I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11449

Survey Item-Response Behavior as an Imperfect Proxy for Unobserved Ability: Theory and Application

Sonja C. Kassenboehmer
Stefanie Schurer

APRIL 2018

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11449

Survey Item-Response Behavior as an Imperfect Proxy for Unobserved Ability: Theory and Application

Sonja C. Kassenboehmer
Monash University and IZA

Stefanie Schurer
The University of Sydney and IZA

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Survey Item-Response Behavior as an Imperfect Proxy for Unobserved Ability: Theory and Application*

We develop and test an economic model of the cognitive and non-cognitive foundations of survey item-response behavior. We show that a summary measure of response behaviour – the survey item-response rate (SIRR) – varies with cognitive and less so with non-cognitive abilities, has a strong individual fixed component and is predictive of economic outcomes because of its relationship with ability. We demonstrate the usefulness of SIRR, although an imperfect proxy for cognitive ability, to reduce omitted-variable biases in estimated wage returns. We derive both necessary and sufficient conditions under which the use of an imperfect proxy reduces such biases, providing a general guideline for researchers.

JEL Classification: J24, C18, C83, I20, J30

Keywords: survey item-response behavior, imperfect proxy variables, behavioral proxy, cognitive ability, personality traits, selection on unobservables

Corresponding author:
Stefanie Schurer
School of Economics
The University of Sydney
Sydney
Australia
E-mail: stefanie.schurer@sydney.edu.au

* A previous version of this manuscript was published under the title "Testing the Validity of Item Non-Response as a Proxy for Cognitive and Non-Cognitive Skills" published as IZA Discussion Paper Nr. 8874 (2015). The authors acknowledge financial support from an Australian Research Council Early Career Discovery Program Grant (DE140100463) and the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027). This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, which is a project initiated and funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research. The findings and views reported in this paper, however, are those of the authors and should not be attributed to either FaHCSIA or the Melbourne Institute. The authors would like to thank participants of the University of Sydney Microeconometrics and Public Policy Seminar, Monash University Workshop on Microeconometrics and Panel Data Analysis, and Colin Cameron, Denzil Fiebig, Cheng Hsiao, David Johnston, Frauke Kreuter, Zhuan Pei, Pravin Trivedi, Peter Siminski, Agne Suziedelyte, Joachim Winter, Mark Wooden, Kenneth Wolpin, and Jeffrey Wooldridge for helpful comments. Special thanks go to Felix Leung for his research assistance during the early stages of the manuscript.

1 Introduction

Survey methodologists have recognized that survey item-response is closely related to respondents' individual attributes (Beatty and Herrmann, 2002; Dillman et al., 2002). Previous studies have highlighted cognitive abilities (e.g. Tourangeau, 2003) and personality traits (Weijters et al., 2010; Wetzel et al., 2016) as key attributes that shape response patterns.[1] The cognitive model of survey item-response behavior concedes that individuals consider question complexity and relative costs before responding to survey questions (Sudman et al., 1996; Schwarz and Oyserman, 2001). Better cognition, diligence, and willingness to cooperate are likely to reduce the (perceived) cost of responding to even highly complex questions, which in turn increases the probability of response

Economists acknowledge cognitive and non-cognitive abilities[2] as two distinct components of adult human capital that can be fostered through education, but that are hard to measure (Schurer, 2017; Kautz et al., 2014; Heckman and Kautz, 2014; Almlund et al., 2011). In this study we ask whether individuals' survey response behavior may provide insights into their abilities. We suggest that the survey item response rate (SIRR) – which is calculated as the fraction of survey items responded to by each individual – could be a convenient proxy for often unobserved or poorly-measured cognitive and non-cognitive ability. Its appeal lies in its simplicity; variation in item response rates are a side-product of every survey. SIRR could prove useful when included as an additional control variable in data applications that are plagued by unobserved confounders.

Previous studies have exploited such paradata – data captured during the process of producing a survey statistic (Kreuter, 2013) – to fix estimation biases due to selection or heaping behavior (Pudney, 2014; Heffetz and Rabin, 2013; Kleinjans and van Soest, 2014; Behaghel et al., 2015). However, there is little empirical evidence on the link between SIRR and standard measures of cognitive and non-cognitive abilities, nor on the potential of SIRR to address ability-related estimation biases.[3]

[1] Other factors that have been suggested are the survey mode, interviewer characteristics, question topic and structure, question difficulty, institutional policies (Dillman et al., 2002).

[2] In this study we use the term non-cognitive abilities interchangeably for personality traits; other terms commonly used in the literature are character traits, soft skills, life skills, or non-cognitive skills.

[3] The economics literature covered in depth survey measurement error, recall bias and unit non-response, but has paid little attention to item non-response with the exception of e.g. Riphahn and Serfling (2005) and Raessler and Riphahn (2006). Most of the research in this area investigates the determinants of non-response to sensitive questions

To guide our empirical investigation, we first develop an economic model of the cognitive and non-cognitive foundations of survey item-response behavior. This theoretical model stipulates that cognitive and non-cognitive abilities affect both the marginal costs and benefits of responding to a survey question, and thus influence the optimal number of survey questions responded to in equilibrium. For instance, higher levels of cognitive ability increase the opportunity cost of each question, reducing the number of questions responded to. Yet, they also decrease the perceived psychic cost of responding by reducing the time needed to answer each question, thus increasing the number of questions responded to. The model predicts that if the decrease in psychic cost outweighs the increase in opportunity cost, we will observe a positive relationship between cognitive ability and survey-item response. Similar predictions are derived for non-cognitive ability.

We test the predictions of the model using nationally-representative survey data from Australia. The advantage of these data is that skills and survey response behaviors were measured by two separate components of the survey; an interviewer-assisted questionnaire with high and non-variable item-response rates, and a self-completion questionnaire characterized by lower and more variable item-response rates. We demonstrate a strong relationship between SIRR calculated from the self-completion questionnaire and standard task-based, interviewer-administered measures of cognitive ability, ceteris paribus. From a statistical perspective, SIRR behaves similarly to cognitive ability: it has a strong individual-fixed component, and is predictive of economic outcomes. This interpretation provides evidence in support of the model, suggesting that the reduced psychic costs produced by higher cognitive ability outweigh the increased opportunity costs. In contrast, SIRR is only weakly associated with non-cognitive skills, suggesting that the combined benefits and reduced psychic costs resulting from an individual's personality traits do not outweigh the increased opportunity costs.

Building on our empirical findings, we argue that SIRR – although likely to be imperfect –

(wages, income, wealth) or adjusts the estimates of the determinants of wages, income or wealth equations to it (Zweimueller, 1992; Riphahn and Serfling, 2005; Bollinger and Hirsch, 2013; Bollinger and David, 2005). Some recent exceptions are Hedengren and Stratmann (2013) and Hitt and Trivitt (2013) who tested explicitly the hypothesis that personality traits are associated with item non-response using US and German longitudinal data. Both studies show evidence of a significant association between conscientiousness, a key component of the Big-Five personality traits, and item non-response. Zamarro et al. (2016) use data from the Understanding America Study, a nationally representative internet panel, to study the validity of innovative measures of character skills based on measures of survey effort. A recent study by Hitt et al. (2016) demonstrates that survey effort of youth is highly predictive of later-life educational attainment, using data from six nationally representative data sets. The authors judge that survey effort is a behavioral measure for conscientiousness.

could be used as a proxy variable for cognitive ability to reduce omitted-variable biases (OVB). To do so, we first present a conceptual framework and derive the necessary and sufficient conditions under which imperfect proxy variables reduce OVB. We illustrate how these conditions can be used as a guideline by researchers for evaluating the risks of employing an imperfect proxy variable approach, a tool of general applicability. This conceptual framework is then applied to a standard wage regression model, the most widely used application by economists to illustrate methodological points,[4] to quantify the bias-reduction potential of SIRR as a proxy for unobserved cognitive ability. Because we observe both the proxy and cognitive ability, we can calculate the exact bias reductions in the returns to education and locus of control, a widely studied non-cognitive skill (see Cobb-Clark, 2015). We demonstrate that SIRR is indeed an imperfect proxy, but it reduces OVB in the wage returns to education and locus of control by roughly 10%. Item-response behaviors for the most cognitively demanding questions, such as those on computer use, time diaries, and household expenditures, are most powerful for bias reduction.

Our study contributes to the human capital and statistical literature in three important ways. First, we complement recent work on identifying task-based or observational measures of ability. This emerging literature raises the question of whether standard instruments used to measure cognitive or non-cognitive ability yield what researchers intend to capture (see Lundberg (forthcoming), Borghans et al. (2016), and Almlund et al. (2011) for a discussion of these issues). Behavioral measures have been used in the literature to circumvent measurement problems or unavailability of standard instruments. Many studies use administrative records on student behavior to proxy non-cognitive abilities including school attendance rates, number of suspensions (Jackson, 2017; Dee and West, 2011; Holmlund and Silva, 2014; West et al., 2016), participation in extracurricular activities (Lleras, 2008), or behavior in class (Heckman et al., 2013). To proxy

[4]Wage regression models are the textbook example to illustrate OVB (see Angrist and Pischke, 2009, on the returns to education). For recent applications, see Pei et al. (2018) and Oster (2017). In the context of the Theory of Human Capital, unobserved cognitive ability is one of the most-often cited factors that cause OVB. The discussion on the confounding role of unobserved ability emerged in the early 1960s when economists debated whether education boosts skills and productivity or whether self-selection into education by ability is the reason for the high returns to education found empirically (see Gronau, 2010, for a review). Concerns over OVB due to unobserved ability are still dominating the empirical literature, giving rise to ever-more innovative methods – that entail their own limitations – to identify a causal effect of education on earnings (e.g. Blackburn and Neumark, 1993; Ashenfelter and Rouse, 1998; Card, 2001; Oreopoulos, 2006; Heckman et al., 2017), health outcomes (Chatterji, 2014; Heckman et al., 2017) and health behaviors such as smoking (Heckman et al., 2017). A similar discussion has emerged in the literature that estimates the causal impact of non-cognitive skills on economic outcomes (see Almlund et al., 2011, for an overview).

conscientiousness, Hitt et al. (2016) use information on survey effort, while Heine et al. (2008) use information on walking speed and the accuracy of clocks. Kautz et al. (2014) suggest that "performance on any task or any observed behavior can be used to measure personality and other skills" (p. 16), and conclude that as long as such a measure predicts behavior and can be implemented in practice, it is useful. With the exception of Hitt et al. (2016) and West et al. (2016), none of these studies have validated their behavioral proxies.

We also contribute to the dormant and predominantly theoretical literature on proxy variable biases (Wickens, 1972; McCallum, 1972; Aigner, 1974; Frost, 1979). We extend this theoretical literature by showing that a strong proxy is neither a necessary nor a sufficient condition for reducing OVB. The necessary condition requires sign equivalence of two key correlation coefficients: the partial correlation between the missing variable and the proxy (strength of the proxy), and the partial correlation between the missing variable and the variable of interest (degree of OVB). The sufficient condition bounds the ratio of the strength of the proxy to the degree of OVB – a term we refer to as the relative strength of the proxy – by terms that only depend on the degree of multicollinearity introduced into the model by the proxy, which is always observable.

There has been a paucity in empirical research testing the validity of proxy variables even though proxies are widely used in microeconomic applications. Yet, testing is important, because an imperfect proxy variable – that is, a proxy that correlates with the key variable from the regression model – may exacerbate biases (see Frost, 1979; Wolpin, 1995, for a formal discussion).[5] Practical guidelines for such testing are non-existent. Todd and Wolpin (2003) rightly suggested that "in some sense, the problem of whether or not to include proxy variables is insoluble, because it involves a comparison between two unknown biases" (p. F15). Nevertheless, we offer a solution to this theoretical dilemma – a practical guideline which can be used to evaluate the potential estimation biases resulting from an imperfect proxy.

Third, we contribute to a broader literature on causal inference in the presence of unobserved confounders (Rosenbaum and Rubin, 1983). Recent work by Oster (2017) and Altonji et al. (2005)

[5] Apart from task-based behavioral proxies which we consider in this study and that have been used in previous research as discussed above, proxies are also widely used as inputs in the educational production function, such as family income to compensate for family investments into children (Todd and Wolpin, 2003). Proxies have also been used to measure unobserved wage offer distributions in a hazard model of unemployment durations, a choice criticized by Wolpin (1995) who showed in detail the implications of an imperfect, or what he refers to as a 'crude' proxy variable in the context of a job-search model.

provide bounds on estimation biases that depend on assumptions about the maximal possible degree of explained variation in the outcome of interest and the relative importance of unobservable over observable selection into treatment.[6] Our approach requires different and arguably less restrictive assumptions if information on the nature of the unobserved confounder is available. Comparing the bias-reduction potential of the proxy variable approach against the methods proposed in Oster (2017), we conclude that in certain situations the proxy variable approach is a good alternative.

Our work is closely related to Pei et al. (2018), who demonstrate that in the presence of poorly measured confounders, an alternative empirical tool kit should be used than the "coefficient-comparison test" to test the identification assumption of a regression strategy. Although our approach is also centered on the problem of poorly measured right-hand-side control variables, we focus on a different implication regarding this concern. Pei et al. (2018) highlight that a sensitivity check with poorly-measured additional control variables may lead to the erroneous conclusion that the identification assumptions of the model hold, misleading researchers into believing that OVB play no role. We emphasize, in contrast, that adding a poorly measured right-hand-side variable may exacerbate OVB.

The remainder of this paper is structured as follows. In Section 2 we lay out an economic model on the cognitive and non-cognitive foundations of item-response behavior. In Section 3 we describe the Household, Income, and Labor Dynamics in Australia (HILDA) survey which we use for testing the predictions of our economic model. In Section 4 we derive necessary and sufficient conditions of a valid but imperfect proxy variable, and translate these into an empirical guideline. Section 5 presents test results of using SIRR as a proxy for cognitive ability in the context of wage returns to education and locus of control. Section 6 concludes. An Online Appendix provides supplementary material.

[6]This literature extends previous approaches that explicitly model unobservable selection into treatment (e.g. Murphy and Topel, 1990; Imbens, 2003) or which vary the set of control variables as sensitivity analyses (Heckman and Hotz, 1989; Dehejia and Wahba, 1999; Gelbach, 2016).

2 An economic model of survey item-response behavior

Why is survey item-response behavior related to cognitive ability and personality traits? The following, brief theoretical framework highlights the relevant behavioral channels and motivate our empirical analysis. We base our theoretical model on Dillman (2000)'s theory on survey participation. This model originates from social exchange theory developed by Homans (1961) and Blau (1964). It assumes that people engage in social interactions because they receive something in return. Individual i replies to question j ($R_{ij} = 1$) if the expected benefits B_{ij} outweigh the expected costs C_{ij}. $E[.]$ indicates expectations:

$$\begin{aligned} R_{ij} &= 1 \text{ if } E[B_{ij}(P_i, FI_i) - C_{ij}(O_{ij}, S_{ij})] > 0, \\ &= 0 \text{ otherwise.} \end{aligned} \tag{1}$$

The expected benefits are a function of personality trait P_i and a financial incentive FI_i if paid to complete the survey as a compensation for lost time. Individuals are more likely to respond to the question, if they are paid a dollar amount for each response. We therefore assume that $\frac{\partial B_{ij}}{\partial FI_{ij}} > 0$, independent of the individual's personality or cognitive ability.[7] Furthermore, personality may affect survey response because some individuals want to contribute to a public good (e.g. they are agreeable or conscientious), or because they derive pleasure from talking about themselves (see Singer and Ye, 2013). We assume that the marginal benefits due to personality trait P are $MB^P = \frac{\partial B_{ij}}{\partial P_i} > 0$.

Survey item response on each question j comes at a cost to the individual. Expected costs $E(C_{ij})$ include both opportunity costs O_{ij} and psychic costs S_{ij}:

$$E(C_{ij}) = E(O_{ij}(W_i(A_i, P_i)) + S_{ij}(T(A_i, P_i))). \tag{2}$$

We assume that the highest-valued opportunity foregone due to the time spent completing the survey is working, which can be quantified by wages W_i. Wages depend on both cognitive

[7] In our empirical example there are no financial incentives paid for completing the self-completion questionnaire, and thus we abstract from financial incentives as determinants of survey item-respose. As an extension to this model, one could argue that the marginal response to financial incentives differs by cognitive ability or personality.

ability A_i and personality trait P_i.[8] Therefore, we assume that the marginal opportunity costs due to cognitive ability are $MC|_O^A = \frac{\partial O_{ij}}{\partial W_i}\frac{\partial W_i}{\partial A_i} > 0$ and due to personality traits are $MC|_O^P = \frac{\partial O_{ij}}{\partial W_i}\frac{\partial W_i}{\partial P_i} > 0$.[9]

Answering survey questions is also mentally demanding and therefore places psychic costs on individuals for the time T_{ij} they have to endure the full survey. These psychic costs depend on the (perceived) time it takes to answer each question. The total time needed to respond is shorter for individuals with higher levels of cognitive ability A and personality trait P. Time to answer is reduced by cognitive ability because an individual has to understand the meaning of the question, recall relevant behavior and information, infer the appropriate answer and map the answer into the response format of the survey (Schwarz and Oyserman, 2001). The higher cognitive ability A is, the lower the mental demands associated with answering a given question. We therefore assume that the marginal psychic costs due to cognitive ability are $MC|_S^A = \frac{\partial S_{ij}}{\partial T_{ij}}\frac{\partial T_{ij}}{\partial A_i} < 0$.

Furthermore, replying to a question may entail revealing information about personal life and carefully and diligently reading and replying to each question (Singer and Presser, 2007). Therefore, the psychic costs will be lower for individuals with certain personality traits, such as openness to experience, extraversion and conscientiousness, because the respondent derives less disutility from every minute spent on answering the question. Such attributes allow individuals to more easily engage with an interviewer, overcome privacy and confidentiality concerns, and patiently and diligently go through a questionnaire. We therefore assume that the marginal psychic costs due to personality trait P are $MC|_S^P = \frac{\partial S_{ij}}{\partial T_{ij}}\frac{\partial T_{ij}}{\partial P_i} < 0$.

The model highlights that the decision to reply to a specific question depends on the trade-offs between perceived costs and benefits, and the trade-offs between taking less time to respond but facing greater opportunity costs for time spent on the survey. In the optimum, the individual will respond to the question when the marginal benefits due to increases in Ability A_i are equal to or greater than the marginal costs due to increases in Ability A_i:

[8] Alternatively, one could assume that opportunity costs could refer to time lost for studying or time spent with children. We will acknowledge this possibility in our empirical specification.

[9] There is sufficient evidence in the literature that wages depend on both cognitive and non-cognitive abilities. See Almlund et al. (2011) for a review of the literature and Elkins et al. (2017) for more recent evidence.

$$MB^A \geqslant MC|_O^A - MC|_S^A. \tag{3}$$

Since $MB^A = 0$, it must follow that $MC|_S^A \geqslant MC|_O^A$. Thus, as long as the marginal psychic cost reductions due to cognitive ability are at least as large as the increases in marginal opportunity costs, then the individual will respond to question j.

In the optimum, the individual will also respond to the question if the marginal benefits are equal to or greater than the marginal costs due to changes in personality trait P_i:[10]

$$MB^P \geqslant MC|_O^P - MC|_S^P, \tag{4}$$

$$MB^P + MC|_S^P \geqslant MC|_O^P. \tag{5}$$

Since $MB^P > 0$, $MC|_O^P > 0$ and $MC|_S^P < 0$ this condition implies that the individual decides to respond to question j if the combined increases in marginal benefits and reductions in psychic costs outweigh the increases in marginal opportunity costs due to personality trait P_i.

In the next two subsections we illustrate implications of the economic model's derivations for the aggregate level, leading to empirically testable predictions.

2.1 Model predictions for cognitive ability

Figure 1 summarizes graphically what happens in the aggregate – over the full duration of the survey – for individuals of different cognitive and non-cognitive abilities. The horizontal axis describes the number of survey questions responded to, where Q_{max} is the maximum number of survey questions (complete survey), and Q_{min} the minimum number of questions necessary to be counted as survey participant. The vertical axis describes the implicit "price" which needs

[10]Note, we are assuming additive separability in the inputs personality trait P and cognitive ability A. This may be a strong assumption, yet we argue that the predictions of the model will be the same, but stronger, if we allowed for interaction effects between personality and cognitive ability. Since the marginal cost responses of personality and cognitive ability are moving into the same direction, we will be facing the same tradeoffs when exploring the case of higher or lower levels of both personality trait P and cognitive ability. The only difference will be if allowing for high levels of cognitive ability and low levels of personality trait P, or vice versa. We leave this special case for future research.

to be paid to the participant for responding to every additional question. The marginal cost (MC) curve is upward sloping, and the marginal benefit (MB) curve is downward sloping, assuming that the benefit for every additional question declines and the cost for every additional survey question increases. Where marginal cost and benefit curves intersect indicates the total number of survey questions responded to by the participant (Q_1^*).

Figure 1(a) shows that cognitive ability affects only the slope of the marginal cost curve. On the one hand, because of its negative impact on psychic costs, it makes the marginal cost curve flatter – a move from $MC1$ to $MC2$ – and thus the optimal number of questions answered is larger (Q_2^*). On the other hand, because of its positive impact on opportunity costs, it makes the marginal cost curve steeper – a move from $MC1$ to $MC3$ – and thus the optimal number of survey questions answered is smaller (Q_3^*).

This model allows for three different predictions. Cognitive ability may be *positively, negatively*, or *not at all* associated with survey item response, depending on the dominance of the psychic cost response over the opportunity cost response. Empirically, we can test for this relationship by estimating a reduced-form OLS model, in which SIRR is regressed on a measure of cognitive ability and other controls. A positive relationship is evidence that psychic cost responses outweigh opportunity cost responses of cognitive ability ($MC2$); a negative relationship is evidence for the reverse ($MC3$); and no association is evidence that psychic cost responses neutralise opportunity cost responses.

2.2 Model predictions for personality

Figure 1(b) shows that personality affects the slope of both the marginal cost and the marginal benefit curves. Similar to cognitive ability, personality trait P steepens the marginal cost curve – a move from $MC1$ to $MC3$ – but also flattens the psychic cost curve, a move from $MC1$ to $MC2$. Personality trait P also flattens the marginal benefit curve – a move from $MB1$ to $MB2$ – so that for every additional survey question, the associated reduction in marginal benefits declines. For instance, a survey participant with high levels of personality trait P ends up responding to all possible questions (Q_{max}), under the assumption that psychic costs outweigh opportunity costs.

Again, the model allows for three predictions. Personality may be *positively, negatively*, or *not at all* associated with survey item response, depending on the dominance of the opportunity cost response over the *combined* psychic cost and benefit responses. Empirically, we can test for this relationship by estimating a reduced-form OLS model, in which SIRR is regressed on a measure of personality and other controls. A positive relationship is evidence that combined psychic cost and benefit responses outweigh opportunity cost responses of personality (MC2); a negative relationship is evidence for the reverse (MC3); and no association is evidence that combined psychic cost and benefit responses neutralise opportunity cost responses.

In the following section we test the economic model of survey item-response behavior before we consider using SIRR as a proxy variable for unobservable confounding ability variables in wage regressions.

3 Testing the economic model of survey item-response behavior

In this section we outline the data we use in the analysis before testing the predictions of the theoretical model and provide further evidence on whether SIRR behaves in a statistically similar way to abilities.

3.1 Data: Household, Income and Labour Dynamics in Australia Survey

We use data from the HILDA survey which is a nationally representative household panel study conducted annually since 2001 (Summerfield, 2010). The survey comprises a household questionnaire, a person questionnaire for each household member, and a self-completion questionnaire. All adult household members aged 15 years and above are invited to respond to an interviewer-assisted (continuing or new-person) questionnaire, in which information on education, employment, or family formation is collected. In addition, each eligible household member is invited to complete a self-completion questionnaire (SCQ) to be filled out in private. This SCQ collects predominantly attitudinal questions. The interviewer collects the completed SCQs at a later date or, if a date cannot be arranged, the household is asked to return the SCQ by mail. Some household members opt to return a completed SCQ before the interviewer conducts the face-to-face interview.[11]

[11]The household form takes about 10 minutes to complete while each person questionnaire takes about 35 minutes. The SCQ takes also about 30 minutes to complete. HILDA pays a financial incentive of 30$ for each completed person questionnaire and a bonus of another 30$ is paid to the household if all eligible household members complete the survey. No financial incentive is paid for the completion of the SCQ. Household and person questionnaires are

For the analysis, we selected a sample of eligible survey participants from Wave 12 (2012), because this was the year when cognitive ability measures were collected. Item-non response is calculated from the SCQ collected in Wave 12. Personality traits were not collected in Wave 12, but in the SCQs in previous years. We attach to Wave 12 information on personality assessments from waves 5 and 9 and locus of control from waves 3, 4, 7, and 11.

In 2012, 13,195 eligible adults were interviewed.[12] Less than 8% of this eligible sample failed to participate in the cognitive ability assessment, reducing the sample to 12,139 individuals. Around 9% of these individuals did not return a SCQ in Wave 12, reducing our sample further to 11,046 individuals. Finally, another 21% of these individuals did not have information on their personality traits or failed to return a SCQ in previous waves, further reducing our estimation sample to 8,763 individuals. Due to missing information in some control variables – as listed in Table 1 – our sample declined by another 1% (8,666 individuals).

3.1.1 Survey item-response behavior

We construct a measure that summarizes each participant's survey item-response behavior on the SCQ in wave 12 by calculating the survey item response rate (SIRR).[13] SIRR is calculated for each individual as the ratio of the number of answered questions to the total number of questions that the respondent was required to respond to. Therefore, the denominator varies across individuals as some participants are asked more questions than others depending on their socio-demographic situation. The number of responses in the SCQ is calculated as the number of times an individual responded to the question instead of refusing to answer. For a small sub-sample of questions, respondents could choose an option "Don't Know". Some argue that refusals and "Don't Know" answers are determined by different mechanisms (e.g. Raessler and Riphahn, 2006), however in our survey this refers to only two set of questions (neighbourhood characteristics, employer entitlements) comprising 6% of all questions (see Table A.1). We therefore conduct the analysis including these questions, but show in robustness checks that our findings are not sensitive to

collected through CAPI. In total there are around 145 interviewers each of which conduct 145 face-to-face interviews. (see Watson, 2011).

[12] Technically, 17,162 adults completed an interview in Wave 12, because a top-up sample of 4,280 eligible adults was added in 2011. However, we cannot use this top-up sample because we do not have Big-Five personality data for this group, which was collected in the years before the top-up sample was added.

[13] We considered calculating an item response proxy also from the interviewer-assisted continuing/new-person questionnaire, but there was too little variation in the response rates to make it a useful proxy variable.

their inclusion.

Figure 2(a) shows that the total range of applicable questions varies from 190 to 266 questions, and there is heaping of required responses with 12 dominant clusters. The most frequent number of required questions are 216 (8.5%), 244 (14%), and 265 (11%). Figure 2(b) shows that the total number of unanswered items in Wave 12 varies between 0 (33%) and 192. The mean non-response count is 4.8 questions. About 90% of the sample failed to respond to up to 10 items. About 1% of the sample refused to answer 50 or more questions. A 1 standard-deviation (SD) increase in item non-response is equivalent to 11 additional questions not responded to in the SCQ. Figure 2(c) shows that on average individuals respond to 98% of the questions, and the minimum response is 15%.

3.1.2 Cognitive ability

The HILDA survey assessed respondents' cognitive ability in Wave 12 as part of the interviewer-assisted survey. This assessment included standard tests to measure memory, executive function, and crystallized intelligence through a Backward-Digit Span Test (BDS), a Symbol-Digit Modalities Test (SDM), and a National Adult Reading Test (NART), respectively (see Wooden, 2013, for an overview). The BDS measures working memory span and is a sub-component of traditional intelligence tests. The interviewer reads out a string of digits which the respondent has to repeat in reverse order. BDS measures the number of correctly remembered sequences of numbers. SDM is a test of executive function, which was originally developed to detect cerebral dysfunction but is now a recognized test for divided attention, visual scanning and motor speed. Respondents have to match symbols to numbers according to a printed key that is given to them. SDM measures the number of correctly matched symbol-number pairs. NART is assessed through a 25-item list of irregular English words, which the respondents are asked to read out loud and pronounce correctly. NART measures the number of correctly pronounced words. On average, sample members score 4 on the BDS, 49 on the SDM, and 14 on the NART tests. Because the range of possible values differs across these three measures, we standardize each measure to mean 0 and SD 1.

3.1.3 Non-cognitive ability

We measure respondents' non-cognitive ability with the Big Five personality traits and locus of control. In waves 5 and 9, HILDA collected an inventory of the Big-Five personality traits based on Saucier (1994) that can be used to construct measures for extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Out of these five, we would expect agreeableness and conscientiousness to be most closely related to survey response behavior because they best capture willingness to cooperate and diligence with tasks. To construct a summary measure for each trait, we use the 28 items used to measure personality on the Big-5 and conduct factor analysis (see Cobb-Clark and Schurer, 2012).

A measure of internal locus of control is derived from seven available items from the Psychological Coping Resources Component of the Mastery Module developed by Pearlin and Schooler (1978) collected in waves 3, 4, 7, and 11. Mastery refers to the extent to which an individual believes that outcomes in life are under her own control. Respondents were asked to report the extent to which they agree with each of seven statements related to the perception of control and the importance of fate (see Table 1). We construct a continuous measure increasing in internal locus of control using factor analysis (see Cobb-Clark and Schurer, 2013; Cobb-Clark et al., 2014).

To minimize measurement error in our constructs of NCS, we follow Cobb-Clark et al. (2014) by averaging the scores for each of the Big-Five personality traits from 2005 and 2009, and the scores for locus of control from 2003, 2004, 2007, and 2011. All personality variables are standardized to mean 0 and standard deviation 1.[14] Table 1 reports summary statistics.

[14]Our estimation results are not sensitive to whether we use time-averaged measures of non-cognitive skills or measures from one particular year. As discussed in Cobb-Clark and Schurer (2012) and Cobb-Clark and Schurer (2013), the Big-Five personality traits and locus of control are relatively stable in adulthood. Their analyses show that small variations over time can be attributed to measurement error and that past measures of non-cognitive skills can yield attenuation biases. Averaging across time reduces the influence of extreme but non-representative variations in any particular year. For similar conclusions in the context of a model of health behavior, see Cobb-Clark et al. (2014). The attenuation bias on the estimated coefficients of interest depends on the variance in the measurement error (difference between true and predicted factor). One could use a method proposed in Croon (2002) - which was applied, for instance, in Gensowski (2018) - to adjust for large measurement errors. This adjustment becomes more important when Cronbach's alpha, a measure of internal consistency, of the respective skill measure is low, i.e. below 0.7. In our sample, Cronbach's alpha of all non-cognitive skill measures are beyond 0.7 and some lie even beyond 0.8 such as locus of control, conscientiousness or openness to experience.

3.2 Test results

3.2.1 Association between SIRR and abilities

In this section we test the predictions of the economic model of the cognitive and non-cognitive foundations of survey-item response behavior. For this purpose, we first estimate for each cognitive and non-cognitive ability measure a model in which SIRR is the dependent variable, and the specific cognitive or non-cognitive ability measures are the independent variables. We then re-estimate the same models with a standard set of control variables (gender, age, education, language background and geographic remoteness), time availability characteristics (out of the labor force, number of children under the age of 14), and interviewer fixed effects.

Figure 3 summarizes our key findings. It plots the linear and non-linear relationships between SIRR (vertical axis) and nine distinct ability measures (horizontal axis). The fitted solid line displays the OLS estimate of the *adjusted* correlation coefficient, the white dashed line plots the non-parametric kernel estimates, to allow for non-linearities, and their 95% confidence intervals (See e.g. Wand and Jones, 1995). The nine figures demonstrate that there is a positive, significant association between SIRR and the three cognitive ability measures (Figures 3(a) to 3(c)) and a positive but weak association between SIRR and four out of the six non-cognitive ability measures (Figures 3(e), 3(f), 3(g) and 3(i)). Most notable is the association between SIRR and the Symbol Digits Modalities (SDM) measure. The raw and adjusted correlation coefficients between SIRR and the SDM measure ranges between 0.22 SD and 0.19 SD. The second largest association is between SIRR and the National Adult Reading Test (NART) measure (0.14 SD). In contrast, none of the adjusted correlation coefficients for non-cognitive abilities are greater than 0.06 SD. These conclusions do not change when including in each regression model all skill measures simultaneously.[15]

Variation in the Symbol Digits Modalities test explains 5% of the total variation in SIRR, while none of the non-cognitive ability measures explain more than 1%. To understand the magnitude of the explanatory power of SDM, it is worthwhile to compare it against the explanatory power

[15] Full estimation results – including all ability measures simultaneously – are reported in Table B.1 (Appendix). To better understand to what extent parameter estimates are sensitive to adding different set of control variables, we add subsequently three blocks of control variables: individual characteristics, interviewer rating of the understanding of survey questions, and interviewer fixed effects.

of interviewer characteristics, which is considered to be one of the main determinants of survey participation. Interviewer fixed effects explain about 2% of the variation in SIRR. Education, another important predictor of SIRR, explains about 0.6%. Overall, the explanatory power of the SDM measure equates to 40% of the maximum explained variation that can be achieved by including all ability measures, control variables, and interviewer ratings of the understanding of the questions (12.4%, column (4) in Table B.1).

SIRR is significantly related to the interviewer's rating of the survey participant's understanding of the questions (Models (3) and (4), Table B.1) and controlling for this variable reduces the association between SIRR and two of the three cognitive ability measures, which is additional evidence that SIRR captures variation in cognitive ability.[16]

In summary, these findings are evidence for the cognitive, and less so for the non-cognitive, foundations of survey response behavior. Because of a significant positive association between cognitive ability and SIRR, we conclude that the decreased psychic costs *outweigh* the increased opportunity costs as a response to increased cognitive ability. In contrast, because of a marginally positive but weak association between personality (openness to experience, agreeableness, conscientiousness or locus of control) and SIRR, we conclude that the combined increased benefit and reduced psychic costs *barely outweigh* the increase in opportunity costs. Because none of the estimated coefficients is negative, we reject the hypothesis that opportunity costs play an important role in survey response behavior. We come back to this point in Section 3.2.2.

Although Figures 3(a) to 3(c) reveal some non-linearities in the relationship between SIRR and the SDM measure, it is predominantly linear between very low and medium survey item-response rates. In the following, we therefore focus on a linear measure of SIRR. However, we also replicate our key findings when using a non-linear measure of SIRR in Section 4. Our results are also robust to a different definition of non-response based on refusals only (excluding "Don't know"-answers). Furthermore, for simplicity, in the following we focus on SDM only and not on all cognitive skill measures separately, because the association between SDM and SIRR is most pronounced, and all cognitive ability measures are highly correlated (correlation coefficients of

[16]The original variable on interviewer rating of the respondent's understanding of the questions is based on five possible categories: very poor, poor, fair, good excellent. We standardise the responses to mean 0 and standard deviation 1, so that the coefficients are comparable.

0.2 to 0.4).[17]

Because a summary measure of SIRR does not reveal whether item-response behavior is associated with cognitive ability on every individual question of the survey, we identify the questions on the 266-item survey which are not or only weakly associated with cognitive ability.[18] Figure C.1 (Appendix) shows that for 92% of all 266 questions we find a significantly higher cognitive ability for respondents than for non-respondents. The non-response rates of these questions are up to 5%. There are seven questions in the survey for which the difference in cognitive ability between respondents and non-respondents is not statistically significant, and only four questions for which respondents score lower on the SDM than non-respondents. Importantly, the top 20 item-non response questions – questions for which more than 20% of the sample did not respond – are not or only weakly associated with differences in cognitive ability. These questions refer to the timing of life events and workplace entitlements (paid paternity leave). For questions referring to the timing of events, respondents largely refused to answer whereas for workplace entitlements a large proportion ticked the box "Don't know", as opposed to simply not responding.[19]

3.2.2 Predictive power of SIRR

We show in Table 2 that SIRR is predictive of wages, educational attainment and health (columns 1, 3 and 5), even after controlling for standard covariates.[20] While of course the estimates for SIRR cannot be interpreted as causal, the results provide evidence that SIRR measures something that is an important determinant of economic outcomes. Importantly, the impact of SIRR is no longer statistically significant once we control for the cognitive ability measures (columns 2, 4 and 6). This latter finding allows us to conclude that SIRR captures facets of cognitive ability, and that the impact of SIRR on economic outcomes operates only through cognitive ability. Our findings from

[17] Our results are also robust to either using each cognitive ability measure separately or using a summary measure of all cognitive ability measures.

[18] A list of all questions on the SCQ and their respective non-response rates is provided in Table A.1 in Appendix A.

[19] Excluding these high non-response items from our SIRR proxy increases the association between SIRR and SDM. This is not surprising because only a small fraction of questions allowed for a Don't Know option as can be seen in Table A.

[20] Each regression model controls for gender, age, years of education (not included in education regression model), casual worker, years of tenure with current firm and its square, state of residence, region, country of birth, Big-5 personality traits, and employment status (in case of the health outcome variable.

the wage regression model provide additional evidence against the opportunity cost hypothesis of survey-item response behavior. SIRR is positively associated with higher wages – a 1 standard deviation increase in SIRR is associated with a 2.6 percent increase in wages – and the association operates trough the impact of SIRR on cognitive ability.

3.2.3 Stability of SIRR over time

Cognitive ability is considered to be a relatively stable trait in adulthood. It is considered to increase moderately in early adulthood and to remain stable until old age when cognitive functioning begins to decline (Schaie, 1988).[21] If SIRR behaves statistically in a similar way to a standard measure of cognitive ability, then individuals should not substantially change their response patterns over time. To test the fixed-trait hypothesis, we exploit the longitudinal nature of our data which allow us to calculate the average inter-temporal correlation coefficients (ITC) of SIRR between any two recording periods across 13 available waves of data. Figure 4 presents these ITCs for a balanced sample of 4,981 individuals who remained in the survey for 13 waves. The horizontal axis reports the starting wave, while the vertical axis reports the ITC.

If SIRR contains a strong element of an individual-specific component, then the ITC between time period t and $t - 10$ should not differ substantially from the correlation calculated between time period t and $t - 1$. Figure 4 demonstrates that the ITCs are the same across the waves and time lags. For instance, the ITC is 0.28 between Wave 1 and Wave 2, while the ITC between Wave 1 and Wave 13 is still 0.25.[22] We also calculated the probability of remaining an "all-item responder", conditional on the respondent being an "all-item responder" in the previous time period. In our data, this probability is over 62%. Since more than 30% of the sample are "all-item responders", a conditional probability of 62% implies that such individuals are twice as likely to respond again to all questions in a given time-period, compared to the average individual in the sample.

[21] At least 50% of the variance in cognitive test performance later in life is accounted for by childhood cognitive ability (Gow et al., 2011; Deary and J. Yang, 2012). Correlations between age 11 and 77 cognitive ability are estimated to be as high as 0.73 (Deary et al., 2000).

[22] The stability of the ITCs across time periods is independent of the starting waves. Figure 4 also reports the ITCs between wave 4, wave 8 and wave 10 and subsequent waves. ITCs are generally stronger for the later waves of the survey.

4 Using SIRR as a proxy variable for cognitive ability: A conceptual framework

Having shown that SIRR is significantly linked to cognitive ability, the question arises as to whether it has the potential to reduce omitted-variable biases (OVB) when used as proxy variable for unobserved cognitive ability, and under what conditions this objective would be achieved. The early theoretical literature on proxy variables suggested that it is always preferable to use even a noisy proxy variable as long as its measurement error is random and uncorrelated with the missing variable and other covariates in the structural model (Wickens, 1972; McCallum, 1972; Aigner, 1974).[23]

Frost (1979) criticized the assumption of random measurement error stating that "in general, the difference between the unmeasurable variable and the proxy variable is not a random variable independent of the true regressors" (p. 323). He highlighted that substantial biases may occur when the proxy variable is imperfect, which means that it is correlated with a key variable of interest in the structural model. Thus, proxy variables should not be used 'indiscriminately' (Frost, 1979, p. 325). The same point was later emphasized by Wolpin (1995), who derived the imperfect proxy variable bias in the context of job-search models.[24]

We depart from the assumption that our proxy variable – the Survey Item Response Rate (SIRR) – is an imperfect proxy because it is likely to correlate with key variables from the structural model, as shown in Section 2. We derive both necessary and sufficient conditions under which such an imperfect proxy variable approach reduces OVB in our setting, complementing the conceptual framework offered by Frost (1979). This conceptual framework is applied to a wage regression model, the standard example for illustrating the benefit of a new method that addresses OVB (e.g. Oster, 2017; Pei et al., 2018). We depart our discussion from a standard Mincer wage regression which assumes that hourly wages (Y_i) are a function of formal training (years

[23]The results presented in Wickens (1972) and McCallum (1972) are based on asymptotic derivations. Aigner (1974) expanded their work by deriving the relative biases in small samples and demonstrating that the tradeoff between the two methods, expressed in terms of mean squared error, depends on the sample size, the proportion of measurement error and the correlation between the missing variable and the main variable of interest. Aigner (1974)'s results indicate that the proxy variable approach is preferable in samples of 50 or more observations, even if the potentials for OVB and measurement error are high.

[24]Wolpin (1995) referred to the imperfect proxy variable as a 'crude' proxy variable. He suggested that the proxy-variable bias cannot be understood in isolation of a theoretical model. His work further demonstrated that a 'crude' proxy variable may confound the interpretation of other important variables in the model (Wolpin, 1995, 1997; Todd and Wolpin, 2003).

of education, X_i), cognitive ability (M_i) and variables (Z_i) that are commonly associated with productivity (e.g. non-cognitive ability, experience, type of work contract, etc):[25]

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 M_i + Z_i' \theta + u_i, \tag{6}$$

where error u_i satisfies strict exogeneity ($E(u_i|X, M, Z) = 0$). The parameter α_1 is of main interest as it measures the wage returns for an extra year of education. We further assume that we cannot observe M_i. As researchers we would have to work with one of the following misspecified models. In Eq. (7) M_i is omitted:

$$Y_i = \beta_0 + \beta_1 X_i + Z_i' \theta' + \mu_i, \tag{7}$$

It is straightforward to show the OVB in β_1, the estimated wage returns of education in the misspecified model (see Appendix D). In an alternative specification we add the variable P_i (in our case: SIRR) as a proxy for M_i:

$$Y_i = \gamma_0 + \gamma_1 X_i + Z_i' \theta'' + \gamma_3 P_i + \nu_i, \tag{8}$$

where $P_i = M_i + \varphi$ is measured with error, and a priori φ could include opportunity costs (hourly wages) or education as modelled in our theoretical model (see Equation 2). In order for the proxy variable approach to lead to unbiased estimates in the estimated returns to education, we would need to ensure that the measurement error φ does not include hourly wages (assumption 1) and education (assumption 2).

Hence, the first assumption implies that the proxy variable is redundant in Equation 6 such that $E(Y_i|X_i, Z_i, M_i, P_i) = E(Y_i|X_i, Z_i, M_i)$. We have tested the redundancy assumption empirically and shown that SIRR does not have a significant impact on wages once we control for cognitive ability (see Table 2), rejecting our theoretical proposition.[26]

[25] A similar model was used in Oster (2017) or Pei et al. (2018), except that we consider also non-cognitive ability as an important input into the wage production function (see Almlund et al., 2011, for a justification).

[26] If the redundancy assumption is violated, then an additional bias emerges. In the linear case, the measurement error in the proxy variable would then be as follows: $P_i = \delta_0 + \delta_1 Y_i + \delta_2 M_i + \delta_3 X_i + \varepsilon_i$. Solving for M_i and substituting it into the earnings function in Equation 6 yields $Y_i = \frac{\alpha_0 - \frac{\beta_0 \alpha 2}{\beta_2}}{1 + \frac{\beta_1 \alpha_2}{\beta_2}} + \frac{\alpha_1 - \frac{\beta_3 \alpha 2}{\beta_2}}{1 - \frac{\beta_1 \alpha_2}{\beta_2}} X_i + \frac{\frac{\alpha_2}{\beta_2}}{1 - \frac{\beta_1 \alpha_2}{\beta_2}} P_i + \frac{\mu_i - \frac{1}{\beta_2} \varepsilon_i}{1 - \frac{\beta_1 \alpha_2}{\beta_2}}$. It can be seen that there would be bias in the estimated coefficients and that P_i would be correlated with the error term through its correlation with ε_i. The OVB could however still be reduced.

The second assumption implies that there is no remaining correlation between the missing variable M_i and education X_i (or all other explanatory variables) conditional on controlling for P_i: $E(M_i|X_i, Z_i, P_i) = E(M_i|P_i)$. If this assumption is violated, i.e. the proxy variable is imperfect, we obtain an imperfect proxy bias (IPB) in γ_1 (see Appendix D). However, even if some residual correlation remains between M_i and X_i, the imperfect proxy could still reduce OVB. In what follows, we derive and discuss the conditions under which an imperfect-proxy variable will reduce OVB. We follow Frost (1979) to express the relative squared biases (IPB^2/OVB^2) in terms of three partial correlation coefficients:[27]

$$\lambda = \frac{(E(\hat{\gamma}_1 - \alpha_1))^2}{(E(\hat{\beta}_1 - \alpha_1))^2} = \frac{(r_{XM|Z} - r_{MP|Z} r_{XP|Z})^2}{r^2_{XM|Z}(1 - r^2_{XP|Z})^2}. \tag{9}$$

For the proxy variable to improve upon OVB, we need to show that $\lambda < 1$. The relative bias depends on the strength of the proxy ($r_{MP|Z}$) and the strength of the relationship between M_i and X_i ($r_{XM|Z}$), which indicates the potential for OVB. It also depends on the correlation between X_i and P_i ($r_{XP|Z}$). Large values for $r_{XP|Z}$ imply that the proxy variable is closer in nature to education rather than the underlying omitted variable. Using the proxy variable thus may introduce a multicollinearity problem.[28] While $r_{MP|Z}$ and $r_{XM|Z}$ are usually unobserved by the researcher, $r_{XP|Z}$ is always observed.

We propose that the relative performance of the proxy variable approach depends on the sign equivalence of $r_{MP|Z}$ and $r_{XM|Z}$ and the ratio between $r_{MP|Z}$ and $r_{XM|Z}$ relative to $r_{XP|Z}$. The necessary and sufficient conditions for an imperfect proxy variable to reduce OVB are as follows (the proofs are shown in Appendix E):

Theorem 1: A necessary condition for the imperfect proxy variable to reduce OVB is that $\text{sign}(r_{MP|Z}) = \text{sign}(r_{XM|Z})$ if $r_{XP|Z} > 0$, and $\text{sign}(r_{MP|Z}) \neq \text{sign}(r_{XM|Z})$ if $r_{XP|Z} < 0$.

[27] Frost (1979) was the first to express the squared relative biases between the imperfect proxy variable and omitted variable approach. He did not discuss the conditions under which the proxy variable holds other than stating that the strength of the proxy variable is not a necessary condition and that a necessary, but not sufficient, condition for the squared biases to be smaller than 1 is if $r^2_{MP|Z} > r^2_{XM|Q}$. We are able to demonstrate that the latter was an erroneous conclusion, by analytically showing both necessary and sufficient conditions, and demonstrating their correctness in a simulation exercise.

[28] In case of using multiple proxies, λ will be calculated on the basis of partial R-squared which measures the additional contribution of the relevant variable in explaining the variation in the dependent variable. Partial correlation coefficients do not tell us about the sign of the correlation between the proxy and cognitive skills, which is relevant in the relative bias calculation. In this case, we calculate λ by using the alternative formula: $\frac{(\delta_{MX|Z} - \delta_{MX|ZP})}{\delta_{MX|Z}}$.

Theorem 2: A sufficient condition for the imperfect proxy variable to reduce OVB is that $r_{XP|Z} < \frac{r_{MP|Z}}{r_{XM|Z}} < \frac{2-r^2_{XP|Z}}{r_{XP|Z}}$ if $r_{XP|Z} > 0$, and $r_{XP|Z} > \frac{r_{MP|Z}}{r_{XM|Z}} > \frac{2-r^2_{XP|Z}}{r_{XP|Z}}$ if $r_{XP|Z} < 0$.

Theorem 1 implies that if $r_{XP|Z} > 0$, then for the imperfect proxy variable approach to improve upon omitting a relevant variable, it must be the case that the relative strength of the proxy variable must be positive. This means that the sign of the partial correlation between the proxy and the omitted variable must be the same sign as the partial correlation between the missing and the main variable of interest in the model (in our case: education). If $r_{XP|Z} < 0$, then the two partial correlation coefficients must be of opposite signs. *Theorem 2* implies that to definitely improve upon the omitted variable approach, the relative strength of the proxy variable must lie within an interval bounded between $r_{XP|Z}$ and $\frac{2-r^2_{XP|Z}}{r_{XP|Z}}$.

We illustrate these trade-offs in a simulation exercise in which we vary $r_{XP|Z}$ and assume $r_{XP|Z} > 0$. Figure 5 depicts the relative strength of the proxy variable ($\frac{r_{MP|Z}}{r_{XM|Z}}$) on the horizontal axis. Although the values for this ratio can become indefinitely large, we restrict its possible values between -0.5 to +3 to make the figures more legible. Negative values on the x-axis indicate that the sign of the partial correlation coefficients are opposite in sign. λ is expressed on the vertical axis, where values smaller than 1 imply a reduction in the OVB, and a value larger than 1 implies an increase in the OVB when including the proxy variable. Figure 5 simulates λ for five possible values for $r_{XP|Z}$ (0.05, 0.20, 0.40, 0.60, 0.80).

The analytical and simulation results emphasize that a strong proxy variable is neither a necessary nor a sufficient condition for reducing OVB. The proxy variable needs to be strong only relative to the potential for the omitted variable problem and relative to the potential of the multicollinearity problem. For instance, if the latter is very small ($r_{XP|Z} = 0.05$), then the IPB is smaller than the OVB as long as the relative strength of the proxy variable is greater than 0.05 and smaller than 39.95. In stark contrast, if the potential for multicollinearity is very large ($r_{XP|Z} = 0.80$), then the relative strength of the proxy variable must lie within a small window of 0.80 and 1.7.

Since $r_{XP|Z}$ is always observed, the researcher can make an informed judgment about whether the necessary and sufficient conditions are likely to hold. For instance, the researcher can evaluate whether the sign equivalence of $r_{MP|Z}$ and $r_{XM|Z}$ is likely to hold and judge whether the width of the bias-improvement window is large enough for certainty. The larger is $r_{XP|Z}$, the less certainty the researcher has about whether the proxy variable helps to reduce OVB. We will return to this insight again in our empirical application.

5 Reducing omitted-variable biases in wage returns to education and locus of control

5.1 The validity of SIRR as a proxy for cognitive ability

Because we have data on the unobserved variable, cognitive ability, we can test whether SIRR reduces omitted-variable biases and calculate the exact bias reductions that SIRR may achieve.[29] We calculate bias reductions for the returns to education and locus of control, a widely studied non-cognitive skill in the context of wage regressions (see Cobb-Clark, 2015, for an overview). SIRR is a valid proxy if λ is smaller than 1. We therefore test that λ is equal to one against the one-sided hypothesis that it is smaller than one:

$$H_0: \quad \lambda = 1, \tag{10}$$

$$H_a: \quad \lambda < 1. \tag{11}$$

If the null hypothesis is rejected against the alternative, we have certainty that including SIRR as a proxy variable for cognitive ability reduces OVB in estimating the wage returns to education and locus of control. Panel A of Table 3 reports for each covariate X (education, and locus of control) the respective partial correlation coefficients, the ratio of the squared biases (λ) and the relative strength of the proxy variable ($\frac{r_{MP|Z}}{r_{XM|Z}}$). The partial correlation coefficients indicate that, in our data scenario, the strength of the proxy is neither weak nor strong ($r_{MP|Z} = 0.12$), the potential for the MCP is medium to low for education and locus of control ($r_{XP|Z} = 0.13$ and 0.04), and the potential for OVB is also medium to low ($r_{XM|Z} = 0.14$ and 0.07).

The results presented in Panel A suggest that SIRR fulfills the necessary condition to be a valid proxy variable for estimating the returns to education and locus of control, because $\text{sign}(r_{MP|Z}) = \text{sign}(r_{XM|Z})$. Furthermore, the sufficient condition is also fulfilled, since the relative strength of the proxy variable (Panel A4.) is always greater than the potential multicollinear-

[29] Technically, our measure of cognitive ability is a proxy for true, underlying cognitive ability. Our calculations are based on the assumption that our cognitive ability measure contains at worst random error.

ity problem (Panel A3.), and it is always smaller than the maximum upper bound ($\frac{2-r^2_{XP|Z}}{r_{XP|Z}}$). Thus, including SIRR as a proxy variable in the wage regression model would yield significant bias reductions for both the estimated returns to education and locus of control. This is also reflected in values for λ that are significantly below 1 (Panel A5.).

Theorems 1 and 2 are also useful for discussing the bias-reduction potential of SIRR, even in the absence of observable information on the omitted variable. Knowledge about the degree of multicollinearity is enough to make an informed risk assessment regarding use of the proxy. We observe in the data that the partial correlation coefficient between education X and proxy P is positive (0.13). To fulfill the necessary condition of sign equivalence (Theorem 1), we need to argue only that the sign of the partial correlation coefficient between the unobserved variable M (cognitive ability) and X (e.g. education) is also positive. Such judgment could be based entirely on insights from previous literature. In our setting, this would be a reasonable and credible assumption.

Furthermore, to fulfill the sufficient condition (Theorem 2), we would need to assess whether it is reasonable to assume that the relative strength of the proxy lies within an interval that can be calculated from knowledge of $r_{XP|Z}$. In our data setting, this interval is: $0.13 < \frac{r_{MP|Z}}{r_{XM|Z}} < 15.5$. In other words, for SIRR to be an *invalid* proxy, the strength of the proxy would have to be either 16 times larger or just one eighth or less than the potential for omitted variable problem ($r_{MP|Z} \geqslant 15.5 r_{XM|Z}$ or $r_{MP|Z} \leqslant 0.13 r_{XM|Z}$). Given that this is a wide interval, one could reasonably assume that the proxy variable approach is associated with a very low risk of exacerbating the bias. To visualize this, our data setting lies between the solid line (multicollinearity potential of 0.05) and the short-dashed-dotted line (multicollinearity potential of 0.20) in Figure 5.

5.2 Magnitude of bias reductions

Now that we have demonstrated that SIRR is a valid proxy for cognitive ability , we are interested in understanding the magnitude by which we can reduce OVB. Panel B1. of Table 3 shows that using SIRR as proxy for cognitive ability significantly reduces OVB in the estimated returns to education and locus of control by 8.9% and 6.5%, respectively.[30] The bias-reduction potential is slightly higher – 10.1% and 7.3%, respectively for education and locus of control – when using

[30]Standard errors for the bias reductions are obtained with the delta method.

indicator variables that represent different levels of SIRR to approximate the non-linear relationship between SIRR and cognitive ability (Panel B2.)[31]. The bias reductions are slightly smaller when the calculation of SIRR is based on refusals only (Panel B3.). They are also robust to using a larger sample spanning ages 20-69 (Panel B4.). Furthermore, the bias reductions are slightly higher if we drop individuals with SIRR below 0.4 (10.3% and 7.3% in Panel B5.) or individuals with a non-English speaking background (10.2% and 6.3% in Panel B6). Panels B7. and B8. show that SIRR also reduces OVB related to aspects of cognitive ability measured by BDS and NART instead of SDM.

Whether bias reductions of up to 10% are large or small depends, of course, on the size of the OVB and the degree of precision required. In the context of wage regression models, the estimation biases in the returns to education are traditionally small, averaging around 10% (Card, 2001). In line with previous research, Table F.1 shows that in our application, one extra year of education is associated with a 7.2% increase in hourly wages, and the OVB is 12.2%. At face value, bias reductions of up to 10% are not large when considering a one-period model. However, if the returns to education were a policy-relevant parameter used in lifecycle simulation studies that factor in multiplier effects over time – typical examples are labor-supply, price or income elasticities – one may consider such OVB reductions as relevant.

5.3 Relative performance of SIRR to other proxy variables

We have shown so far that SIRR is a valid proxy for cognitive ability, reducing OVB in the returns to education and locus of control by up to 10%. In this section we evaluate the relative performance of SIRR against four alternative proxy variables, including the interviewer's ratings of the participants' understanding of the questions, minutes taken for completing the personal questionnaire or the household questionnaire, and days elapsed until the self-completion questionnaire is returned. Table 4 shows the estimated λ and the associated OVB reduction.[32] All proxy variables are standardized to mean 0 and SD 1.

We find that only the interviewer's rating of the participants' understanding of the questions

[31]To approximate this non-linear relationship, we use four dummy variables that indicate item response rates within the 5th percentile, between the 5th and 10th percentile, between the 10h and 25th percentile, and above the 25th percentile; the base category is 100% response rate.

[32]Full estimation results are provided upon request.

comes close to the bias-reduction potential of SIRR, reducing OVB in the estimated returns to education and locus of control by 6.1% and 12.5%. Although this suggests that interviewer ratings could be useful proxies for unobserved cognitive ability, such measures are not by-products of standard survey collection procedures but rather questions which actively have to be included as part of the survey. The main advantage of SIRR is that it can be easily calculated by researchers in any survey.

5.4 Heterogeneity by block of questions

Our SIRR measure captures the average relationship between item response and cognitive ability (SDM) across 266 survey questions. However, we have demonstrated in Figure C.1 a large degree of heterogeneity in the relationship between cognitive ability and item-response behavior across survey items. In this section we identify the block of questions with the highest potential for bias reduction. Figure 6 scatter plots the bias reduction in the returns to education for each individual survey question when used as a binary proxy variable (vertical axis) and the difference in cognitive ability between respondents and non-respondents (horizontal axis). Grey dots indicate that the bias reduction is not statistically significant at the 5% level whereas black dots indicate significant bias reductions.

We find a positive relationship between bias reduction and the difference in cognitive ability, with a correlation coefficient of 0.47 (blue fitted line). This indicates that the greater the difference in cognitive ability, the larger is the bias reduction in the returns to education. Furthermore, there are 31 variables for which each individual question reduces OVB in the estimated returns to education by over 1.5% and up to 3.3%. These are the seven questions on the usefulness of computer use to learn more skills[33], 14 questions on weekly time use[34], three questions on household expen-

[33] SCQ:B25a Computer use - My level of computer skills meets my present needs, SCQ:B25b Computer use - I feel comfortable installing or upgrading computer soft, SCQ:B25c Computer use - Computers have made it possible for me to get more done, SCQ:B25d Computer use - Computers have made it easier for me to get useful information, SCQ:B25e Computer use - Computers have helped me to learn new skills other than, SCQ:B25f Computer use - Computers have helped me to communicate with people, SCQ:B25g Computer use - Computers have helped me reach my occupational career.

[34] SCQ:C7e Household decisions - The way children are raised SCQ:C7d Household decisions - The number of hours your partner/spouse spends in, SCQ:B24f Hours per week - Playing with your children, SCQ:B24b Hours per week - Travelling to and from a place of paid employment, SCQ:B24a Hours per week - Paid employment, SCQ:B24c Hours per week - Household errands SCQ:B24d Hours per week - Housework, SCQ:B24h Hours per week - Volunteer/Charity work, SCQ:B24f Minutes per week - Playing with your children SCQ:B24b Minutes per week - Travelling to and from a place of paid employment, SCQ:B24a Minutes per week - Paid employment SCQ:B24c

ditures[35], 12 questions on timing of life events[36] and five questions on achievement motivation and satisfaction regarding domestic life[37]. These are also the questions that are most strongly associated with differences in SDM between respondents and non-respondents. For instance, non-respondents for the time use and household expenditure information score on average more than 10 points lower on the SDM test than responding individuals, which corresponds to a 0.8 SD difference in SDM. Bundling these 31 high-yield response questions into a continuous summary proxy for ability-related item response reduces the bias in the returns to education significantly by 8.6%. Bundling all questions that individually yield positive and significant bias reductions in the estimated returns to education yield an overall bias reduction of 9% (full estimation results are provided upon request).

5.5 Relative performance of proxy-variable approach to other methods for bias reduction

Recent literature proposes to address OVB by calculating bias-adjusted estimates of the treatment effect of interest using information about the likely degree of self-selection into treatment (Oster, 2017; Altonji et al., 2005). Researchers using these methods have to make assumptions on: (1) the amount of variation in the outcome variable that can be explained by observable and unobservable characteristics[38] and (2) the degree of selection on unobservables relative to selection on observables. In the absence of proxies, the advantage of this method is that it makes no assumption about the number and type of unobservable covariates. However, the approach is based on the very strong assumption that the relationship between the variable of interest and observable characteristics is informative of the relationship between the variable of interest and

Minutes per week - Household errands, SCQ:B24d Minutes per week - Housework, SCQ:B24h Minutes per week - Volunteer/Charity work

[35]SCQ:C9i Monthly household expenditure - Children's clothing and footwear SCQ:C9r Annual household expenditure - Education fees, SCQ:C9h Monthly household expenditure - Women's clothing and footwear

[36]SCQ:B23h How long ago life event happened - Death of spouse/child - no answer SCQ:B23k How long ago life event happened - Victim of violence - no answer, SCQ:B23h Life events in past year: Death of spouse or child SCQ:B23p Life events in past year: Fired or made redundant SCQ:B23u Life events in past year: Changed residence

[37]SCQ:B21b Achievement motivation - I like situations where I can find out how ca, SCQ:B21e Achievement motivation - I am afraid of tasks that I cannot work out o, SCQ:B13g Satisfaction with: Relationship with step parents, SCQ:B13b Satisfaction with: Children, SCQ:B13d Satisfaction with: Relationship with step children.

[38]Altonji et al. (2005) suggests that R-squared equals 1 while Oster, 2017 argues it is highly unlikely that R-squared ever reaches 1, among other because of measurement error in the outcome variable. Oster (2017) provides a rule-of-thumb for a maximum possible R-squared that is about 1.3 times larger than the R-squared obtained from a model that suffers from OVB.

unobservable characteristics.

We assess how the proxy variable approach to reduce OVB compares to the method suggested in Oster (2017).[39] Table 5 shows that the true returns to education (7.63%, column 1) are over-estimated when excluding cognitive ability (7.90%, column 2). Out of all the estimated effects of the wage returns to education, the one using the proxy variable method produces an estimate (7.82%, column 5) that is closest to the true effect. In comparison, under Oster (2017) in columns 3 and 4, the effect is either 7.41% (assuming an R-squared that is about 1.3 times larger than the R-squared obtained from a model that suffers from OVB: $R^2_{max} = 1.3 * \widetilde{R} = 0.29$), or -5.9% (assuming $R^2_{max} = 1$).[40] We conclude that in certain situations the proxy variable approach is a good if not better alternative to the bias-adjusted methods proposed in Altonji et al. (2005) and Oster (2017), for example if information on the nature of omitted confounders and measures of potential proxies are available.

6 Conclusions

We propose that a simple summary measure of survey item-response behavior, the survey item response rate (SIRR), reveals important information on survey respondents' cognitive abilities. We are the first to demonstrate the usefulness of SIRR to reduce cognitive ability related omitted-variable biases in the wage returns to education and locus of control, a widely studied non-cognitive skill. Furthermore, we are the first to provide both necessary and sufficient conditions under which the use of an imperfect proxy variable, a proxy variable that correlates with key control variables in the structure model, reduces biases. Researchers can use the conditions as an empirical guideline to evaluate the risks associated with their own proxy-variable estimation strategy.

More specifically, we show that SIRR behaves in a statistically similar manner to cognitive ability, which is in line with the cognitive interpretation of our economic model of survey item-

[39]For simplicity, we base the estimations on the population of 4,432 working individuals with positive hourly wages in Wave 12 and estimate coefficients of interest with OLS. All regression models control for Big 5 personality traits, gender, age, a measure for casual worker, years of tenure with current firm, state of residence, region (other urban, bounded locally, rural balance), country of birth (english speaking, other), number of children in household in various age brackets (0-4 years, 5-9 years, 10-14 years).

[40]This finding is consistent with Oster (2017) who proposed that under $R^2_{max} = 1$ treatment effects may be underestimated.

response behavior. SIRR (i) varies positively with cognitive ability as well as the interviewers' assessment of the respondents' understanding of the question over and above the influence of a range of confounding factors; (ii) has an individual-specific fixed component in a similar manner to cognitive ability; and (iii) is predictive of standard economic outcomes. The empirical findings are only weakly supportive of SIRR capturing personality-related willingness to complete survey questions. There is also no evidence to support the prediction of the economic model that high opportunity cost considerations may dominate non-response.

In our data application, SIRR reduces omitted-variable biases in the returns to education and locus of control by roughly 10%. We show that the bias reductions are robust to a range of sensitivity tests such as using a non-linear measure of SIRR and alternative sample restrictions (wider age cutoffs, exclusion of outliers in SIRR, and exclusion of sample members from non-English speaking backgrounds). We also show that in certain situations – for example if information on the nature of omitted confounders and measures of proxies are available – the proxy variable approach is a good alternative to the bias-adjusted methods that rely on assumptions about the likely degree of self-selection into treatment (Oster, 2017; Altonji et al., 2005).

We also show that our proxy performs well relative to other proxies, such as the minutes taken for completing the personal and household questionnaire and days taken to return the self-completion questionnaire. We identify a subset of survey questions for which item response is most strongly associated with differences in cognitive ability. These are questions that require recall ability and tedious coding such as questions on household expenditures or time-use that are widely used as inputs or outcomes of economic decision models over the lifecycle (see Aguiar et al., 2012, for an overview). Survey data on these types of questions are more complete for participants with better cognitive function. Researchers working with time-use or household-expenditure data may need to account for this self-selection as was proposed and demonstrated in Heffetz and Rabin (2013) in the context of life satisfaction.

Our conceptual framework extends both Frost (1979) and Wolpin (1995), who acknowledged that the use of imperfect or crude proxy variables may exacerbate estimation biases, an observation that was missing in the early work of Wickens (1972), McCallum (1972), and Aigner (1974). We derive both necessary and sufficient conditions under which an imperfect proxy variable reduces omitted-variable biases. A minimum condition for a valid proxy variable is sign equivalence between two correlation coefficients, namely the partial correlation between the missing variable and the proxy (strength of the proxy) and the partial correlation between the missing variable and the variable of interest (potential for omitted variable problem). A sufficient condition is that the ratio of the strength of the proxy to the potential for omitted-variable problem – which we refer to as the relative strength of the proxy – needs to be bound by terms that only depend on the partial correlation between the variable of interest and the proxy (potential for multicollinearity). These bounds can easily be calculated. Although these conditions do not allow for a water-tight testing procedure, they equip researchers with an empirical guideline for informed risk evaluation of using an imperfect proxy variable approach. Such guidelines can be used in a wide array of empirical settings.

References

Aguiar, M., Hurst, E., Karabarbounis, L., 2012. Recent developments in the economics of time use. Annual Review of Economics 4, 373–398.

Aigner, D.J., 1974. MSE dominance of least squares with errors of observation. Journal of Econometrics 2, 365–372.

Almlund, M., Lee Duckworth, A., Heckman, J.J., Kautz, T., 2011. Personality psychology and economics. In S.M. Eric A. Hanushek, L. Woessmann (Eds.), Handbook of the Economics of Education, vol. 4. North-Holland, pp. 1–181.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. Journal of Political Economy 113, 151–184.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton.

Ashenfelter, O., Rouse, C., 1998. Income, schooling, and ability: evidence from a new sample of identical twins. The Quarterly Journal of Economics 113, 253–284.

Beatty, P., Herrmann, D.J., 2002. To answer or not to answer: Decision processes related to survey item nonresponse. In R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little (Eds.), Survey Nonresponse. John Wiley and Sons, New York, pp. 71–86.

Behaghel, L., Crepon, B., Gurgand, M., Le Barbanchon, T., 2015. Please call again: Correcting non-response bias in treatment effect models. Review of Economics and Statistics Forthcoming.

Blackburn, M., Neumark, D., 1993. Omitted-ability bias and the increase in returns to schooling. Journal of Labor Economics , 521–544.

Blau, P.M., 1964. Exchange and power in cosial life. Wiley, New York.

Bollinger, C., David, M.H., 2005. I Didn't Tell and I Won't Tell: Dynamic Response Error in the SIPP. Journal of Applied Econometrics 20, 563–569.

Bollinger, C., Hirsch, B., 2013. Is earnigns response ignorable. Review of Economics and Statistics 95, 407–416.

Borghans, L., Golsteyn, B.H.H., Heckman, J.J., Humphries, J.E., 2016. What grades and achievement tests measure. PNAS 113, 13354–13359.

Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica 69, 1127–1160.

Chatterji, P., 2014. Education and health: Disentangling causal relationships from associations. In A.J. Culyer (Ed.), Encyclopedia of Health Economics. Elsevier, San Diego, pp. 250 – 258.

Cobb-Clark, D.A., 2015. Locus of control and the labour market. IZA Journal of Labor Economics 4.

Cobb-Clark, D.A., Kassenboehmer, S., Schurer, S., 2014. Healthy habits: What explains the connection between diet, exercise, and locus of control? Journal of Economic Behavior & Organization 98, 1–28.

Cobb-Clark, D.A., Schurer, S., 2012. The stability of big-five personality traits. Economics Letters 115, 11–15.

Cobb-Clark, D.A., Schurer, S., 2013. Two economists' musings on the stability of locus of control. The Economic Journal 123, F358–F400.

Croon, M., 2002. Using predicted latent scores in general latent structure models. In G.A. Marcoulides, I. Moustaki (Eds.), Latent Variable and Latent Structure Models. Lawrence Erlbaum Associates, Inc., pp. 195–223.

Deary, I.J., J. Yang, e.a., 2012. Genetic contributions to stability and change in intelligence from childhood to old age. Nature 482, 212–215.

Deary, I.J., Whalley, L., Lemmon, H., Crawford, J., Starr, J., 2000. The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. Intelligence 28, 49–55.

Dee, T.S., West, M., 2011. The non-cognitive returns to class size. Educational Evaluation and Policy Analysis 33, 23–46.

Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American Statistical Association 94, 1053–1062.

Dillman, D.A., 2000. Mail and Internet Surveys: The Tailored Design Method. Wiley, New York.

Dillman, D., Eltinge, J., Groves, R., Little, R., 2002. Survey nouresponse in design, data collection, and analysis. In J.L.E.R.J.A.L. R. M. Groves D. A. Dillman (Ed.), Survey Nonresponse. Wiley, New York, pp. 3–26.

Elkins, R.K., Kassenboehmer, S.C., Schurer, S., 2017. The stability of personality traits in adolescence and young adulthood. Journal of Economic Psychology 60, 37–52.

Frost, P.A., 1979. Variables and spcification bias. The Review of Economics and Statistics 61, 323–325.

Gelbach, J.B., 2016. When do covariates matter? And which ones, and how much? Journal of Labor Economics 34, 509–543.

Gensowski, M., 2018. Personality, IQ, and lifetime earnings. Labour Economics In print.

Gow, A., Johnson, W., et al., 2011. Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. Psychological Aging 26, 232–240.

Gronau, R., 2010. Zvi Griliches' contribution to the theory of human capital. In Contributions in Memory of Zvi Griliches. National Bureau of Economic Research, pp. 275–297.

Heckman, J., Hotz, J., 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. Journal of the American Statistical Association 84, 862–874.

Heckman, J.J., Humphries, J.E., Veramendi, G., 2017. Returns to education: The causal effects of education on earnings, health, and smoking. Journal of Political Economy Forthcoming.

Heckman, J.J., Kautz, T., 2014. Fostering and measuring skills: Interventions that improve character and cognition. In J.J. Heckman, T. Kautz (Eds.), The Myth of Achievement Tests: The GED and the Role of Character in American Life. University of Chicago Press.

Heckman, J., Pinto, R., Savelyev, P.A., 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. American Economic Review 103, 1–35.

Hedengren, D., Stratmann, T., 2013. The dog that didn't bark: Item non-response effects on earnings and health. Unpublished manuscript, George Mason University.

Heffetz, O., Rabin, M., 2013. Conclusions regarding cross-group differences in happiness depend on difficulty of reaching respondents. American Economic Review 103, 3001–3021.

Heine, S., Buchtel, E.E., Norenzayan, A., 2008. What do cross-national comparisons of personality traits tell us? The case of conscientiousness. Psychological Science 19, 309–313.

Hitt, C., Trivitt, J., Cheng, A., 2016. When you say nothing at all: The surprising predictive power of student effort on surveys. Economics of Education Review 52, 105–119.

Hitt, C., Trivitt, J., 2013. Don't know? or don't care? predicting educational attainment using survey item response rates and coding speed tests as measures of conscientiousness. Working paper, University of Arkansas Department of Education Reform.

Holmlund, H., Silva, O., 2014. Targeting non-cognitive skills to improve cognitive outcomes: Evidence from a remedial education intervention. Journal of Human Capital 8, 126–160.

Homans, G., 1961. Social Behavior: Its Elementary Forms. Harcourt Brace Jovanovich, New York.

Imbens, G., 2003. Sensitivity to exogeneity assumptions in program evaluation. American Economic Review 93, 126–132.

Jackson, C.K., 2017. What do test scores miss? the importance of teacher effects on non-test score outcomes. Journal of Political Economy Forthcoming.

Kautz, T., Heckman, J.J., Diris, R., ter Weel, B., Borghans, L., 2014. Fostering and measuring skills - improving cognitive and non-cognitive skills to promote lifetime success. OECD Education Working Papers no. 110, OECD Publishing, Paris.

Kleinjans, K.J., van Soest, A., 2014. Rounding, focal point answers and nonresponse to subjective probability questions. Journal of Applied Econometrics 29, 567–585.

Kreuter, F., 2013. Improving Surveys with Paradata: Analytic Uses of Process Information. John Wiley & Sons.

Lleras, C., 2008. Do skills and behaviours in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. Social Science Research 37, 888–902.

Lundberg, S., forthcoming. Non-cognitive skills as human capital. In C. Hulten, V. Ramey (Eds.), Implications for Future U.S. GDP Growth. NBER & CRIW.

McCallum, B.T., 1972. Relative asymptotic bias from errors of omission and measurement. Econometrica 40, 757–758.

Murphy, K., Topel, R., 1990. Efficiency wages reconsidered: Theory and evidence. Advances in the Theory and Measurement of Unemployment , 204–240.

Oreopoulos, P., 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. American Economic Review 96, 152–175.

Oster, E., 2017. Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics 0, 1–18.

Pearlin, L.I., Schooler, C., 1978. The structure of coping. Journal of Health and Social Behavior 19, 2–21.

Pei, Z., Pischke, J.S., Schwandt, H., 2018. Poorly measured confounders are more useful on the left than on the right. Journal of Business and Economic Statistics Forthcoming.

Pudney, S., 2014. Heaping and leaping: Survey response behaviour and the dynamics of self-reported consumption expenditure. ISER Working Paper 2008-09, University of Essex.

Raessler, S., Riphahn, R.T., 2006. Survey item nonresponse and its treatment. Allgemeines Statistisches Archiv 90, 217–232.

Riphahn, R.T., Serfling, O., 2005. Item non-response on income and wealth questions. Empirical Economics 30, 521–538.

Rosenbaum, P., Rubin, D., 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society, Series B 45, 212–218.

Saucier, G., 1994. Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. Journal of Personality Assessment 63, 506–516.

Schaie, C.H.K.W., 1988. Stability and change in adult intelligence: Ii. simultaneous analysis of longitudinal means and covariance structures. Psychology and Aging 3, 122–130.

Schurer, S., 2017. Does education strengthen life skills of adolescents? IZA World of Labor 366.

Schwarz, N., Oyserman, D., 2001. Asking questions about behavior: Cognition, communication, and questionnaire construction. American Journal of Evaluation 22, 127–160.

Singer, E., Presser, S., 2007. Privacy, confidentiality, and respondent burden as factors in telephone survey nonresponse. In J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. d. Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, R.L. Sangster (Eds.), Advances in Telephone Survey Methodology. Wiley, Hoboken, NJ, pp. 449–470.

Singer, E., Ye, C., 2013. The nonresponse challenge to surveys and statistics. The Annals of the American Academy of Political and Social Science 645, 112–141.

Sudman, S., Bradburn, N., Schwarz, N., 1996. Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. Jossey Bass Publishers, San Francisco.

Summerfield, M., 2010. User manual HILDA Release 9. Tech. rep., Melbourne Institute of Applied Economic and Social Research - The University of Melbourne, Melbourne.

Theil, H., 1957. Specification errors and the estimation of economic relationships. Review of the International Statistical Institute 2, 41–51.

Todd, P., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. The Economic Journal 113, F3–F331.

Tourangeau, R., 2003. Cognitive aspects of survey measurement and mismeasurement. International Journal of Public Opinion Research 15, 3–7.

Wand, M.P., Jones, M., 1995. Kernel smoothing. Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Watson, N., 2011. Methodology for the hilda top-up sample. HILDA Project Technical Paper Series 1/11, Melbourne Institute of Applied Economic and Social Research.

Weijters, B., Geuens, M., Schillewaert, N., 2010. The stability of individual response styles. Psychological Methods 15, 96–11.

West, M.R., Kraft, M., Finn, A., Martin, R., Duckworth, A., Gabrieli, C., Gabrieli, J., 2016. Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. Educational Evaluation and Policy Analysis 20, 1–23.

Wetzel, E., Luedtke, O., Baehnke, I.Z.J., 2016. The stability of extreme response style and acquiescence over 8 years. Assessment 23, 279–291.

Wickens, M.R., 1972. A note on the use of proxy variables. Econometrica 40, 759–760.

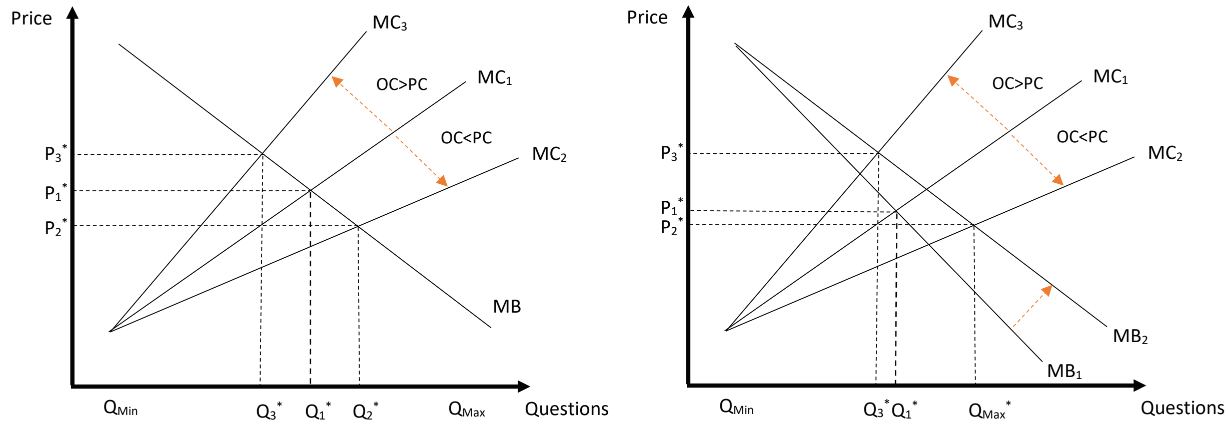
Wolpin, K.I., 1995. Empirical methods for the study of labor force dynamics. In F. Welch (Ed.), Fundamentals of Pure Applied Economic Series, vol. 60: Labor Economics Section. Harwood Academic Publishers, Luxembourg.

Wolpin, K.I., 1997. Determinants and consequences of the mortality and health of infants and children. In M. Rosenzweig, O. Stark (Eds.), Handbook of Population and Family Economics, vol. 60. Elsevier, Amsterdam.

Wooden, M., 2013. The measurement of cognitive ability in wave 12 of the hilda survey. HILDA Project Discussion Paper Series 1/13, Melbourne Institute of Applied Economic and Social Research.

Zamarro, G., Cheng, A., Shakeel, M.D., Hitt, C., 2016. Comparing and validating measures of character skills: Findings from a nationally representative sample. Unpublished manuscript, University of Arkansas Department of Education Reform.

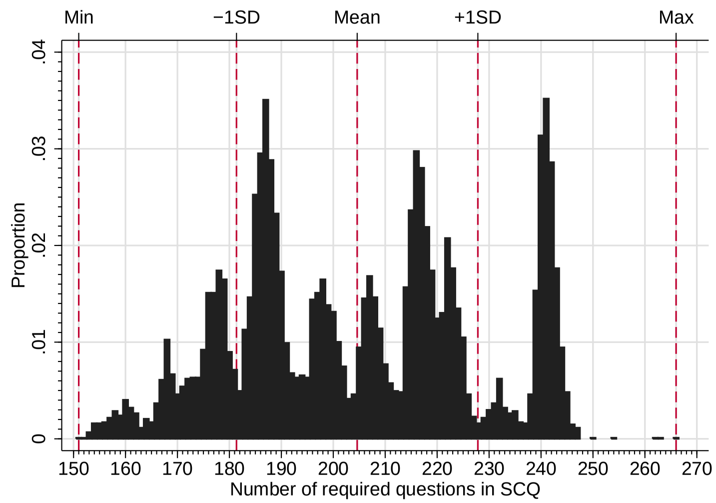
Zweimueller, J., 1992. Survey non-response and biases in wage regressions. Economics Letters 39, 105–109.



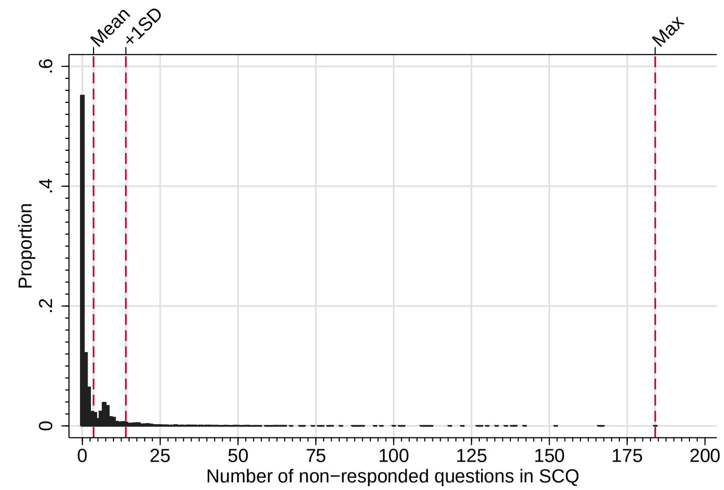
(a) Cognitive ability

(b) Personality

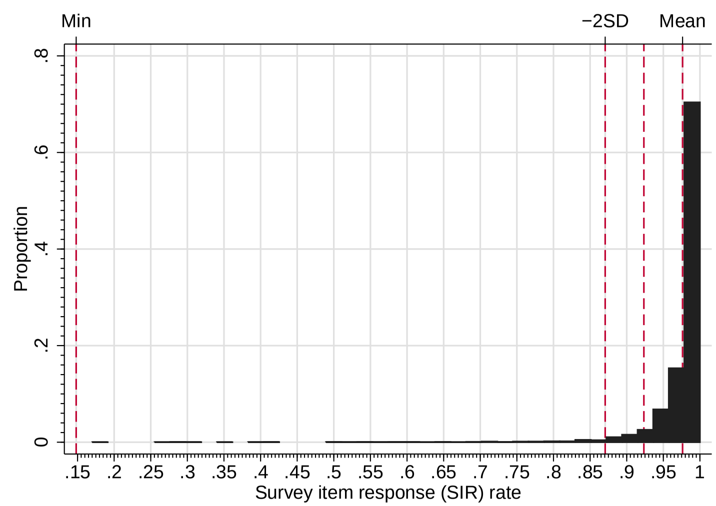
Figure 1: Ability changes the slope of marginal cost and benefit curves of responding to survey questions. The total number of questions Q a participant responds to is where marginal cost (**MC**) and benefit curves (**MB**) intersect. (a) Cognitive ability steepens marginal costs curve because it increases opportunity costs, but flattens marginal costs curve because it reduces psychic costs. At MC3 opportunity cost increases outweigh psychic cost decreases for increases ($OC > PC$) in cognitive ability; at MC2 psych cost decreases outweigh opportunity cost increases ($OC < PC$) for increases in cognitive ability;(b) Personality traits flatten marginal benefits curve from MB1 to MB2, but steepen marginal cost curve because they increase opportunity costs (MC1 to MC3), and flatten marginal costs curve, because of their reduction of psychic costs (MC1 to MC2). Model assumes additive separability of cognitive ability and personality traits.



(a) Number of required responses

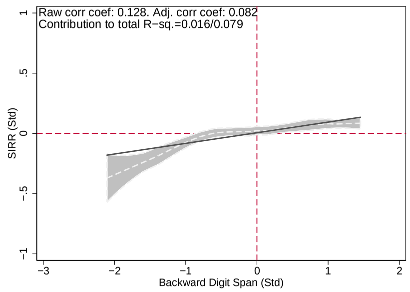


(b) Number of non-responded questions

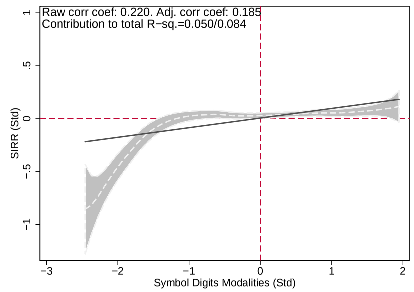


(c) SIRR

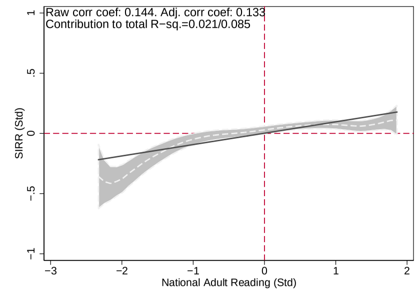
Figure 2: Distribution of required number of questions and of count of non-responded questions in SCQ (wave 12)



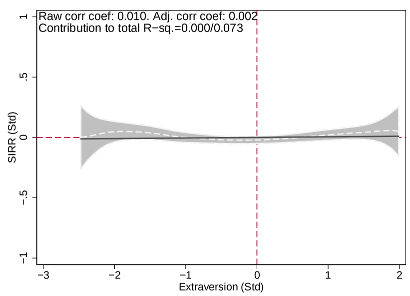
(a) BDS



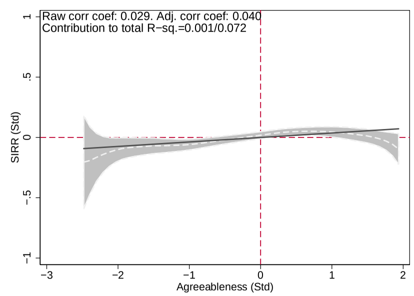
(b) SDM



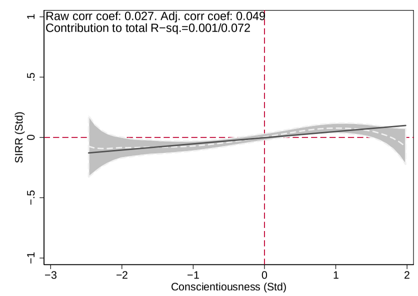
(c) NART 25



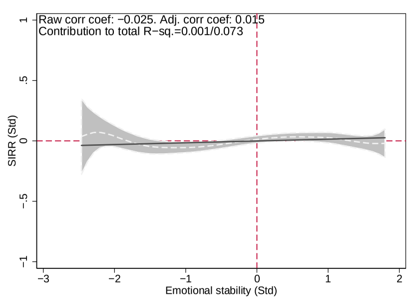
(d) Extraversion



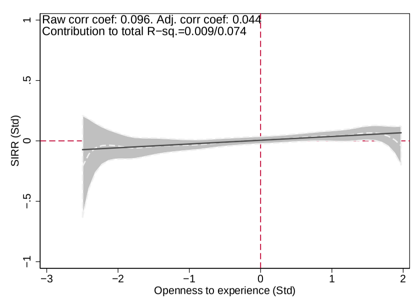
(e) Agreeableness



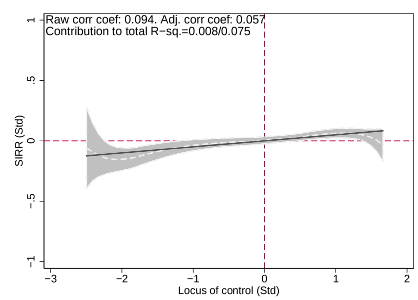
(f) Conscientiousness



(g) Emotional stability



(h) Openness to experience



(i) Locus of control

Figure 3: Relationship between survey item-response rate (SIRR) and cognitive or non-cognitive ability measures

s Note: Multivariate non-parametric (third degree) kernel regressions (white dashed line) and OLS regressions (grey solid line). Each model controls for a cubic polynomial in age, gender, education level, language background, geographic remoteness, employment status, number of young children in household, and interviewer characteristics. Full estimation results – including all ability measures simultaneously – are reported in Table B.1.

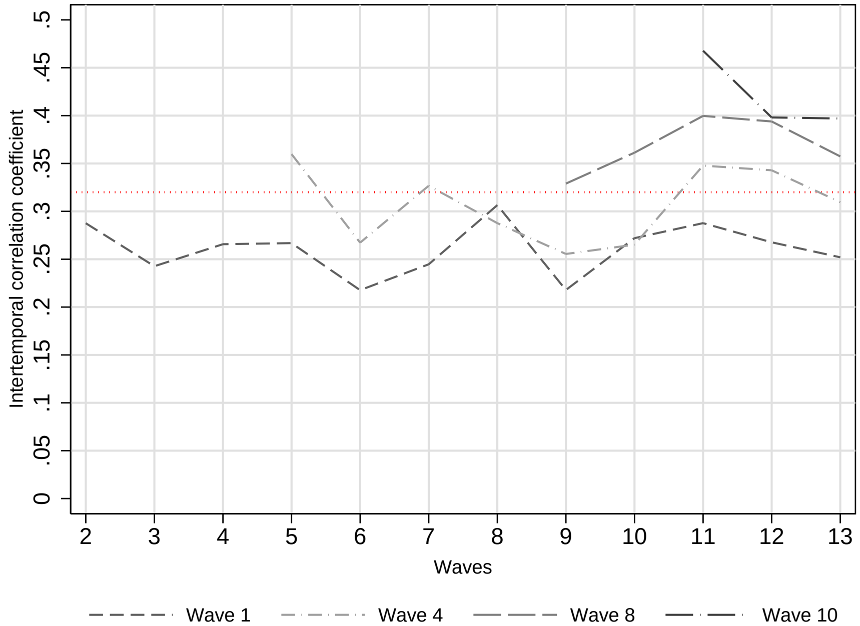


Figure 4: Average inter-temporal correlation coefficient (ITC) of SIRR across waves. Sample: individuals who stayed from wave 1 to 13 in the longitudinal survey (N=4,981)

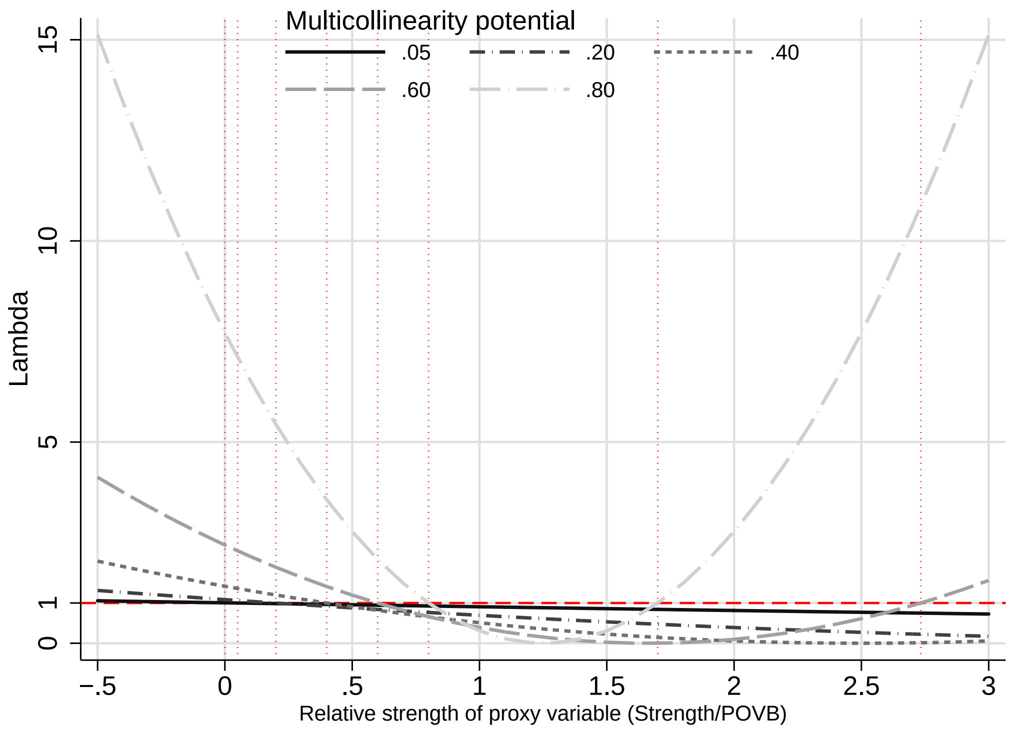


Figure 5: Simulation of lambda for five different values of the degree of the potential of multicollinearity problem (PMCP). The horizontal axis measures the relative strength of the proxy variable. A negative sign of the relative strength indicates a violation of the necessary condition of a valid proxy variable.

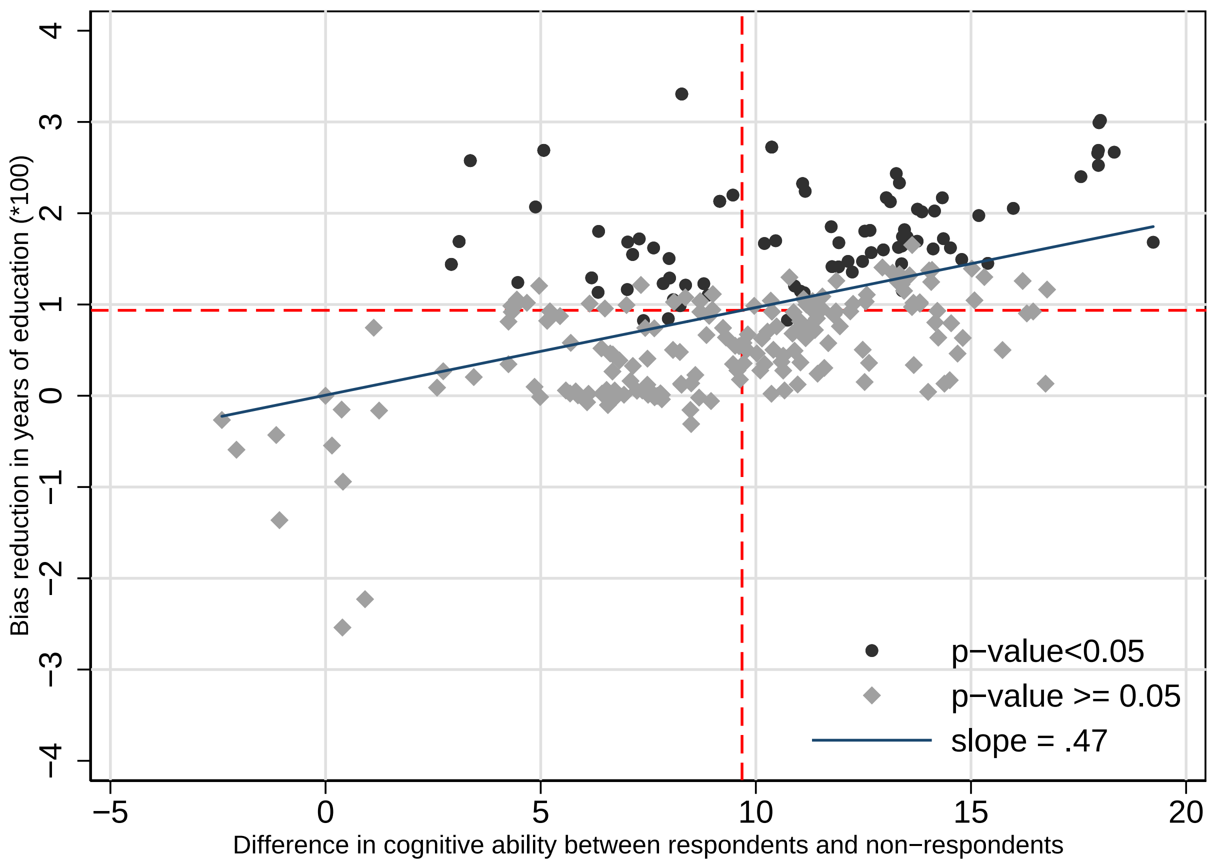


Figure 6: Scatter plot of percent bias reduction in coefficients on returns to education by difference in cognitive ability between respondents and non-respondents. Cognitive ability is measured with Symbol Digits Modalities test

Table 1: Summary statistics for variables used in analysis (N=8,666)

Variable definition	Mean	SD	Min	Max
Proxy variables derived from paradata				
Number of non-responded questions	3.59	10.37	0	184.00
Required count	204.60	23.23	151	266.00
SIRR (Refuse and Don't Know)	0.98	0.05	0.15	1.00
SIRR (Refuse)	0.98	0.05	0.15	1
Interviewer rating of understanding	4.77	0.47	1	5.00
Minutes spent on answering personal quest.	7.45	5.51	1	106
Minutes spent on answering HH quest.	41.08	12.56	13	128
Cognitive skill measures				
National Adult Reading Test	14.23	5.25	0	25
Backward Digits Span Test	3.96	1.40	0	7
Symbol Digits Modalities Test	48.81	12.94	0	104
Non-cognitive skill measures				
Locus of control				
Little control over life	2.76	1.56	1	7
No way to solve problems	2.56	1.60	1	7
Cannot change the important things	2.60	1.55	1	7
Feel helpless	2.45	1.53	1	7
Pushed around	2.59	1.60	1	7
Future depends on me	5.54	1.54	1	7
Can do just about everything	5.32	1.45	1	7
Big-Five personality traits				
Extraversion	4.43	1.08	1	7
Agreeableness	5.37	0.90	1	7
Conscientiousness	5.10	1.02	1	7
Emotional stability	5.26	1.05	1	7
Openness to experience	4.19	1.05	1	7
Variables used as controls and/or in wage regression				
Hourly wage	32.17	19.43	0.9	333
Australian	0.80	0.40	0	1
English-speaking background	0.10	0.30	0	1
Non-English speaking background	0.10	0.30	0	1
Age	47.85	17.84	18	100
Male	0.46	0.50	0	1
Years of education	12.35	2.23	8	17
Casual worker	0.12	0.32	0	1
Years spent with firm	5.23	8.16	0	58
Number of children ages 0 to 4	0.18	0.51	0	4
Number of children ages 5 to 9	0.17	0.49	0	3
Number of children ages 10 to 14	0.17	0.48	0	4
Major urban area	0.62	0.48	0	1

Continued on next page

Table 1: Summary statistics for variables used in analysis (N=8,666)

Variable definition	Mean	SD	Min	Max
Other urban area	0.22	0.42	0	1
Bounded locality	0.03	0.16	0	1
Rural balance	0.13	0.33	0	1
New South Wales	0.29	0.46	0	1
Victoria	0.25	0.43	0	1
Queensland	0.21	0.41	0	1
South Australia	0.09	0.28	0	1
Western Australia	0.10	0.29	0	1
Tasmania	0.03	0.17	0	1
Northern Territory	0.01	0.08	0	1
Australian Capital Territory	0.02	0.15	0	1

Table 2: Predictive power of SIRR

	Log hourly wages (-0.1 , 6.1)		High school (0 , 1)		Health (0 , 100)	
	Including SIRR (1)	Including CS & SIRR (2)	Including SIRR (3)	Including CS & SIRR (4)	Including SIRR (5)	Including CS & SIRR (6)
SIRR (Std)	0.0262**	0.0155	0.0297***	0.00573	0.606**	0.313
	(0.0114)	(0.0115)	(0.00527)	(0.00512)	(0.259)	(0.262)
Symbol Digits Modalities (Std)		0.0485***		0.0345***		1.744***
		(0.0103)		(0.00605)		(0.283)
National Adult Reading (Std)		0.0425***		0.130***		0.571**
		(0.00918)		(0.00538)		(0.262)
Backward Digit Span (Std)		0.00262		0.00778		-0.259
		(0.00719)		(0.00481)		(0.222)
Observations	5780	5780	8666	8666	8588	8588
Adjusted R^2	1110	1165	0.123	0.203	0.252	0.256

Note: Dependent variables are log hourly wages (models 1,2), high school degree (models 3,4) and SF-36 general health component (Models 5, 6). Models (1) and (2) are estimated with a Tobit sample selection model, with left censored observations set to missing. Exclusion restriction for selection equation are number of children in age bracket 0-4, 5-10, 11-14. The Mills ratio is 0.05 (t-value of 1.1) for model (1) and 0.08 (t-value of 1.6) for model (2). Estimation samples in models (1) and (2) are based on working age population, ages 24 to 64. All other models are estimated with OLS with no age restriction. Models 3 and 4 are linear probability models. All regression models control for gender, age, years of education (except models (2) and (3)), a dummy for casual worker, years of tenure with current firm and its square, state of residence dummies, region dummies (other urban, bounded locality, rural balance), language background, Big 5 personality traits. Models (3) and (4) additionally control for labor force status (unemployed, out of the labor force). Standard errors are reported in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3: Estimates of λ and percent bias reduction

	Education (Years)	Locus of control (Std)
Panel A: Partial correlation coefficients		
1. Strength of proxy variable: $r_{MP\|Z}$	0.117	0.117
2. Omitted-variable potential: $r_{XM\|Z}$	0.144	0.070
3. Multicollinearity potential: $r_{XP\|Z}$	0.128	0.040
4. Relative strength of proxy variable: $\frac{r_{MP\|Z}}{r_{XM\|Z}}$	0.810	1.675
5. Ratio of the squared biases: λ^{a}	0.830	0.873
Panel B: Bias reduction in percent of omitted-variable bias[b]		
1. Linear SIRR (N=5,780)	8.9***	6.5***
2. Non-linear SIRR[c] (N=5,780)	10.1***	7.3***
3. SIRR based on refusals only (N=5,780)	7.9***	5.6***
4. Age range 20-69 sample (N=7,215)	10.0***	8.3***
5. Drop if SIRR $<=$ 0.4 (N=5,771)	10.3***	7.2***
6. Drop Non-Engl. background (N=5,183)	10.2***	6.3***
7. Alternative CA: BDS (N=5,780)	5.4***	11.4***
8. Alternative CA: NART 25 (N=5,780)	2.7***	5.6***

Note: $r_{MP|Z}$ $r_{XP|Z}$ and $r_{XM|Z}$ are the partial correlation coefficients of cognitive skills M and the proxy variable P (strength of proxy), of the education or locus of control variable X and the proxy variable P (multicollinearity potential), and of the education or non-cognitive skill variable X and cognitive skills M (omitted-variable potential), netting out the effect of all other control variables used in the wage regression model. [a] λ measures the squared relative bias, and stars measure the null hypothesis that the ratio is equal to one, against the alternative hypothesis that the ratio is smaller than 1. [b] Bias reduction is calculated as the percentage change in the omitted-variable bias, calculated as: $\frac{(OVB-IPB)}{OVB} \times 100$. Standard errors for λ and the bias reduction are calculated with the delta method. [c] Non-linear measure of SIRR is captured by four dummy variables that indicate non-response rates within the 5th percentile, between the 5th and 10th percentile, between the 10th and 25th percentile, and above the 25th percentile. The comparison group is zero non-response. * p$<$0.10 ** p$<$0.05, *** p$<$0.01.

Table 4: Relative performance of SIRR as proxy variable for cognitive skills against alternative proxy variables

Proxy variable (Std)	Years Education		Locus of control	
	λ^{a}	BR^{b}	λ	BR
1. Survey item response	0.83	8.9***	0.87	6.5***
2. Interviewer rating understanding	0.88	6.1***	0.77	12.5***
3. Minutes spent on personal questionnaire	0.99	0.53	1.00	-0.05
4. Minutes spend on household questionnaire	0.98	1.2	0.90	4.9***
5. Days elapsed until SCQ returned	1.23	-11.2***	1.60	-26.5***

Note: All proxy variables are standardised to mean 0 and standard deviation 1. Alternative proxy variables derived from para data are: 2. Interviewer rating of the understanding of questions; 3. Minutes spent on completing personal questionnaires; 4. Minutes spent on completing the household questionnaire; 5. Days elapsed between receiving and completing the self-completion questionnaires. ^{a} λ measures the squared relative bias, and stars measure the null hypothesis that the ratio is equal to one, against the alternative hypothesis that the ratio is smaller than 1. ^{a} Bias reduction is calculated as the percentage change in the omitted-variable bias, calculated as: $\frac{(OVB-IPB)}{OVB} \times 100$.

Table 5: Bias-adjusted wage returns to education under various assumptions

	True effect β^* (Full model)	β^*+OVB (Omitting Cogn. Ability)	Oster (2017) method		β^*+IPB (Including SIRR)
			$R^2_{max} = 1.3R^2$ $\delta = 1$	$R^2_{max} = 1$ $\delta = 1$	
	(1)	(2)	(3)	(4)	(5)
β	0.0763	0.0790	0.0741	-0.5919	0.0782

Note: All models are estimated with 4,024 working individuals with positive hourly wages in Wave 12. The dependent variable is log hourly wages. Reported are the coefficients (labeled "β") for years of education from OLS models. All regression models control for gender, age, years of education, a dummy for casual worker, years of tenure with current firm and its square, state of residence dummies, region dummies (other urban, bounded locality, rural balance), language background, Big-5 personality traits. Model (1) additionally controls for cognitive ability and shows the true effect of education (β^*) under the assumption that all relevant variables are controlled for. Model (2) shows the effect of education when cognitive skills are omitted and thus the estimate suffers from omitted-variable bias (OVB). Models (3) and (4) show the bias adjusted effects as suggested by Oster (2017). Model (3) shows the effect assuming that the degree of selection on unobservables is equal the degree of selection on observables ($\delta = 1$) and the maximum R-squared that can be achieved is 1.3 times the R-Squared of Model (2) ($R_{max} = 1.3 * 0.22 = 0.29$) as suggested in Oster (2017). Model (4) assumes $\delta = 1$, but a maximum R-squared of 1. Model (5) controls for SIRR and therefore shows an effect that potentially includes an imperfect proxy bias. The calculation of the bias adjusted effects are based on the STATA command psacalc written by Emily Oster and available through ssc.

A Response behavior on self-completion questionnaire in unrestricted sample

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:B23m How long ago life event happened - Jailed - no answer	0.816	0	0.816	1.801	174
SCQ:B23c How long ago life event happened - Reconciled - no answer	0.578	0	0.578	1.045	287
SCQ:B23h How long ago life event happened - Death of spouse/child - no answer	0.575	0	0.575	2.069	322
SCQ:B23o How long ago life event happened - Retired - no answer	0.499	0	0.499	1.037	557
SCQ:B23v How long ago life event happened - Home destroyed - no answer	0.483	0	0.483	0.876	362
SCQ:B23n How long ago life event happened - Family member jailed - no answer	0.473	0	0.473	1.677	391
SCQ:B23k How long ago life event happened - Victim of violence - no answer	0.467	0	0.467	2.333	407
SCQ:B23t How long ago life event happened - Financial worsening - no answer	0.371	0	0.371	1.669	590
SCQ:D3b Workplace entitlements: Paid paternity leave	0.335	0.326	0.008	-2.539	9485
SCQ:B23p How long ago life event happened - Fired - no answer	0.329	0	0.329	1.568	645
SCQ:B23s How long ago life event happened - Financial improvement - no answer	0.311	0	0.311	0.924	601
SCQ:B23a How long ago life event happened - Married - no answer	0.289	0	0.289	1.163	494
SCQ:B23r How long ago life event happened - Promoted - no answer	0.266	0	0.266	0.151	1121
SCQ:B23e How long ago life event happened - Gained family - no answer	0.259	0	0.259	0.501	746
SCQ:D3a Workplace entitlements: Paid maternity leave	0.255	0.248	0.008	-2.229	9485
SCQ:B23b How long ago life event happened - Separated - no answer	0.254	0	0.254	1.975	700
SCQ:B23l How long ago life event happened - Victim of property crime - no answer	0.248	0	0.248	1.348	738
SCQ:D3c Workplace entitlements: Special leave for caring for family members	0.216	0.208	0.008	-0.546	9485
SCQ:D3g Child care facilities or subsidised child care expenses	0.211	0.201	0.01	-1.363	9485
SCQ:B23d How long ago life event happened - Pregnancy - no answer	0.205	0	0.205	0.170	980
SCQ:B23j How long ago life event happened - Death of friend - no answer	0.205	0	0.205	1.010	2029
SCQ:B23f How long ago life event happened - Injury to self - no answer	0.191	0	0.191	0.452	1594
SCQ:B23g How long ago life event happened - Injury to relative/family - no answe	0.153	0	0.153	0.919	2676
SCQ:D3d Workplace entitlements: Permanent part-time work	0.151	0.142	0.009	-0.941	9485
SCQ:D3e Workplace entitlements: Home-based work	0.148	0.138	0.01	-0.152	9485
SCQ:B23i How long ago life event happened - Death of other rel - no answer	0.146	0	0.146	1.042	2180
SCQ:B23q How long ago life event happened - Changed jobs - no answer	0.146	0	0.146	0.277	2159
SCQ:C6b Financial risk - assumed	0.139	0	0.139	1.719	2755
SCQ:B23u How long ago life event happened - Moved - no answer	0.123	0	0.123	0.681	2635
SCQ:B12b Neighbourhood: Neighbours doing things together	0.117	0.11	0.006	-0.430	15380
SCQ:B24i Hours per week - Caring for disabled spouse/relative	0.104	0	0.104	1.440	15380
SCQ:B24i Minutes per week - Caring for disabled spouse/relative	0.104	0	0.104	1.440	15380
SCQ:B12a Neighbourhood: Neighbours helping each other out	0.1	0.095	0.005	-0.592	15380
SCQ:B13d Satisfaction with: Relationship with step children	0.099	0	0.099	2.199	2191
SCQ:B24g Hours per week - Playing with other children	0.098	0	0.098	1.690	15380
SCQ:B24g Minutes per week - Playing with other children	0.098	0	0.098	1.690	15380
SCQ:B24h Hours per week - Volunteer/Charity work	0.093	0	0.093	2.576	15380
SCQ:B24h Minutes per week - Volunteer/Charity work	0.093	0	0.093	2.576	15380
SCQ:B12j Neighbourhood: Burglary and theft	0.092	0.087	0.006	-0.265	15380
SCQ:B13g Satisfaction with: Relationship with step parents	0.088	0	0.088	2.054	2683

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:D3f Workplace entitlements: Flexible start/finish times	0.085	0.077	0.008	0.746	9485
SCQ:B3 Number of cigarettes usually smoked each week	0.077	0	0.077	1.027	2933
SCQ:B24f Hours per week - Playing with your children	0.076	0	0.076	2.689	15380
SCQ:B24f Minutes per week - Playing with your children	0.076	0	0.076	2.689	15380
SCQ:C2b Could not pay the mortgage or rent on time	0.062	0	0.062	1.205	15380
SCQ:C2c Pawned or sold something	0.06	0	0.06	1.019	15380
SCQ:C2e Was unable to heat home	0.06	0	0.06	1.241	15380
SCQ:C2g Asked for help from welfare/community organisations	0.06	0	0.06	0.985	15380
SCQ:C2d Went without meals	0.059	0	0.059	1.054	15380
SCQ:C2a Could not pay electricity, gas or telephone bills on time	0.057	0	0.057	0.918	15380
SCQ:C2f Asked for financial help from friends or family	0.057	0	0.057	1.021	15380
SCQ:C9i Monthly household expenditure - Childrens clothing and footwear	0.052	0	0.052	1.685	10477
SCQ:C9d Weekly household expenditure - Public transport and taxis	0.05	0	0.05	1.133	10477
SCQ:C7e Household decisions - The way children are raised	0.048	0	0.048	2.169	6714
SCQ:C9c Weekly household expenditure - Cigarettes and tobacco	0.047	0	0.047	0.520	10477
SCQ:B24b Hours per week - Travelling to and from a place of paid employment	0.046	0	0.046	2.131	15380
SCQ:B24b Minutes per week - Travelling to and from a place of paid employment	0.046	0	0.046	2.131	15380
SCQ:C9h Monthly household expenditure - Womens clothing and footwear	0.043	0	0.043	1.504	10477
SCQ:B25g Computer use - Computers have helped me reach my occupational (career)	0.042	0	0.042	2.401	15380
SCQ:B13h Satisfaction with: Relationship with most recent former spouse/partner	0.041	0	0.041	0.930	4689
SCQ:B24e Hours per week - Outdoor tasks	0.041	0	0.041	1.548	15380
SCQ:B24e Minutes per week - Outdoor tasks	0.041	0	0.041	1.548	15380
SCQ:C9g Monthly household expenditure - Mens clothing and footwear	0.041	0	0.041	1.053	10477
SCQ:C9p Annual household expenditure - Home repairs/renovations/maintenance	0.041	0	0.041	0.994	10477
SCQ:C9r Annual household expenditure - Education fees	0.041	0	0.041	1.620	10477
SCQ:B24c Hours per week - Household errands	0.04	0	0.04	3.306	15380
SCQ:B24c Minutes per week - Household errands	0.04	0	0.04	3.306	15380
SCQ:C7d Household decisions - The number of hours your partner/spouse spends in	0.04	0	0.04	2.024	8298
SCQ:B25c Computer use - Computers have made it possible for me to get more done	0.039	0	0.039	2.689	15380
SCQ:B25e Computer use - Computers have helped me to learn new skills other than	0.039	0	0.039	2.992	15380
SCQ:B25f Computer use - Computers have helped me to communicate with people	0.039	0	0.039	2.524	15380
SCQ:B25a Computer use - My level of computer skills meets my present needs	0.038	0	0.038	3.018	15380
SCQ:B25d Computer use - Computers have made it easier for me to get useful infor	0.038	0	0.038	2.658	15380
SCQ:C9b Weekly household expenditure - Alcohol	0.038	0	0.038	0.824	10477
SCQ:B25b Computer use - I feel comfortable installing or upgrading computer soft	0.037	0	0.037	2.669	15380
SCQ:B12h Neighbourhood: People being hostile and aggressive	0.036	0.029	0.006	0.269	15380
SCQ:B12i Neighbourhood: Vandalism and deliberate damage to property	0.035	0.03	0.005	-0.163	15380
SCQ:B14a The way child care tasks are divided between you and your partner	0.035	0	0.035	1.258	4927
SCQ:B24a Hours per week - Paid employment	0.035	0	0.035	2.325	15380
SCQ:B24a Minutes per week - Paid employment	0.035	0	0.035	2.325	15380
SCQ:C9m Annual household expenditure - Fees paid to health practitioners	0.034	0	0.034	1.229	10477
SCQ:C9l Annual household expenditure - Other insurance (home/contents/motor vehi	0.033	0	0.033	1.212	10477
SCQ:C9o Annual household expenditure - Electricity bills, gas bills and other he	0.033	0	0.033	1.164	10477
SCQ:B12e Neighbourhood: Homes and gardens in bad condition	0.031	0.025	0.006	0.812	15380

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:C9q Annual household expenditure - Motor vehicle repairs/maintenance	0.031	0	0.031	1.078	10477
SCQ:B13e Satisfaction with: Children in household get along with each other	0.03	0	0.03	1.220	6195
SCQ:C9n Annual household expenditure - Medicines, prescriptions, pharmaceuticals	0.03	0	0.03	0.739	10477
SCQ:B24d Hours per week - Housework	0.029	0	0.029	2.724	15380
SCQ:B24d Minutes per week - Housework	0.029	0	0.029	2.724	15380
SCQ:C9e Weekly household expenditure - Meals eaten out	0.029	0	0.029	0.987	10477
DV: [SCQ:B] Height in centimetres	0.029	0	0.029	0.346	15380
SCQ:C7c Household decisions - The number of hours you spend in paid work	0.028	0	0.028	1.494	10728
SCQ:C9j Monthly household expenditure - Telephone rent, calls and internet charg	0.028	0	0.028	1.228	10477
SCQ:C9f Monthly household expenditure - Motor vehicle fuel	0.027	0	0.027	1.102	10477
SCQ:C9k Annual household expenditure - Private health insurance	0.026	0	0.026	1.292	10477
SCQ:B12g Neighbourhood: Teenagers hanging around on the streets	0.025	0.019	0.006	0.580	15380
SCQ:B17c To what extent has your relationship met your original expectations	0.025	0	0.025	0.888	9866
DV: [SCQ:B] Weight in kilograms	0.025	0	0.025	0.821	15380
SCQ:C3a Difficulty in raising (waves 1-8) $2,000(waves9+)$3,000 for an emerge	0.024	0	0.024	0.639	15380
SCQ:B17a How good is your relationship compared to most	0.023	0	0.023	1.129	9866
SCQ:B17d How much do you love your spouse/partner	0.023	0	0.023	0.760	9866
SCQ:B17f How well does your spouse meet your needs	0.023	0	0.023	0.765	9866
SCQ:B18 Share of work around the house	0.023	0	0.023	1.203	15380
SCQ:C4 Savings habits	0.023	0	0.023	0.925	15380
SCQ:C6a Financial risk prepared to take	0.023	0	0.023	1.698	15380
SCQ:B17b How often do you wish you had not got married / got into this relations	0.022	0	0.022	0.952	9866
SCQ:B17e How many problems are there in your relationship	0.022	0	0.022	0.848	9866
SCQ:C7g Household decisions - Savings, investment and borrowing	0.022	0	0.022	1.852	13562
SCQ:C9a Weekly household expenditure - Groceries	0.022	0	0.022	0.845	10476
SCQ:E1 Has parenting responsibilities for any children aged 17 years or less	0.022	0	0.022	1.451	15380
SCQ:C3b How would obtain (waves 1-8) $2,000(waves9+)$3,000 - No Answer	0.022	0	0.022	1.290	13385
SCQ:C7f Household decisions - Social life and leisure activities	0.021	0	0.021	1.472	14129
SCQ:B13c Satisfaction with: Partners relationship with children	0.019	0	0.019	1.803	8181
SCQ:C5 Most important when planning savings and spending	0.019	0	0.019	0.999	15380
SCQ:C7b Household decisions - Making large household purchases	0.018	0	0.018	1.060	13845
SCQ:B12c Neighbourhood: Traffic noise	0.017	0.007	0.01	0.873	15380
SCQ:B12f Neighbourhood: Rubbish and litter lying around	0.017	0.011	0.006	0.956	15380
SCQ:B14b The way household tasks are divided between you and your partner	0.017	0	0.017	1.621	10028
SCQ:C7a Household decisions - Managing day-to-day spending and paying bills	0.016	0	0.016	1.114	14234
SCQ:B12d Neighbourhood: Noise from airplanes, trains or industry	0.015	0.009	0.007	1.214	15380
SCQ:B13f Satisfaction with: Relationship with parents	0.015	0	0.015	1.682	10811
SCQ:B19 How often get together socially with friends/relatives not living with y	0.015	0	0.015	1.415	15380
SCQ:E4j Work-family balance: My work has a positive effect on my children	0.015	0	0.015	0.025	3510
SCQ:E4m Work-family balance: Worry about children while at work	0.015	0	0.015	-0.102	3510
SCQ:B5 Standard drinks usually have per day	0.014	0	0.014	0.277	12538
SCQ:B21c Achievement motivation - I feel uneasy about undertaking a task if I a	0.014	0	0.014	1.332	15380
SCQ:E4c Work-family balance: Work makes me feel competent	0.014	0	0.014	0.026	3510
SCQ:E4e Work-family balance: Having both work and family responsibilities challe	0.014	0	0.014	0.121	3510

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:E4f Work-family balance: Time working less enjoyable/more pressured	0.014	0	0.014	-0.025	3510
SCQ:E4i Work-family balance: Working makes me feel good about myself, which is g	0.014	0	0.014	0.058	3510
SCQ:E4k Work-family balance: Better appreciate time spent with children	0.014	0	0.014	-0.002	3510
SCQ:E4l Work-family balance: Working makes me a better parent	0.014	0	0.014	0.021	3510
SCQ:E4p Work-family balance: Thinking about the children interferes with my perf	0.014	0	0.014	0.058	3510
SCQ:A3i Physical Functioning: Walking 100 metres	0.013	0	0.013	0.627	15380
SCQ:A4c Role-physical: Were limited in the kind of work	0.013	0	0.013	1.721	15380
SCQ:B13a Satisfaction with: Partner	0.013	0	0.013	0.795	10919
SCQ:B13b Satisfaction with: Children	0.013	0	0.013	2.240	10095
SCQ:B21d Achievement motivation - When confronted by a difficult problem, I pre	0.013	0	0.013	1.259	15380
SCQ:B21g Achievement motivation - Even when nobody is watching, I feel anxious	0.013	0	0.013	1.356	15380
SCQ:B21h Achievement motivation - I am attracted to tasks that allow me to test	0.013	0	0.013	1.412	15380
SCQ:B21i Achievement motivation - I start feeling anxious if I do not understan	0.013	0	0.013	1.146	15380
SCQ:B23r Life events in past year: Promoted at work	0.013	0	0.013	1.298	15380
SCQ:D1 Is in paid work	0.013	0	0.013	1.393	15380
SCQ:E4g Work-family balance: Miss out on home/family activities	0.013	0	0.013	0.005	3510
SCQ:E4h Work-family balance: Family time less enjoyable/more pressured	0.013	0	0.013	0.055	3510
SCQ:E4n Work-family balance: Too little time or energy to be aspirational parent	0.013	0	0.013	0.064	3510
SCQ:E4o Work-family balance: Miss out on the rewarding aspects of being parent	0.013	0	0.013	0.023	3510
SCQ:A3e Physical Functioning: Climbing one flight of stairs	0.012	0	0.012	1.037	15380
SCQ:B15 Sexual Identity	0.012	0	0.012	0.360	15380
SCQ:B21b Achievement motivation - I like situations where I can find out how ca	0.012	0	0.012	2.015	15380
SCQ:B21e Achievement motivation - I am afraid of tasks that I cannot work out o	0.012	0	0.012	2.435	15380
SCQ:B21f Achievement motivation - I enjoy situations that make use of my abilit	0.012	0	0.012	1.812	15380
SCQ:E4b Work-family balance: Work gives my life more variety	0.012	0	0.012	0.011	3510
SCQ:E4d Work-family balance: Have to turn down work/opportunities	0.012	0	0.012	0.011	3510
SCQ:A4d Role-physical: Had difficulty performing work or other activities	0.011	0	0.011	0.337	15380
SCQ:A5c Role-emotional: Didnt do work/other activities as carefully as usual	0.011	0	0.011	0.637	15380
SCQ:A11c Health: Expect my health to get worse	0.011	0	0.011	0.355	15380
SCQ:A11d Health: My health is excellent	0.011	0	0.011	0.720	15380
SCQ:B4 Drink alcohol	0.011	0	0.011	0.098	15380
SCQ:B20f There is someone who can always cheer me up when Im down	0.011	0	0.011	1.373	15380
SCQ:B21a Achievement motivation - In difficult situations where a lot depends o	0.011	0	0.011	1.447	15380
SCQ:B23g Life events in past year: Serious injury/illness to family member	0.011	0	0.011	1.406	15380
SCQ:B23k Life events in past year: Victim of physical violence	0.011	0	0.011	1.599	15380
SCQ:E4a Work-family balance: Work makes me a more rounded person	0.011	0	0.011	0.011	3510
SCQ:A3a Physical Functioning: Vigorous activities	0.01	0	0.01	0.242	15380
SCQ:A3d Physical Functioning: Climbing several flights of stairs	0.01	0	0.01	0.753	15380
SCQ:A3f Physical Functioning: Bending kneeling or stooping	0.01	0	0.01	0.666	15380
SCQ:A3h Physical Functioning: Walking half a kilometre	0.01	0	0.01	0.502	15380
SCQ:A4b Role-physical: Accomplished less than would like	0.01	0	0.01	0.634	15380
SCQ:A11a Health: Get sick a little easier than other people	0.01	0	0.01	0.576	15380
SCQ:A11b Health: As healthy as anybody I know	0.01	0	0.01	1.088	15380
SCQ:B16 Married or in a long-term relationship	0.01	0	0.01	1.007	15380

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:B20d I dont have anyone that I can confide in	0.01	0	0.01	1.149	15380
SCQ:B20e I have no one to lean on in times of trouble	0.01	0	0.01	1.018	15380
SCQ:B20g I often feel very lonely	0.01	0	0.01	1.152	15380
SCQ:B23e Life events in past year: Birth/adoption of new child	0.01	0	0.01	1.107	15380
SCQ:B23f Life events in past year: Serious personal injury/illness	0.01	0	0.01	0.953	15380
SCQ:B23h Life events in past year: Death of spouse or child	0.01	0	0.01	2.126	15380
SCQ:B23p Life events in past year: Fired or made redundant	0.01	0	0.01	2.170	15380
SCQ:B23q Life events in past year: Changed jobs	0.01	0	0.01	1.626	15380
SCQ:A3c Physical Functioning: Lifting or carrying groceries	0.009	0	0.009	0.743	15380
SCQ:A3g Physical Functioning: Walking more than one kilometre	0.009	0	0.009	0.543	15380
SCQ:A3j Physical Functioning: Bathing or dressing yourself	0.009	0	0.009	0.743	15380
SCQ:A4a Role-physical: Cut down the amount of time spent on work or other activi	0.009	0	0.009	0.924	15380
SCQ:A5a Role-emotional: Cut down the amount of time spent on work/other activiti	0.009	0	0.009	0.902	15380
SCQ:A5b Role-emotional: Accomplished less than would like	0.009	0	0.009	1.301	15380
SCQ:B2 Smokes cigarettes or other tobacco products	0.009	0	0.009	0.124	15380
SCQ:B20a People dont come to visit me as often as I would like	0.009	0	0.009	1.378	15380
SCQ:B20b I often need help from other people but cant get it	0.009	0	0.009	0.803	15380
SCQ:B20c I seem to have a lot of friends	0.009	0	0.009	0.631	15380
SCQ:B20i When somethings on my mind, just talking with the people I know can mak	0.009	0	0.009	1.472	15380
SCQ:B22b Cognitive activities - Read books	0.009	0	0.009	0.703	15380
SCQ:B22d Cognitive activities - Do puzzles (like crosswords or Sudoku) or play	0.009	0	0.009	0.177	15380
SCQ:B22f Cognitive activities - Write (e.g., reports, letters, stories or journ	0.009	0	0.009	0.985	15380
SCQ:B22g Cognitive activities - Attend educational lectures or courses	0.009	0	0.009	0.579	15380
SCQ:B23b Life events in past year: Separated from spouse	0.009	0	0.009	1.243	15380
SCQ:B23c Life events in past year: Got back together with spouse	0.009	0	0.009	1.024	15380
SCQ:B23d Life events in past year: Pregnancy	0.009	0	0.009	1.009	15380
SCQ:B23i Life events in past year: Death of close relative/family member	0.009	0	0.009	1.320	15380
SCQ:B23j Life events in past year: Death of a close friend	0.009	0	0.009	1.247	15380
SCQ:B23m Life events in past year: Detained in jail	0.009	0	0.009	1.641	15380
SCQ:B23n Life events in past year: Close family member detained in jail	0.009	0	0.009	1.693	15380
SCQ:B23o Life events in past year: Retired from the workforce	0.009	0	0.009	1.741	15380
SCQ:B23s Life events in past year: Major improvement in finances	0.009	0	0.009	1.652	15380
SCQ:B23t Life events in past year: Major worsening in finances	0.009	0	0.009	1.820	15380
SCQ:B23v A weather related disaster (flood, bushfire, cyclone) damaged or destro	0.009	0	0.009	1.744	15380
SCQ:C1 Prosperity given current needs and financial responsibilities	0.009	0	0.009	0.942	15380
SCQ:A3b Physical Functioning: Moderate activities	0.008	0	0.008	0.915	15380
SCQ:A9g Vitality: Felt worn out	0.008	0	0.008	0.922	15380
SCQ:B20h I enjoy the time I spend with the people who are important to me	0.008	0	0.008	1.031	15380
SCQ:B20j When I need someone to help me out, I can usually find someone	0.008	0	0.008	0.924	15380
SCQ:B22e Cognitive activities - Play other games, such as board games or comput	0.008	0	0.008	0.761	15380
SCQ:B22h Cognitive activities - Arts or crafts or other artistic activities (e.	0.008	0	0.008	0.672	15380
SCQ:B22i Cognitive activities - Go to museums or art galleries	0.008	0	0.008	0.815	15380
SCQ:B23l Life events in past year: Victim of a property crime	0.008	0	0.008	1.609	15380
SCQ:B23u Life events in past year: Changed residence	0.008	0	0.008	2.045	15380

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:F1 Self completion questionnaire: sex	0.008	0	0.008	0.548	15380
SCQ:A1 Self-assessed health	0.007	0	0.007	0.205	15380
SCQ:A9f Mental Health: Felt down	0.007	0	0.007	0.704	15380
SCQ:A10 Social functioning: Physical/emotional problems interfered with social a	0.007	0	0.007	0.407	15380
SCQ:B8 Currently an active member of a sporting/hobby/community based club or as	0.007	0	0.007	-0.058	15380
SCQ:B22a Cognitive activities - Watch television programs or movies	0.007	0	0.007	0.480	15380
SCQ:B22c Cognitive activities - Read magazines or newspapers	0.007	0	0.007	0.464	15380
SCQ:B23a Life events in past year: Got married	0.007	0	0.007	0.973	15380
SCQ:E2b I often feel tired, worn out or exhausted from meeting the needs of my c	0.007	0	0.007	0.133	4684
SCQ:E2c I feel trapped by my responsibilities as a parent	0.007	0	0.007	0.462	4684
SCQ:E2d I find that taking care of my children is much more work than pleasure	0.007	0	0.007	0.133	4684
SCQ:A9b Mental Health: Been a nervous person	0.006	0	0.006	0.828	15380
SCQ:A9c Mental Health: Felt so down in the dumps nothing could cheer you up	0.006	0	0.006	0.348	15380
SCQ:A9e Vitality: Have a lot of energy	0.006	0	0.006	0.504	15380
SCQ:A9h Mental Health: Been a happy person	0.006	0	0.006	0.463	15380
SCQ:E2a Being a parent is harder than I thought it would be	0.006	0	0.006	0.043	4684
SCQ:E3 Do fair share of looking after children	0.006	0	0.006	0.504	4684
SCQ:F2 Self completion questionnaire: age group	0.006	0	0.006	0.347	15380
SCQ:A2 Health compared to one year ago	0.005	0	0.005	0.132	15380
SCQ:A9a Vitality: Feel full of life	0.005	0	0.005	0.371	15380
SCQ:A9d Mental Health: Felt calm and peaceful	0.005	0	0.005	0.365	15380
SCQ:A9i Vitality: Felt tired	0.005	0	0.005	0.505	15380
SCQ:B11 Preference to continue living in area	0.005	0	0.005	-0.071	15380
SCQ:D2e The company I work for will still be in business 5 years from now	0.005	0	0.005	-0.017	9485
SCQ:B1 How often participate in physical activity	0.004	0	0.004	0.048	15380
SCQ:B9 How often feel rushed or pressed for time	0.004	0	0.004	-0.156	15380
SCQ:B10 Spare time that dont know what to do with	0.004	0	0.004	-0.310	15380
SCQ:D2d I have a secure future in my job	0.004	0	0.004	-0.020	9485
SCQ:D2f I worry about the future of my job	0.004	0	0.004	0.009	9485
SCQ:D2k I have a lot of say about what happens on my job	0.004	0	0.004	0.266	9485
SCQ:D2o I can decide when to take a break	0.004	0	0.004	0.228	9485
SCQ:D2p My job requires me to do the same things over and over again	0.004	0	0.004	0.089	9485
SCQ:D2q My job provides me with a variety of interesting things to do	0.004	0	0.004	0.120	9485
SCQ:D2r My job requires me to take initiative	0.004	0	0.004	0.136	9485
SCQ:A6 Social functioning: Has physical/emotional health interfered with normal	0.003	0	0.003	0.493	15380
SCQ:A7 Bodily pain in last 4 weeks	0.003	0	0.003	0.387	15380
SCQ:A8 How much did pain interfere with normal work	0.003	0	0.003	0.548	15380
SCQ:D2c I get paid fairly for the things I do in my job	0.003	0	0.003	0.438	9485
SCQ:D2g My job is complex and difficult	0.003	0	0.003	0.327	9485
SCQ:D2h My job often requires me to learn new skills	0.003	0	0.003	0.305	9485
SCQ:D2i I use many of my skills and abilities in my current job	0.003	0	0.003	0.047	9485
SCQ:D2j I have a lot of freedom to decide how I do my own work	0.003	0	0.003	0.059	9485
SCQ:D2l I have a lot of freedom to decide when I do my work	0.003	0	0.003	-0.013	9485
SCQ:D2m I have a lot of choice in deciding what I do at work	0.003	0	0.003	0.276	9485

Continued on next page

Table A.1: Questions and non-response

Variable in SCQ: Ordered by largest to smallest non-response proportion	NR (%)	DK (%)	Ref (%)	Bias reduction Returns to Education	N
SCQ:D2n My working times can be flexible	0.003	0	0.003	0.022	9485
SCQ:D2s I have to work fast in my job	0.003	0	0.003	-0.039	9485
SCQ:D2t I have to work very intensely in my job	0.003	0	0.003	-0.013	9485
SCQ:D2a My job is more stressful than I had ever imagined	0.002	0	0.002	0.022	9485
SCQ:D2b I fear that the amount of stress in my job will make me physically ill	0.002	0	0.002	0.058	9485
SCQ:D2u I dont have enough time to do everything in my job	0.002	0	0.002	0.162	9485
SCQ:C8 Has responsibility for payment of household bills such as electricity, ga	0	0	0	0.000	15380

B Testing the economic model of survey item-response behavior

Table B.1: Determinants of survey item-response behavior

Dependent variable SIRR (Std)	No further controls	Individ. charact.	Interviewer rating	Interviewer FE
Locus of control (Std)	0.045***	0.030**	0.027**	0.025**
	(0.012)	(0.012)	(0.012)	(0.012)
Extraversion (Std)	-0.021**	-0.013	-0.014	-0.012
	(0.011)	(0.011)	(0.011)	(0.011)
Agreeableness (Std)	0.032***	0.045***	0.045***	0.045***
	(0.011)	(0.012)	(0.012)	(0.012)
Conscientiousness (Std)	0.023**	0.029**	0.028**	0.029**
	(0.011)	(0.012)	(0.011)	(0.012)
Emotional stability (Std)	-0.035***	-0.026**	-0.026**	-0.024*
	(0.012)	(0.012)	(0.012)	(0.012)
Openness to experience (Std)	0.031***	0.011	0.011	0.010
	(0.012)	(0.012)	(0.012)	(0.012)
Backward Digit Span (Std)	0.028**	0.023**	0.023**	0.023*
	(0.012)	(0.011)	(0.011)	(0.012)
National Adult Reading (Std)	0.078***	0.076***	0.066***	0.066***
	(0.012)	(0.013)	(0.013)	(0.014)
Symbol Digits Modalities (Std)	0.195***	0.147***	0.136***	0.140***
	(0.011)	(0.015)	(0.015)	(0.015)
Age		-0.077***	-0.077***	-0.079***
		(0.013)	(0.013)	(0.013)
Age square		0.002***	0.002***	0.002***
		(0.000)	(0.000)	(0.000)
Age cube		-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)
Male		0.097***	0.097***	0.097***
		(0.022)	(0.022)	(0.023)
Years of education=9		0.282***	0.277***	0.280***
		(0.074)	(0.073)	(0.074)
Years of education=10		0.225***	0.220***	0.212***
		(0.059)	(0.059)	(0.060)
Years of education=11		0.250***	0.246***	0.240***
		(0.069)	(0.069)	(0.070)
Years of education=12		0.305***	0.297***	0.288***
		(0.057)	(0.057)	(0.057)
Years of education=15		0.323***	0.314***	0.313***
		(0.064)	(0.064)	(0.065)
Years of education=16		0.324***	0.319***	0.311***
		(0.071)	(0.071)	(0.072)
Years of education=17		0.294***	0.292***	0.291***
		(0.075)	(0.075)	(0.076)
Engl. speaking background (Base: Australia born)		-0.008	-0.012	-0.011
		(0.035)	(0.035)	(0.036)
Non-Engl. speaking background		-0.064*	-0.050	-0.038
		(0.037)	(0.037)	(0.038)
Other Urban (Base: Major city)		0.011	0.010	0.025
		(0.026)	(0.026)	(0.043)
Bounded Locality		0.040	0.048	0.038
		(0.066)	(0.066)	(0.074)
Rural Balance		0.053*	0.048	0.052
		(0.032)	(0.032)	(0.042)
Number of persons aged 0-4 years		0.064***	0.065***	0.064***
		(0.022)	(0.022)	(0.022)
Number of persons aged 5-9 years		0.039*	0.039*	0.032
		(0.022)	(0.022)	(0.023)
Number of persons aged 10-14 years		-0.001	-0.001	0.001
		(0.022)	(0.022)	(0.023)
Unemployed (Base: employed)		0.033	0.043	0.043
		(0.064)	(0.064)	(0.064)
Out of the labor force		0.084***	0.085***	0.087***
		(0.029)	(0.029)	(0.029)
Interviewer rating of underst. questions (Std)			0.067***	0.088***
			(0.011)	(0.013)
Observations	8666	8658	8658	8658
R^2	0.070	0.103	0.107	0.124

Note: Each column represents a different estimation model with survey item-response rate (SIRR), a standardized measure of survey response behavior (mean 0, SD 1), as the dependent variable. Column (1) includes all cognitive and non-cognitive ability measures simultaneously (all standardized to mean 0, SD 1). Column (2) adds all individual characteristics. Column (3) adds interviewer rating of the respondent's understanding of the questions, which is based on five possible categories: very poor, poor, fair, good excellent (standardized to mean 0 and SD 1). Column (4) adds interviewer fixed effects (140 dummy variables). Standard errors are reported in parentheses. *** $p<0.10$ ** $p<0.05$, *** $p<0.01$.

C Proportion of non-response by question

Figure C.1 summarizes information on the proportion of non-respondents for each individual question on the vertical axis, and the difference in SDM between respondents and non-respondents on the horizontal axis. Large values on the vertical axis indicate a greater number of non-respondents. A positive value on the horizontal axis indicates that respondents score higher on SDM, while a negative value indicates that they score lower. Light-grey diamond symbols indicate that the difference in SDM for the specific question is not statistically significant at the 5% level or better.

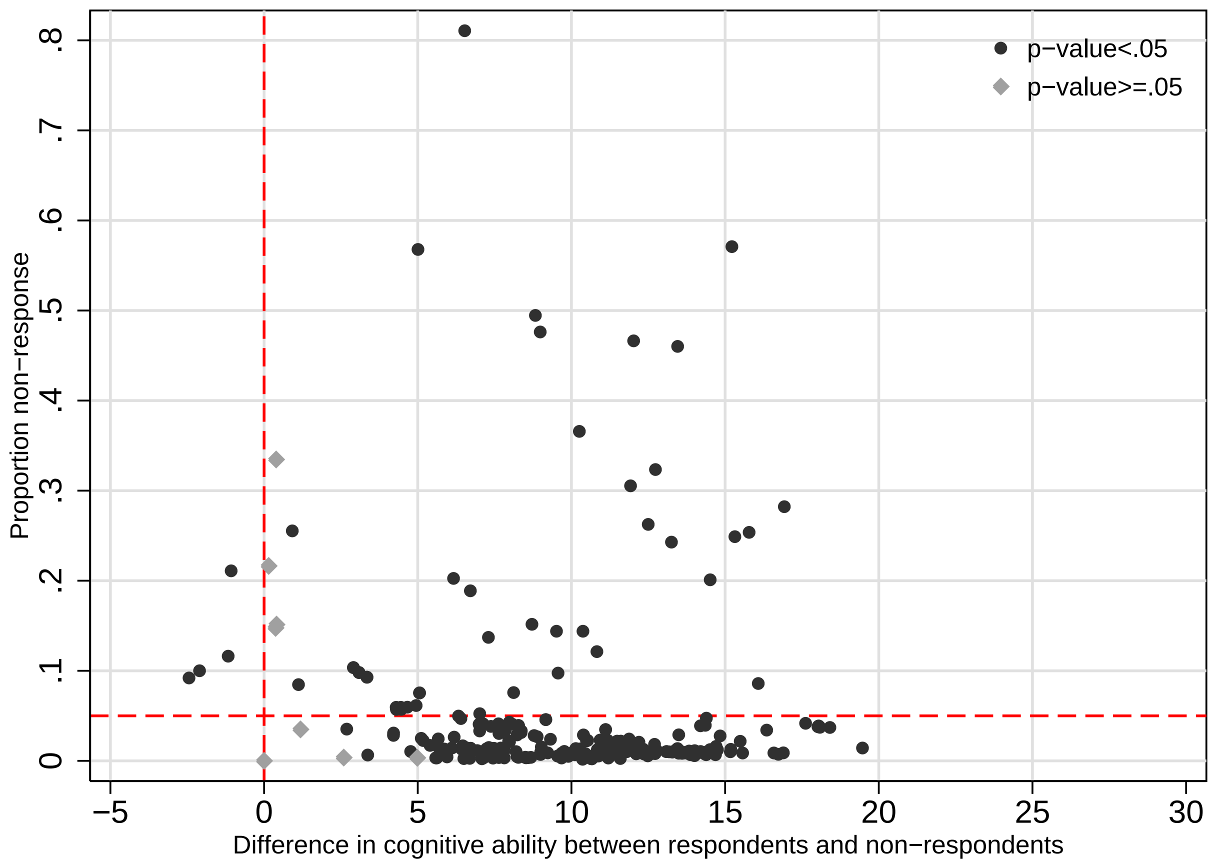


Figure C.1: Summary of proportion of non-response for each item on the self-completion survey (vertical axis) and difference in cognitive skills (Symbol Digits Modalities Test) between respondents and non-respondents for each item (horizontal axis).

D Omitted-variable bias (OVB) and Imperfect Proxy Bias (IPB)

The OVB in the estimated returns of education ($\hat{\beta}_1$) will be:[41]

$$E(\hat{\beta}_1) - \alpha_1 = \alpha_2 \delta_{MX|Z} = \alpha_2 \frac{S_{M|Z}}{S_{X|Z}} [r_{XM|Z}], \tag{1}$$

where $E()$ is the expectation operator, subscript Z denotes all other control variables (Z_i), $\delta_{MX|Z}$ is the regression coefficient for X_i in a regression of M_i on X_i and Z_i; $S_{M|Z}$ is the standard deviation of the errors of the regressions of M_i on Z_i; $S_{X|Z}$ is the standard deviation of the error in the regression X_i on Z_i. Finally, $r_{XM|Z}$ is the partial correlation coefficient for X_i and M_i, given Z_i.

The IPB in the estimated returns to education ($\hat{\gamma}_1$) will be:

$$\begin{aligned} E(\hat{\gamma}_1) - \alpha_1 &= \alpha_2 \delta_{MX|ZP} = \alpha_2 \frac{S_{M|ZP}}{S_{X|ZP}} [r_{XM|ZP}], \\ &= \alpha_2 \frac{S_{M|Z}}{S_{X|Z}} \left[\frac{r_{XM|Z} - r_{MP|Z} r_{XP|Z}}{1 - r^2_{XP|Z}} \right], \end{aligned} \tag{2}$$

where $\delta_{MX|ZP}$ is the regression coefficient for X_i in a regression of M_i on X_i, Z_i, and P_i; $S_{M|ZP}$ is the standard deviation of the errors of the regressions of M on Z_i and P_i; $S_{X|ZP}$ is the standard deviation of the error in the regression X_i on Z_i and P_i; $r_{XM|ZP}$ is the partial correlation coefficient for X_i and M_i, given Z_i and P_i. The partial correlation coefficient $r_{MP|Z}$ measures the association between M_i and P_i, given Z_i, and thus uncovers the strength of the proxy variable. Finally, $r_{XP|Z}$ is the partial correlation coefficient for X_i and P_i, given Z_i, which indicates the potential for a multicollinearity problem.

[41] See the specification bias analysis that was originally developed in Theil (1957).

E Proof of theorems

Let $A = r_{MP|Z}$, $B = r_{XM|Z}$, and $C = r_{XP|Z}$. A, B, and C are bound between -1 and $+1$. To achieve a bias reduction, λ has to be smaller than 1:

$$\lambda = \frac{(r_{XM|Z} - r_{MP|Z} r_{XP|Z})^2}{r^2_{XM|Z}(1 - r^2_{XP|Z})^2} = \frac{(B - A \times C)^2}{B^2(1 - C^2)^2} < 1. \tag{1}$$

Because λ will depend on the sign of A and B, we transform the primitive to be a function of C and $K = \frac{A}{B}$ alone. To do so, we divide both numerator and denominator by $\frac{1}{B^2}$ to get:

$$\lambda = \frac{(1 - K \times C)^2}{(1 - C^2)^2} < 1. \tag{2}$$

Because $(1 - C^2)^2 > 0$, we multiply Eq. (2) by $(1 - C^2)^2$:

$$(1 - K \times C)^2 < (1 - C^2)^2. \tag{3}$$

The term $(1 - X \times C)$ can be positive or negative. To be able to square root both sides of Eq. (3), we need to distinguish two cases:

Case 1: Let $(1 - K \times C) > 0$ and take square roots on both sides:

$$(1 - K \times C) < 1 - C^2, \tag{4}$$
$$C^2 < K \times C. \tag{5}$$

If $C > 0$, then $C < K$. Because $C > 0$ it must be true that $K > 0$. Since $K = \frac{A}{B}$, it must also be true that $sign(A) = sign(B)$. In contrast, if $C < 0$, then $C > X$. Because $C < 0$ it must be true that $K < 0$. Since $K = \frac{A}{B}$, it must also be true that $sign(A) \neq sign(B)$.

Case 2: Let $(1 - K \times C) < 0$ and take square roots on both sides:

$$-(1 - K \times C) < 1 - C^2, \tag{6}$$
$$K \times C < 2 - C^2. \tag{7}$$

If $C > 0$, then $K < \frac{2-C^2}{C}$. Since $(1 - K \times C) < 0$, it follows that $\frac{1}{C} < K$ and therefore $K > 0$. Therefore, $sign(A) = sign(B)$. If $C < 0$, then $K > \frac{2-C^2}{C}$. Since $(1 - K \times C) < 0$, it follows that $\frac{1}{C} > K$ and $K < 0$. Therefore, $sign(A) \neq sign(B)$. Thus, we have proven the necessary and sufficient condition of a valid proxy variable:

Theorem 1: A necessary condition for the proxy variable to be valid is that $sign(A) = sign(B)$ if $C > 0$, and $sign(A) \neq sign(B)$ if $C < 0$.

Theorem 2: A sufficient condition for the proxy variable to be valid is that $C < \frac{A}{B} < \frac{2-C^2}{C}$ if $C > 0$, and $C > \frac{A}{B} > \frac{2-C^2}{C}$ if $C < 0$.

F Returns to education and non-cognitive ability with and without controlling for cognitive ability

Table F.1: Wage regression model

DV: Log hourly wages	True model	CA omitted	SIRR included	SIRR & CA included
Years of education	0.0721***	0.0809***	0.0801***	0.0718***
	(0.0037)	(0.0036)	(0.0036)	(0.0037)
Locus of control (Std)	0.0408***	0.0455***	0.0450***	0.0407***
	(0.0090)	(0.0091)	(0.0091)	(0.0090)
SIRR (Std)			0.0262**	0.0155
			(0.0114)	(0.0115)
Observations	5780	5780	5780	5780

Note: Model estimated with Tobit sample selection model, with left censored observation set to missing. Exclusion restriction for selection equation are number of children in age bracket 0-4, 5-10, 11-14. Estimation sample is based on working age population, ages 24 to 64. All regression models control for gender, age, years of education, a dummy for casual worker, years of tenure with current firm and its square, state of residence dummies, region dummies (other urban, bounded locality, rural balance), language background and Big 5 personality traits. Standard errors are reported in parentheses. *** $p<0.10$ ** $p<0.05$, *** $p<0.01$.

G Simulation of λ

To show the dynamics in λ, we reformulate algebraically λ so that it is expressed in terms of the relative strength of the proxy $\frac{A}{B} = \frac{r_{CP|Q}}{r_{EC|Q}}$ and the potential multicollinearity problem ($C = r_{EP|Q}$).

$$\lambda = \frac{(r_{EC|Q} - r_{CP|Q} r_{EP|Q})^2}{r^2_{EC|Q}(1 - r^2_{EP|Q})^2} = \frac{(B - A \times C)^2}{B^2(1 - C^2)^2}. \tag{8}$$

We first expand the parentheses in both numerator and denominator:

$$\lambda = \frac{B^2 - 2 \times A \times B \times C + A^2 \times C^2}{B^2 - 2 \times B^2 \times C^2 + B^2 \times C^4}. \tag{9}$$

We then multiply both numerator and denominator with $\frac{\frac{1}{B^2}}{\frac{1}{B^2}}$ and by defining $\frac{A}{B} = X$; simplifying terms we obtain:

$$\lambda = \frac{(1 - X \times C)^2}{(1 - C^2)^2}. \tag{10}$$

Expressing λ in terms of only two unknowns allows us to simulate λ over two unknowns. We simulate the dynamics in λ for five possible values C (0.05, 0.20, 0.40, 0.60, 0.80) and continuous values of X, which we allow to range between -3 and +3 in steps d of 0.1. Small values on X imply that the relative strength of the proxy variable is low, while large values on X imply that the relative strength is very high.