At least one sibling did not	59	[53, 65]	55	[46, 64]
	Both siblings	71	[69, 74]	63	[60, 65]
Occupation level as reported by younger of siblings [ANZCO]	Never worked			58	[50, 66]
	Managers	74	[69, 80]	56	[45, 67]
	Professionals	78	[73, 83]	80	[76, 85]
	Technicians & Trades	71	[66, 76]	60	[48, 73]
	Community & Personal Service	71	[60, 82]	60	[52, 68]
	Clerical & admin.	62	[50, 74]	71	[65, 77]
	Sales	65	[52, 78]	56	[46, 65]
	Machinery operators & Drivers	71	[63, 78]	37	[21, 53]
	Labourers	59	[50, 67]	66	[58, 73]
	Unable to be determined	40	[31, 50]	39	[31, 46]
	Don't know	63	[51, 74]	38	[25, 51]
Parent country of birth	Australia	68	[66, 71]	61	[58, 64]
	Other English-speaking country	71	[64, 79]	59	[52, 67]
	Non-English speaking country	69	[63, 74]	67	[61, 73]
Sibling pairs (N)		1,453		1,484	
Unique families (N)		1,199		1,232	
Log-pseudolikelihood		-856.3		-900.9	
Wald chi2		93.6 (20 df) p<0.001		144.2 (21df) p<0.001	

Table A 3 Multinomial logistic regression of parent-child correspondence on parent education- predicted probabilities (P) and 95% Confidence intervals (CI)

		Father-child education report						Mother-child education report					
		Same		Different		Child does not know		Same		Different		Child does not know	
		P	95% CI	P	95% CI	P	95% CI	P	95% CI	P	95% CI	P	95% CI
Child gender	Male	69	[66, 71]	22	[20, 25]	9	[7, 10]	59	[56, 61]	31	[29, 33]	11	[9, 12]
	Female	68	[65, 71]	23	[21, 26]	9	[7, 10]	63	[60, 65]	30	[27, 32]	8	[7, 9]
Lived with target parent at age 14	No	55	[48, 63]	32	[24, 39]	13	[8, 18]	59	[50, 67]	21	[13, 28]	20	[13, 27]
	Yes	69	[67, 71]	22	[21, 24]	8	[7, 10]	61	[59, 62]	30	[29, 32]	9	[8, 10]
Living at home	No	64	[61, 67]	26	[23, 29]	10	[8, 12]	59	[57, 62]	32	[29, 34]	9	[7, 11]
	Yes	72	[69, 74]	20	[18, 23]	8	[6, 9]	62	[59, 64]	29	[26, 31]	9	[8, 11]
Age group	15-18	66	[62, 69]	21	[18, 25]	13	[10, 16]	57	[54, 61]	31	[28, 34]	12	[9, 14]
	19-24	69	[66, 72]	22	[20, 25]	8	[7, 10]	61	[59, 64]	30	[27, 32]	9	[7, 10]
	25+	70	[66, 74]	25	[21, 29]	5	[4, 7]	63	[60, 66]	30	[27, 33]	8	[6, 9]
Parent: Country of birth	Australia	70	[68, 73]	21	[19, 23]	9	[7, 10]	61	[59, 63]	30	[28, 32]	9	[8, 10]
	English Speaking	66	[60, 72]	25	[20, 31]	9	[5, 12]	60	[54, 65]	30	[25, 34]	11	[8, 14]
	Other	62	[57, 67]	29	[24, 33]	9	[6, 12]	59	[55, 63]	31	[27, 35]	10	[7, 12]
Child: Mode of interview	Prior to 2009	68	[64, 71]	24	[21, 27]	9	[7, 11]	58	[55, 61]	33	[30, 36]	9	[7, 10]
	Face-to-face & CAPI	70	[64, 77]	22	[16, 28]	8	[5, 11]	65	[60, 70]	25	[20, 30]	10	[7, 13]
	Telephone & CAPI	69	[57, 82]	14	[5, 24]	16	[6, 26]	63	[51, 74]	23	[13, 32]	15	[6, 23]
Other adults present (child)	Yes	72	[69, 75]	22	[19, 25]	6	[5, 8]	66	[63, 69]	28	[25, 30]	6	[5, 8]
	No	66	[64, 69]	23	[21, 26]	10	[9, 12]	57	[55, 59]	32	[30, 34]	11	[10, 13]
Other adults present (parent)	Yes	68	[66, 71]	24	[21, 26]	8	[6, 9]	62	[60, 65]	29	[26, 32]	9	[7, 10]
	No	68	[66, 71]	22	[20, 24]	9	[8, 11]	59	[57, 62]	31	[29, 33]	10	[8, 11]
Parent: Education level	University	86	[83, 89]	11	[9, 14]	2	[1, 4]	79	[75, 82]	19	[15, 22]	3	[1, 4]
	Other post-school qual.	69	[66, 72]	22	[19, 24]	9	[7, 11]	60	[56, 63]	32	[28, 35]	9	[7, 10]
	Year 12	40	[32, 47]	48	[40, 56]	12	[7, 17]	38	[33, 44]	49	[44, 54]	13	[9, 16]
	Year 11 or below	60	[56, 64]	27	[23, 30]	13	[11, 16]	59	[57, 62]	29	[26, 31]	12	[10, 14]
Number of parent-child pairs		1,901						3,778					
Number of parents		1,432						2,661					
Log pseudolikelihood		-1382.88						-3178.18					
Wald chi2		213.55 (p<0.001)						251.59 (p<0.001)					

Table A 4 Sibling-pair correspondence on fathers and mothers education: Predicted probabilities and 95% Confidence intervals from logistic regression

		Father's education		Mother's education	
		Predicted probability	95 % CI	Predicted probability	95 % CI
Age difference between siblings	1-2 years	69	[65, 72]	67	[63, 70]
	3 years	73	[68, 78]	69	[63, 74]
	4 years	77	[71, 83]	70	[63, 77]
	5+ years	73	[68, 78]	64	[58, 70]
Number of siblings in family	2	73	[71, 76]	68	[65, 71]
	3	69	[63, 74]	67	[61, 72]
	4+	63	[54, 73]	63	[51, 75]
Gender	Both male	72	[67, 77]	67	[62, 71]
	Both female	73	[68, 77]	71	[66, 75]
	Mixed	70	[67, 74]	65	[61, 69]
Living at home at time of interview	At least one sibling did not	65	[62, 69]	62	[59, 66]
	Both siblings	81	[78, 84]	75	[71, 79]
Lived with parent at age 14	At least one sibling did not	74	[68, 80]	58	[48, 69]
	Both siblings	71	[68, 74]	68	[65, 70]
Education level as reported by younger of siblings	University	86	[82, 90]	83	[79, 88]
	Other-post school qualification	72	[68, 76]	65	[60, 70]
	Year 12	59	[50, 68]	55	[47, 63]
	Year 11 or below	70	[64, 75]	70	[66, 75]
	Don't know	48	[39, 57]	44	[36, 53]
Parent country of birth	Australia	72	[69, 75]	68	[65, 71]
	Other English-speaking country	68	[60, 75]	61	[53, 69]
	Non-English speaking country	71	[65, 76]	67	[61, 72]
Sibling pairs (N)		1,314		1,385	
Unique families (N)		1,075		1,139	
Log pseudolikelihood		-758.35		-719.52	
Wald chi(12)		47.37 (p<0.001)		108.73 (p<0.001)	