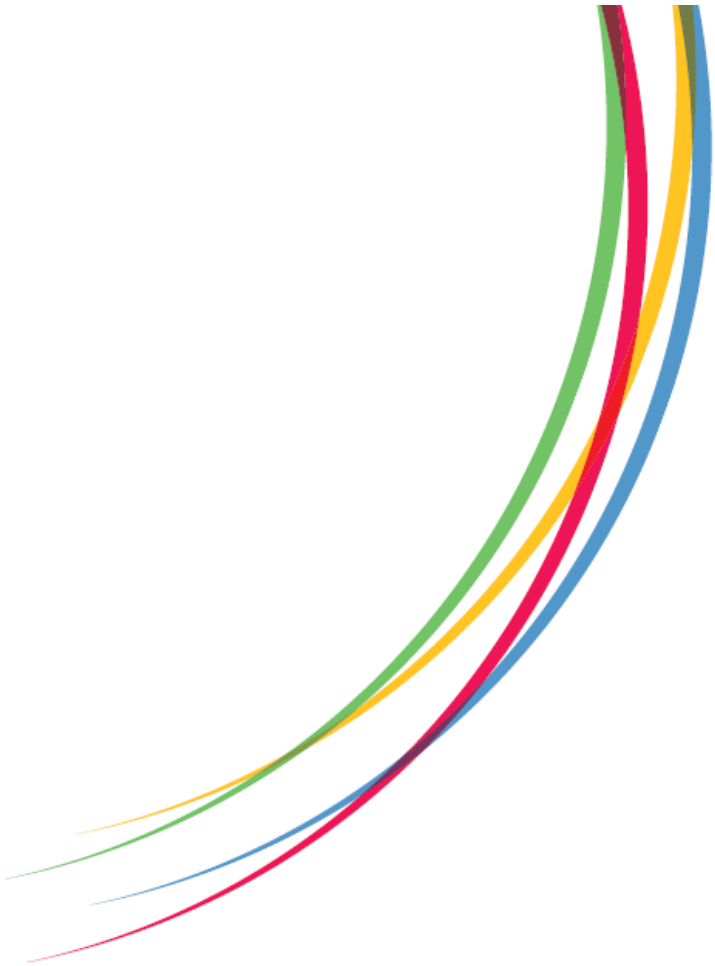


Understanding Society

Working Paper Series

No. 2017 – 08

August 2017

Data Quality in the Understanding Society Youth Self-Completion Questionnaire

Aigul Mavletova

Laboratory for Comparative Social Research, National Research University Higher School of Economics, Department of Sociology, Moscow, Russia

Peter Lynn

Institute for Social and Economic Research, University of Essex

Data Quality in the Understanding Society Youth Self-Completion Questionnaire

Aigul Mavletova & Peter Lynn

Non-Technical Summary

At each annual wave of data collection for Understanding Society, all sample members aged 10 to 15, plus other people in that age range in the households of sample members are asked to complete a paper self-completion questionnaire. These questionnaires are either collected by field interviewers or mailed back by the sample households. Around three-quarters of young people complete the questionnaire each year.

The data collected by means of these questionnaires is often used by analysts in conjunction with the data collected in the main Understanding Society interviews from the parents of the young people, or other household members such as adult siblings. The data can also be used to track the attitudes and achievements of sample members over time, by linking to the main interview data once the young person turns 16. However, little has been documented about the quality of the data collected via these paper questionnaires. This paper aims to fill that gap by reporting on some aspects of data quality.

We consider six measures of data quality: 1) the extent to which young people omit to answer questions; 2) the extent to which they answer “don’t know”; 3) the extent to which they provide rounded answers to questions requiring a numeric answer; 4) the extent to which they provide the same answer to each question in a list of related questions with the same response options (“non-differentiation”); 5) the extent to which sensitive behaviours are reported; and 6) the extent to which answers are consistent (plausibly related) over time.

Data Quality in the Understanding Society Youth Self-Completion Questionnaire

Aigul Mavletova & Peter Lynn

Abstract

This article addresses some questions about the determinants and nature of response quality in self-completion surveys of young people. Every young person aged 10-15 in Understanding Society is asked to complete a self-administered paper questionnaire. Data from waves 1 to 4 (2009-2013) are analysed. We find lower data quality amongst children who live in households with lower socio-economic status. We find some support for both increasing social desirability bias and the panel conditioning effect. We find no evidence to support a hypothesis that increased difficulty in contacting the household would be associated with reduced data quality in the youth survey.

Key words: item non-response; longitudinal surveys; non-differentiation; paper self-completion questionnaires; rounding

JEL classifications: C81, C83

Author contact details: amavletova@hse.ru

Acknowledgements: This paper was conceived and initiated during a research visit to ISER by the first author. The contribution of the first author is funded by the Russian Academic Excellence Project '5-100'. The contribution of the second author is funded by the UK Economic and Social Research Council award for *Understanding Society*: the UK Household Longitudinal Study.

Data Quality in the Understanding Society Youth Self-Completion Questionnaire

Aigul Mavletova & Peter Lynn

1. Introduction

There is a growing interest in children as informants in both qualitative and quantitative research. Punch (2002) suggests that there are two opposite methodological views on research among children. There are those who believe that children can be treated as competent respondents and therefore that the same survey methods can be employed as among adults. The second approach considers children as different from adults and suggests that methods should be different from those that are appropriate for adult respondents. While the latter approach uses mostly observational data and projective techniques (such as drawings and writing), the first approach uses quantitative research methods. Based on the children's intellectual development scheme proposed by Piaget, researchers suggest that starting from age 10 or 11 children can answer standardized questionnaires, though these questionnaires should be adapted to the cognitive abilities of particular age groups, which clearly increase with age (Borgers, De Leeuw, and Hox, 2000; Smith, 2008). Since there is a growing interest in childrens' well-being, victimization, bullying, and risky behaviours, children are included as respondents in a number of repeated cross-sectional studies (e.g. the British Crime Survey or the National Crime Victimization Survey) and cohort panels (e.g. the Longitudinal Study of Young People in England, the National Child Development Study, or the National Longitudinal Survey of Youth). In 1994, the British Household Panel Survey (BHPS) was the first European household panel to ask youth household members aged 11-15 to complete the survey. In 2010-2011 the BHPS participants joined a larger panel, Understanding Society: the UK Household Longitudinal Study (UKHLS). The UKHLS continued to collect data annually from youth in the sample while extending the age range to 10-15. Every household member within this age range is asked to complete a self-administered paper questionnaire.

Though there are several papers about data quality among children and adolescents in both cross-sectional surveys (see Amato and Ochiltree, 1987; Beebe et al, 1998; Fuchs,

2005; König, 2011; Mavletova, 2015a; Vogl 2012, 2013) and longitudinal surveys (Borgers, Hox, and Sikkel, 2003, 2004; Fendrich and Kim, 2001; Fendrich and Rosenbaum, 2003; Halpern-Manners, Warren, and Torche, 2014; Haunberger, 2014; Mavletova, 2015b; Smith and Platt, 2013; Vaillancourt, 1977), to our knowledge none of the studies measured data quality in household panel studies among youth. Using data from a survey with the household panel survey design, but including a youth self-completion component, allows us to use data from the interviews with parents as covariates in the analysis of data quality in the youth survey data. Furthermore, the large sample size of UKHLS allows us to identify age effects in terms of differences between single year cohorts, while relatively high wave-on-wave response rates provides good sample sizes for studying panel conditioning effects. We develop a number of hypotheses regarding the correlates and nature of data quality in the youth survey data and test them using multiple measures of quality. We find that data quality is lower among children who live in the households with lower socio-economic status, lower parental involvement in education, lower expectations, and a less close parent–child relationship, where quality is measured by item nonresponse rates, rate of "don't know" responses, tendency to rounding, and extent of inconsistent responses. By studying residual variances from structural equation models we find a panel conditioning effect in an item about how happy the respondent is with school. Using mixed effects logit models, we find no evidence of a change over waves in social desirability bias. Finally, we find no evidence of a relationship between household reluctance to participate in the survey and data quality among youth.

2. Literature Review and Hypotheses

In this section we briefly review the results of past methodological research on the quality of survey data provided by children and adolescents. Data quality depends highly on the age and cognitive skills of the children. There are several different indicators of data quality. We review some, which we will use in this paper, in particular, item nonresponse rates, don't know responses, non-differentiation, data consistency, rounding, data reliability, and panel conditioning. Since there is more literature on panel conditioning among adults, we will review panel conditioning effect among both adults and youth.

Item nonresponse rates

Borgers, De Leeuw, and Hox (2000) found that boys and younger children produce a higher item nonresponse rate (INR). Based on analysis of the Child Longitudinal Study

among respondents 8-11 Haunberger (2014) also found that younger respondents produce higher INR in a face-to-face interview. However, Mavletova (2015a) found no evidence of age or gender effects on INR in a web survey among respondents 7-15. Those with higher academic achievements produce lower INR (Haunberger, 2014; Mavletova, 2015a).

Borgers and Hox (2001) found a higher probability of item nonresponse at the end rather than at the beginning of the questionnaire. Also, a higher number of response options induced higher INR among children, while knowledge questions and the length of the introductory text in the question reduced INR.

Don't know

Since expressing of "don't know" responses may be socially undesirable for young respondents they tend to give an answer rather than state they "don't know" (König, 2010; Scott, 2008). Waterman, Blades, and Spencer (2001) found that the tendency to provide an answer depends on the question format. In an experiment, respondents were offered some open-ended and some closed questions which respondents could not answer based on the information researchers provided. Most of the children answered the closed questions but stated they don't know in response to the open-ended questions. There is some research evidence that those with a lower academic performance and those who complete the survey in the presence of third parties produce a higher rate of "don't know" responses in web surveys (Mavletova, 2015a). Vogl (2012) found that older respondents (11-year-olds) produced slightly lower number of "don't know" responses compared to the 5, 7- and 9-year olds in a face-to-face and telephone semi-structured interviews, but the difference was not statistically significant. She also found that most of the "don't know" responses are due to a lack of knowledge or failure of memory and only a few are due to comprehension difficulties or avoiding the answer in sensitive questions.

Rounding and straightlining (non-differentiation)

Smith and Platt (2013) compared closed and open-ended responses in a question on frequency of alcohol consumption among youth 10-15. They found that an open-ended format leads to rounding (5, 10, 15, etc.) and produced a higher item non-response rate.

Mavletova (2015a) found that those with lower cognitive abilities and who completed the survey in the presence of third parties tended to straightline more in the grid questions (select the same response category for each item in the scale).

Reliability

Borgers, De Leeuw, and Hox (2000) found lower internal consistency of the scales among children 9-10 compared to children and adolescents 11-15. Vaillancourt (1977) conducted a three-wave panel survey of 1,000 children 9-15. She found lower test-retest reliability among younger respondents and in the questions on political attitudes. While test-retest reliability in the questions on political attitudes varied from 0.25 to 0.62, test-retest reliability in factual questions varied from 0.31 to 0.67. Mavletova (2015b) conducted a two-wave experiment comparing a text-only, visual, and gamified web survey among 737 respondents 7-15 y.o. She found the lowest test-retest correlations in the negatively formulated questions (0.23-0.29) and the highest in the knowledge questions on maternal and paternal level of education. She also found lower test-retest reliability in the gamified than in the text-only and visual surveys. Borgers, Hox, and Sikkel (2003, 2004) conducted a two-wave web survey among 91 respondents 8-16. They found higher reliability in fully-labeled questions, lower reliability in questions with a midpoint and no effect of negatively formulated questions on test-retest reliability.

Panel conditioning

Several studies showed a positive panel conditioning among adults, when respondents have lower item non-response rates, less severe rounding, and higher reliability in subsequent waves (Rendtel et al., 2004; Schräpler, 2003; Sturgis, Allum, and Brunton-Smith, 2009). These changes appear to be due to greater understanding of the survey process and response tasks, when later waves have higher data quality on some indicators.

Other studies showed panel conditioning effect on a number of substantive measures of respondents' behaviour. Answering survey questions makes participants think about question matters which may change both behaviour and their responses in subsequent waves. There is some evidence of changes in responses or behaviour while answering survey questions on illicit behaviours. Fendrich and Kim (2001) found lower level of reporting of lifetime marijuana and cocaine use in 1988, 1992 and 1994 compared to 1984 in the National Longitudinal Survey of Youth. About 40% of participants denied lifetime cocaine use in one of the following waves. Black and married respondents had higher probability of denying. Fendrich and Rosenbaum (2003) as well as Torche, Warren, and Halpern-Manners (2012) found similar results among youth in different panels and

countries when respondents tend to decrease level of reporting of deviant behaviour in the subsequent waves.

Hypotheses

Below we outline our main research hypotheses taking into consideration survey data we have. These concern each of the data quality measures discussed above. In addition to hypothesising that these phenomena should be present in surveys of children, since we have data from the parental questionnaires, we examine how the phenomena are associated with relevant covariates relating to such variables as household income, parental education, parental attitudes towards education, parental involvement in their children's education, and parent-child relationships. These covariates are particularly relevant as previous studies have shown them to be associated with the academic achievements of children, which in turn are expected to be associated with the quality of survey data provided. The influence on children's academic achievements has been shown for parental education (Haveman and Wolfe, 1995), household income (Dahl and Lochner, 2012; White, 1982), parental attitudes towards education (Davis-Kean, 2005; Seginer, 1983), parental involvement in education (such as attending school meetings and discussing things that child studies); and the quality of parent–child relationship (such as spending time together and warmth of the relationship) (Davis-Kean, 2005). Mother's education has a stronger effect on offspring's education than father's education. To examine the associations with children's data quality we take advantage of the rich structure of our data set, which includes interviews with parents and household-level information. Most covariates are taken from either the household interview or the mother's interview.

Hypothesis 1. Data quality is lower among children who live in households with lower socio-economic status, lower parental involvement in the education, lower expectations, and a less close parent–child relationship.

In line with studies of panel conditioning in adults, discussed above, we expect a negative panel conditioning effect, i.e. increasing social desirability bias (a lower level of reporting of sensitive behaviours) and increasing measurement errors across waves).

Hypothesis 2. There is a negative panel conditioning effect, namely increasing social desirability bias and increasing measurement error.

Some research results showed that reluctant or late respondents tend to produce higher measurement error compared to early respondents (Triplett, Blair and Kang, 1996), while some others papers found no effect (Hox, de Leeuw, and Chang, 2012) or showed that the variation might be explained by the differences in cognitive abilities of late and early respondents (Kaminska, Mccutcheon, and Billiet, 2010). We suggest that similar to the findings regarding adults, hard-to-reach households would produce lower data quality among children.

Hypothesis 3. The more difficult it was to contact the household, the lower data quality is among children.

3. Data

Understanding Society is a multi-purpose longitudinal study based on a sample of around 100,000 individuals representing the UK population. It provides a major data resource for research in the social sciences (Buck & McFall 2012). This includes a representative sample of the UK population in 2009, ethnic minority boost, and the BHPS sample. The BHPS sample was included from the second wave of the UKHLS. At each annual wave, data are collected from each current member of the household of each sample member. Our analysis uses data from four waves of the study: the first wave was conducted in 2009-2010, the second in 2010-2011, the third in 2011-2012, and the fourth in 2012-2013. Every child within the age range 10-15 is asked to complete a paper self-completion questionnaire. The youth questionnaire self-completion rate was 74-75% in participating households in all four waves (see Table 1). There were 4,899 completed interviews in the first wave; 5,020 in the second wave; 4,427 in the third wave; and 4,049 in the fourth wave. In each wave about 50% of youth respondents were boys and the average age of the respondents was about 12.5 years old in all waves. (see Table 1, unweighted statistics).

Questionnaires

The total number of items varied from 88 to 104 in different waves (see Table 2). Overall the questions focus on health behaviour, school, friends, psychological well-being, bullying at school, doing sports, and relationships with family. Some questions were included in each wave, while some were asked biennially. In the second and fourth waves there were a higher number of open-ended items, items with a “don’t know” response option, and

sensitive items about smoking, alcohol, and drugs (Table 2). Most of the open-ended questions required a number as an input (e.g. current weight, number of books, number of hours on homework). Copies of all questionnaires are available at www.understandingsociety.ac.uk/documentation.

Table 1: Descriptive Statistics

	Wave 1 (2009-2010)	Wave 2 (2010-2011)	Wave 3 (2011-2012)	Wave 4 (2012-2013)
Youth questionnaire self-completion rate (in productive households)	74%	74%	75%	75%
N (number of completed interviews)	4,899	5,020	4,427	4,049
Mean age	12.5 (SD=1.71)	12.5 (SD=1.67)	12.5 (SD=1.70)	12.6 (SD=1.69)
Males	49.8%	50.5%	50.2%	50.5%
Socio-economic status and household variables				
Gross household income (month before interview, UK pounds)	3,710.9 (SD=2,652.11)	3,926.9 (SD=2,676.50)	4,202.6 (SD=2,665.60)	4,266.2 (SD=2,711.97)
Own home	62.3%	65.6%	67.0%	68.4%
Mother has diploma in higher education	28.2%	29.8%	30.9%	32.6%
Parents are married and live together	63.8%	64.4%	67.9%	67.3%
Mother's employment status: employed/self-employed	64.4%	65.3%	67.1%	68.8%
Mother's race: white	74.4%	81.3%	80.1%	79.5%
Urban area	80.8%	76.7%	76.6%	75.9%

Table 2: Basic Statistics on Youth Questionnaire across Waves

	Wave 1	Wave 2	Wave 3	Wave 4
Total number of items	88	89	104	101
Sensitive items on smoking, alcohol, and drugs	6	24	9	24
Open-ended items	3	10	6	10
Number of items with “don’t know” / “not sure” response category	4	17	4	19

4. Procedures and measures

We measured data quality based on the following indicators (see the Appendix for more details):

Hypothesis 1:

- The overall item nonresponse rates (INRs)
- Selecting the “don’t know” response category
- Rounding
- Inconsistency

Hypothesis 2:

- Panel conditioning effect
- Social desirability bias

Hypothesis 3:

- Reluctance to participate in the survey

5. Results

Item nonresponse rates, don’t know responses, and rounding

The mean INRs were 2.1% in the first wave (SD=4.9), 4.7% in the second wave (SD=6.5), 3.4% in the third wave (SD=4.4), and 5.4% in the fourth wave (SD=8.6). The second and fourth waves had higher item nonresponse rates. The highest INRs were in the cognitively demanding open-ended questions in the third wave: total amount of received money to

spend on oneself last week (82.4%) and earned money if respondent had a paid job (52.7%). However, most questions with high INR were in the second and fourth waves. For questions about the respondent's height and weight the INR varied from 37% to 42%. Other questions with high INR were sensitive items in which respondents were expected to evaluate the ease of obtaining cannabis (INR 33-35%), perceived risk of trying amphetamine (INR varied from 21% to 26%) and ecstasy (INR varied from 19% to 22%). Respondents might not have enough knowledge to evaluate the risk of taking some of these drugs.

We ran a multivariate linear regression to predict the overall INR in each wave. Some effects in the more cognitively demanding second and fourth waves were different from the first and third waves. The age effect was about five times larger: each additional year reduced the INR on average by 0.15-0.17 percentage points in the first and third waves and by 0.69-0.82 percentage points in the second and fourth waves (see Table 4). Maternal education and mother's employment had a significant effect in the second and fourth waves. Children with higher educated (β=-0.56-0.77) and employed mothers (β=-0.50-0.60) produced lower INRs. There was no gender effect in the more cognitively demanding waves, while boys had a higher INR in the first and third waves (β=0.29-035).

Household income had no consistent effect, while living in a house owned by the household produced a lower INR. Stronger parent-child relationships decreased INR: the more often parents discuss books at home or take children to museums and theatres, the lower INR children produced. The difference was up to 1.8 percentage points. Mother's race had also a significant effect on INR. Caucasian children had lower INR by 0.76-1.35 percentage points. The effect was twice higher in the more cognitively demanding second and fourth waves. Mothers' expectations to complete A level / exams had no significant effect on INR.

Table 4: Predicting Item Nonresponse Rates

	Wave 1	Wave 3	Wave 2	Wave 4
Intercept	5.27 (0.56)***	5.46 (0.56)***	16.18 (0.88)***	18.04 (1.03)***
Respondent's characteristics				
Males	0.35 (0.11)**	0.29 (0.11)**	-0.16 (0.16)	0.19 (0.18)
Age	-0.17 (0.04)***	-0.15 (0.03)**	-0.69 (0.05)***	-0.82 (0.06)***
Socio-economic status				
Gross household income (month before interview)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)*
Own home	-0.27 (0.14)*	-0.32 (0.14)*	-0.40 (0.20)*	-0.05 (0.23)
Mother has diploma in higher education	-0.04 (0.13)	-0.14 (0.13)	-0.77 (0.18)***	-0.56 (0.21)**
Mother expectations (waves 1 and 3)				
Importance for your child to complete A level exams: very important	-0.07 (0.13)	-0.11 (0.13)		
Parental involvement in education				
My parents are interested in how I do at school: always or nearly always	-0.20 (0.16)	-0.18 (0. 15)		
My parents come to school parent evenings: always or nearly always	-0.20 (0.16)	-0.01 (0.17)		
Mother helps with homework: once a week or more often	-0.45 (0.13)***	-0.06 (0.13)		
Someone at home helps with homework			-0.15 (0.26)	-0.17 (0.28)
Parent–child relationship				
Spending time together on leisure activities with mother: several times a week or almost every day	0.18 (0.14)	0.18 (0.13)		
Quarrel with child: less than once a week	0.01 (0.11)	-0.03 (0.11)		
The child talks with mother about things that matter to him/her: most days	0.25 (0.12)*	-0.07 (0.12)		
Discussing books at home, discussing TV programmes, buying books as gifts etc.			-0.07 (0.03)**	-0.10 (0.03)***
Household socio-demographic variables				
Parents are married and live together	-0.28 (0.13)*	-0.08 (0.13)	-0.26 (0.19)	-0.08 (0.22)
Number of children under 15 in the household	0.10 (0.06)	0.08 (0.06)	-0.03 (0.06)	0.27 (0.09)**
Mother's employment status: employed/self-employed	-0.15 (0.14)	0.04 (0.13)	-0.50 (0.19)**	-0.60 (0.22)**
Mother's race: white	-0.76 (0.15)***	0.02 (0.15)	-1.35 (0.21)***	-1.24 (0.23)***
Urban area	-0.07 (0.15)	-0.22 (0.14)	-0.05 (0.21)	-0.29 (0.24)
N	4,193	3,030	3,150	2,640
R square	0.032***	0.018***	0.095***	0.131***

About 9% of respondents selected the “don’t know” response category in the second wave. We ran a multivariate logistic regression to predict the selection of the response. Boys (OR[odds ratios]=1.67, p<0.001, results not shown) and younger respondents (OR=0.81, p<0.001) had higher odds ratios of selecting the “don’t know” category. Maternal higher education (OR=0.49, p<0.001), home ownership (OR=0.55, p<0.001) and stronger parent-child relationships (OR=0.93, p<0.001) decreased the odds ratios of selecting the “don’t know” response. Caucasian children had higher odds ratios of stating they don’t know the answer (OR=1.64, p<0.01).

The proportion of those who used round numbers (5, 10, 15 or 20) in the question on the number of close friends varied from 25% to 29% in four waves: 28.5% in the first wave, 26.6% in the second wave, 29.2% in the third wave, and 25.1% in the fourth wave. Boys and older respondents tend to use round numbers more frequently than girls and younger respondents (see Table 5). A higher mother’s involvement in the education and closer relationship produced lower odds ratios of using round numbers. Contrary to expectations living in the household with married parents increased the odds of using round numbers.

About 9% of respondents denied smoking cigarettes and 30% denied drinking alcohol subsequent to a previous response indicating that they had smoked cigarettes or drunk alcohol respectively. Older respondents (OR=1.33-1.57, data not shown) and Caucasian children (1.68-1.81) had higher odds of providing inconsistent responses across waves (results not shown). Greater parental involvement in education when they come to parent evenings (OR=0.63-0.78), are interested in the child’s academic achievements (OR=0.73) or help to do homework (OR=0.67) as well as closer parent-child relationship (OR=0.95) decrease the odds of producing inconsistent responses. Those who live in a home owned by the household and with both parents also have lower odds of providing inconsistent responses.

Table 5: Predicting Rounding (Odds Ratios)

	Wave 1	Wave 3	Wave 2	Wave 4
Intercept	0.07*** (0.04-0.15)	0.04*** (0.02-0.09)	0.09*** (0.03-0.21)	0.09*** (0.03-0.22)
Respondent's characteristics				
Male	1.28*** (1.10-1.47)	1.29** (1.11-1.49)	1.34** (1.14-1.58)	1.28** (1.09-1.50)
Age	1.11*** (1.06-1.16)	1.19*** (1.14-1.25)	1.11** (1.05-1.17)	1.06* (1.01-1.12)
Socio-economic status				
Gross household income (month before interview)	1.00	1.00	1.00	1.00
Own home	1.01 (0.84-1.21)	0.98 (0.81-1.18)	1.08 (0.88-1.32)	1.32** (1.07-1.62)
Mother has diploma in higher education	0.93 (0.78-1.10)	0.93 (0.75-1.15)	1.10 (0.91-1.32)	0.94 (0.73-1.22)
Mother expectations				
Importance for your child to complete A level exams: very important	1.06 (0.90-1.26)	1.00 (0.85-1.18)		
Parental involvement in education				
My parents are interested in how I do at school: always or nearly always	1.14 (0.93-1.41)	1.00 (0.82-1.22)		
My parents come to school parent evenings: always or nearly always	1.07 (0.86-1.32)	1.15 (0.92-1.44)		
Mother helps with homework: once a week or more often	0.82* (0.69-0.97)	0.94 (0.79 -1.11)		
Someone at home helps with homework			0.88 (0.67-1.14)	1.22 (0.95-1.57)
Parent–child relationship				
Spending time together on leisure activities with mother: several times a week or almost every day	1.19 (0.99-1.42)	1.08 (0.90-1.29)		
Quarrel with child: less than once a week	0.94 (0.81-1.09)	0.78** (0.67-0.91)		
The child talks with mother about things that matter to him/her: most days	0.88 (0.815-1.09)	0.85* (0.73-0.99)		
Discussing books at home, discussing TV programmes, buying books as gifts etc.			1.01 (0.98-1.04)	1.00 (0.98-1.03)
Household socio-demographic variables				
Parents are married and live together	1.23* (1.04-1.46)	1.34** (1.12-1.61)	1.09 (0.90-1.32)	0.85 (0.70-1.02)
Number of children under 15 in the household	1.03 (0.96-1.10)	0.95 (0.88-1.02)	1.02 (0.96-1.09)	1.04 (0.96-1.13)
Mother (self-)employed	1.04 (0.87-1.24)	1.01 (0.84-1.21)	1.05 (0.86-1.28)	1.27* (1.04-1.55)
Mother's race: white	1.06 (0.87-1.28)	0.91 (0.74-1.12)	1.21 (0.98-1.50)	1.00 (0.81-1.25)
Urban area	1.00 (0.83-1.21)	0.89 (0.74-1.12)	0.75* (0.60-0.93)	0.91 (0.75-1.10)
N	3,709	3,578	2,942	3,220
Cox & Snell R square; Nagelkerke R Square; -2 Log likelihood	0.017; 0.024; 4,365	0.031; 0.044; 4,180	0.013; 0.019; 3,372	0.013; 0.019; 3,654

Panel conditioning

We measured panel conditioning with respect to two items: how happy children are with school work and how happy they are with school overall. The standardized factor loadings for the variable of school work varied from 0.735 to 0.831, while for being happy with school the loadings were lower and varied from 0.513 to 0.577. We expected an increase in the correlations between residual variances from wave to wave as an indicator of a panel conditioning effect. We found no pattern in the correlations between residual variances in the item on school work, but the expected pattern for the item on being happy with school. There was an increase in the correlations between residual variances from wave to wave: the correlation between the first and second waves was 0.218, between the second and third waves 0.334, and between the third and fourth waves 0.440. This suggests an increasing panel conditioning effect.

Table 6: Correlations between residual variances

School work	Wave 1	Wave 2	Wave 3	Wave 4
Wave 1		0.311	0.018	-
Wave 2			-0.030	0.092
Wave 3				-0.039
School	Wave 1	Wave 2	Wave 3	Wave 4
Wave 1		0.218	0.064	-
Wave 2			0.334	0.331
Wave 3				0.440

Social desirability

Each sensitive item had a higher level of reporting amongst the balanced panel in each subsequent wave except for the level of lifetime drinking: while 50.3% of the participants reported lifetime drinking in the third wave, 44.8% reported it in the fourth wave (see Table 7). These differences could of course be caused by ageing (each respondent is three years older at wave 4 than they were at wave 1), so to identify any panel conditioning effect we ran mixed effects logit models to predict the level of reporting of sensitive behaviour, controlling for a number of socio-demographic variables including age. We found that panellists were less willing to report lifetime drinking in the third wave compared to the fourth wave (OR=2.21, $p<0.001$). It is consistent with our previous finding that 30% denied drinking alcohol in the subsequent waves. No social desirability bias was found in

other variables. Overall, we found some evidence of increasing social desirability bias in the panel.

Table 7: Level of Reporting of Sensitive Behaviours

	Wave 1	Wave 2	Wave 3	Wave 4
Have ever drunk alcohol	10.4%	19.4%	50.7%	44.8%
N	*948*	*945*	*954*	*948*
Have ever smoked	1.5%	3.6%	8.1%	9.9%
N	*953*	*946*	*949*	*946*
Played truant in the last 12 months	4.9%	5.4%	6.9%	10.2%
N	*952*	*943*	*944*	*941*

Reluctance

We conducted a number of regression models to measure whether reluctant households produce higher INR or higher tendency to provide rounded numeric answers. We found almost no evidence. Only in the second wave children from more reluctant households produced higher INR (β=0.07, SE=0.04, $p<0.05$). No other significant effects were found.

6. Discussion and Conclusion

We have examined several indicators of data quality in the paper self-completion questionnaires answered by young persons aged 10-15 in *Understanding Society*. The questionnaires in four study waves varied in total number of items, number of open-ended questions, and number of sensitive items. A number of questions had a high item nonresponse rate. More than a third of the respondents did not answer the questions about their height and weight. Up to a third did not provide an evaluation of the ease of obtaining some drugs or the perceived risk of trying them. Overall item nonresponse rates varied from 2.1% to 5.4% in different waves. The second and fourth waves had more cognitively demanding questions. As a result, they had higher item nonresponse rates. While previous studies found mixed evidence of the difference in item nonresponse rates between boys and girls, we found higher item nonresponse rates among boys in less difficult waves and no difference in the waves with more cognitively demanding questions. It seems that the difference may depend on the types of questions used in a survey. The

age effect was significant in each wave, but was five times larger in more cognitively demanding waves.

In accordance with findings of Haunberger (2014), König (2010) and Scott (2008) we found that children would either skip the question or give an answer rather than explicitly state they don't know the answer. Though in some waves up to 20% of the questions had an explicit "don't know" response category we found the only question which produced "don't know" responses in the second wave ("At the moment, young people can leave school at 16. What would you most like to do when you are 16?"). This question was reformulated at subsequent waves.

We tested three main hypotheses. First, we expected that household variables such as socio-economic status, parental involvement in the education, education expectations, and parent–child relationship can be used as proxies for measuring data quality among youth. We measured item nonresponse rates, tendency to provide rounded numeric answers, don't know responses, and inconsistent responses. As expected, stronger parent-child relationship (when parents discuss books and TV programmes at home, and talk to them about important matters) and higher mother's involvement in education were associated with better data quality. Children from higher socio-economic households also produced higher data quality. Living in a home owned by the household had a positive effect on data quality. Mother's higher education increased data quality in some of the questions: it had a significant effect on "don't know" responses and item nonresponse rates in cognitively demanding waves. While older respondents produced lower item nonresponse rates, they gave more often inconsistent responses throughout waves and tended to use rounding more often. Overall, girls produced better data quality than boys.

Second, we expected a negative panel conditioning effect with an increase of social desirability bias and measurement errors in the subsequent waves. We found some evidence for a social desirability bias and for a panel conditioning effect with an increase of correlations between residual variances over time.

Third, we suggested that the more difficult it was to contact the household, the lower data quality would be in the youth questionnaires. Some previous studies had shown that reluctant or late respondents tend to produce higher measurement error compared to early respondents. We expected to find similar patterns among youth. However, we found no evidence in support of that hypothesis.

What are the implications of our findings? Overall, we found that some data quality indicators and the relation between the variables varied depending on the types of questions and number of cognitively demanding items in the questionnaire. Children from less educated and affluent households and younger respondents are mostly affected in more cognitively demanding waves. Younger age groups have difficulties answering questions about cannabis, ecstasy, and amphetamine. More than 40% of the respondents 10-12 did not give an answer to the question in which they were expected to evaluate how difficult they think it would be to get cannabis. Almost 30% of that age group did not give an answer to items evaluating the risk of taking cannabis, ecstasy or amphetamine. It seems that such questions are quite difficult for this age group and may need to include some age or knowledge filters.

We found that up to 30% of respondents denied lifetime drinking in one of the subsequent waves. It seems that the more sensitive a behaviour is, the higher the level of inconsistent responses. Only 9% of the respondents denied lifetime smoking in the subsequent waves. We found some evidence of an increase in social desirability bias in reporting on lifetime drinking in the fourth wave of the study. If we compare these statistics with other panel data Fendrich and Kim (2001) reported 40% of respondents denying lifetime cocaine use in one of the following waves in the National Longitudinal Survey of Youth.

We suggest that there might be some ways to improve data quality especially among younger respondents and those from households with lower socio-economic status and lower parental involvement in the child's education. Analysts should be aware that both measurement error and the tendency to provide a valid response could be associated with age, gender, socio-economic status, and the nature of parent-child relationships/involvement. However, the nature of this association may vary between items and between surveys, depending on overall survey content and context.

References

Alwin, D. F. (2007). The margins of error: A study of reliability in survey measurement. Wiley-Blackwell.

Amato, P.R, and Ochiltree, G. (1997). Interviewing children about their families: A note on data quality. *Journal of Marriage and the Family*, 49, 669-675.

Beebe, T. J., Harrison, P.A., Mcrae, Jr., J. A., Anderson, R. E., and Fulkerson, J.A. (1998). An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly,* 62, 623–632

Borgers, N., Hox J., and Sikkel, D. (2003). Response quality in survey research with children and adolescents: The effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research*, 15(1), 83-94.

Borgers, N., Hox, J., and Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity*, 38, 17–33.

Borgers, N., and Hox, J. (2001). Item nonresponse in questionnaire research with children. *Journal of Official Statistics*, 17(2), 321-335.

Borgers, N., De Leeuw, E., and Hox, J.P. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin of Sociological Methodology,* 66, 60–75.

Buck, N. and McFall, S. (2012). Understanding Society: design overview. *Longitudinal and Life Course Studies* 3(1), 5-17.

Cantor, D. (2008). A review and summary of studies on panel conditioning. In *Handbook of Longitudinal Research: Design, Measurement and Analysis,* Edited by S. Menard. Burlington, Academic Press pp.123-138.

Cernat, A. (2015). The Impact of mixing modes on reliability in longitudinal studies. *Sociological Methods & Research*, 44(3): 427-457

Clausen, A. (1968). Response validity: Vote report. *Public Opinion Quarterly*, 41: 56-64

Dahl, G. B., and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5), 1927-56.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294–304.

Fendrich, M., and Kim, J. Y.S. (2001). Multiwave analysis of retest artifact in the National Longitudinal Survey of Youth drug use. *Drug and Alcohol Dependence,* 62: 239–253.

Fendrich, M., and Rosenbaum, D. P. 2003. Recanting of substance use reports in a longitudinal prevention study. *Drug and Alcohol Dependence,* 70:241-53.

Fuchs, M. (2005). Children and adolescents as respondents. Experiments on question order, response order, scale effects and the effect of numeric values associated with response options. *Journal of Official Statistics*, 21(4), 701-725.

Haunberger, S. (2014). Item nonresponse in face-to-face interviews with children. *Journal of Official Statistics*, 30, 459–477.

Haveman, R., and Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature*, 33, 1829–1878.

Houtenville, A.J., and Conway, K.S. (2008). Parental effort, school resources, and student achievement. *Journal of Human Resources,* 43 Issue: 2 (2008-04-01) I 437-453

Hox, J.J., de Leeuw, E.D., and Chang, H.-T. (2012). Nonresponse versus measurement error: Are reluctant respondents worth pursuing? *Bulletin de Methodologie Sociologique*, 113, 5-19.

Kaminska, O., Mccutcheon, A.L., and Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74(5), 956–984.

König, S. (2011). *Online-Forschung mit Kindern*. Budrich UniPress.

Lam, G. (2014). A theoretical framework of the relation between socioeconomic status and academic achievement of students. *Education*, 134(3), 326-331

Lord, F. M, and Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison- Wesley.

Mavletova, A. (2015a). Web surveys among children and adolescents: Is there a gamification effect? *Social Science Computer Review*, online first.

Mavletova, A. (2015b) A gamification effect in longitudinal web surveys among children and adolescents. *International Journal of Market Research*, 57 (3), p. 413–438.

Punch, S. (2002). Research with children: The same or different from research with adults? Childhood, 9(3): 321-341.

Rendtel, U., Nordberg, L., Jäntti, M., Hanish, J., and Basic, E. (2004). Report of quality of income data. Chintex. Working paper 21.

Schräpler, J-P. (2003). Respondent behaviour in panel studies: a case study for income-nonresponse by means of the British Household Panel Study. ISER Working Paper Series, No. 2003-08.

Scott, J. (2008). Children as respondents: The challenge for quantitative methods. In P. Christensen, and A. James (Eds.), *Research with children: Perspectives and practices* (pp. 87-108). New York: Routledge.

Seginer, R. (1983) Parents' educational expectations and children's academic achievements: A literature review. *Merrill-Palmer Quarterly,* 29(1), 1-23.

Smith, K., and Platt, L. (2013). How do children answer questions about frequencies and quantities? Evidence from a large-scale field test. Centre for Longitudinal Studies.

Sturgis, P., Allum, N., and Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In *Methodology of Longitudinal Surveys,* Edited by Peter Lynn. John Wiley & Sons, pp. 113-126.

Torche, F., Warren, J.R., and Halpern-Manners, A. (2012). 'Panel conditioning in a Longitudinal Study of Chilean Adolescents' Substance Use: Evidence from an experiment. *Social Forces*, 90, 891-918.

Triplett T., Blair, J., and Kang, Y.C. (1996). Initial cooperators vs. converted refusers: Are there response behavior differences? Proceedings of the Survey Research Methods Section, American Statistical Association, 1038-41.

Vogl, S. (2012). Children between the age of 5 and 11: what “don’t know” answers tell us. *Quality & Quantity*, 46(4), 993–1011.

Vogl, S. (2013). Telephone versus face-to-face interviews: A mode comparison of semi-structured interviews with children. *Sociological Methodology*, 43(1), 138-182.

Waterman, A.H., Blades, M., and Spencer, C. (2001). Interviewing children and adults: the effect of question format on the tendency to speculate. *Applied Cognitive Psychology*, 15, 521–531.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461-481.

Appendix

Procedures and measures

We measured data quality based on the following indicators:

Hypothesis 1:

- The overall item nonresponse rates.

The overall item nonresponse rate was calculated as the number of missing items divided by the number of items a particular respondent was expected to answer. Conditional questions were included only if a respondent was expected to answer them. We conducted a multivariate linear regression to predict the overall item nonresponse rates in each wave. In addition, we analyzed separately items which produced a high item nonresponse rate.

- Selecting the “don’t know” response category.

Though there were a number of items with the “don’t know” response category in each wave respondents tended to give an answer or skip the question rather than explicitly say they don’t know. Only 0.1% of respondents selected the “don’t know” response in the first wave and 8.9% in the second. No “don’t know” responses were selected in the third and fourth waves. The only question in which the “don’t know” category was selected was the close-ended question about the plans at age 16: “At the moment, young people can leave school at 16. What would you most like to do when you are 16?”

Based on the results of the second wave the question was reformulated: “The age young people must stay in education or training differs somewhat across the UK. What would you most like to do when you have completed your final GCSE/Standard Grade year at around age 16?” The responses were also changed, providing more details for each response category. As a result, no respondents selected the “don’t know” response category in the subsequent waves. We conducted a multivariate logistic regression to predict the selection of the “don’t know” response in the second wave.

- Rounding

In each wave respondents were asked about the number of close friends they could talk to if they were in some kind of trouble. Using a multivariate logistic regression we predicted rounding, that is responses that are a multiple of 5 (5, 10, 15, or 20). Those respondents

who reported to have more than 20 friends (about 5% of the sample), missed the question or typed in “0” were excluded from the analysis.

- Inconsistency

Two questions were used to measure response consistency across waves. In each wave respondents were asked if they have ever smoked cigarettes and if they have ever had an alcoholic drink. Responses were deemed to be inconsistent if a respondent answered at one wave that they had ever done one of these things but then answered at a later wave that they had never done this thing. We ran multivariate logistic regressions to predict inconsistent responses for each of the two questions separately. We included only those respondents who completed at least two waves of the study.

To test hypothesis 1, for each data quality indicator we ran models which included as predictors some variables from the mother’s interview such as mother’s expectations, socio-demographic variables (race, level of education, employment status), parent–child relationship reported by mother, and some household variables. Some of the questions were asked biennially. As a result, the models are similar in the first and third waves and in the second and fourth waves. Below we outline the variables included in the analysis. We describe if variables were reported by mother or her child and in which waves they were reported.

- *Socio-economic status:* gross household income (month before interview); owning or renting home by the household (own home=1); mother’s level of education (diploma in higher education=1, less than higher education=0).

- *Mother expectations* (waves 1 and 3): evaluation the importance for her child to complete A level exams (or Higher Grades in Scotland) (very important=1, lower than “very important”=0).

- *Parental involvement in education* (waves 1 and 3) was measured by such items as “my parents are interested in how I do at school” (always or nearly always=1, less often=0), “my parents come to school parent evenings” (always or nearly always=1, less often=0) reported by children and help with homework reported by mother (once a week or more often=1, less often=1).

Parental involvement in education (waves 2 and 4) was measured by the only variable “help with homework” reported by respondents (someone at home helps with homework=1, does not help with homework=0).

- *Parent–child relationship* (waves 1 and 3): spending time together on leisure activities with child reported by mother (several times a week or almost every day=1, less often=0), talking with child about things that matter to him/her (most days=1, less often=0), and quarrel with child (less than once a week=1, more often=0).

Parent–child relationship (waves 2 and 4): based on six statements ("we discuss books at home", "we discuss TV programmes we have watched at home", "my parents/other adults at home buy me books as gifts", "my parents/other adults take me to museums or art galleries", "my parents/other adults take me to watch sporting events", and "my parents/other adults take me to the theatre or to see a dance performance or classical music") in which children were expected to evaluate the frequency from "never" (1) to "often" (4) the overall summed score was computed for each respondent. The sum scale varied from 6 (very poor parent–child relationship) to 24 (very good relationship).

- *Household socio-demographic variables*: marital status of the parents (parents are married and live together=1, other=0); number of children under 15 years old in the household; mother's employment status (employed/self-employed=1, other=0); mother's race (white=1, other=0); and the area respondent lives in (urban area=1, rural area=0).

We used unweighted data in all models.

Hypothesis 2:

- Panel conditioning effect

The latent construct of happiness with school was measured by two indicators: how happy children are with school work and how happy they are with school overall (7-point scale). We run structural equation models to measure a panel conditioning effect. We focus on the correlations between residual variances. An increase from wave to wave in the correlations between residual variances would indicate a panel conditioning effect. The analysis is based on those respondents who completed all four waves (N=956). Since the age has an effect on responses we controlled for age in each wave.

We compared three models. The first model (basic model) set correlations between residual variances equal to zero (see Fig.1a). The second and third models assumed autoregressive change (Alwin, 2007; Cernat, 2015). In the second model, the true score for happiness with school at time (i) is influenced only by a true score at time (i-1). The correlations between residual variances at time (i) and at time (i-1) were estimated (see Fig.1b). The third model assumed that the true score for happiness with school at time (i)

is influenced by the true score at time (i-1) and at time (i-2). The correlations between residual variances at time (i) and at time (i-1), as well as at time (i) and at time (i-2) were estimated (see Fig.1c).

Fig.1a. Model 1: no correlations between residual variances

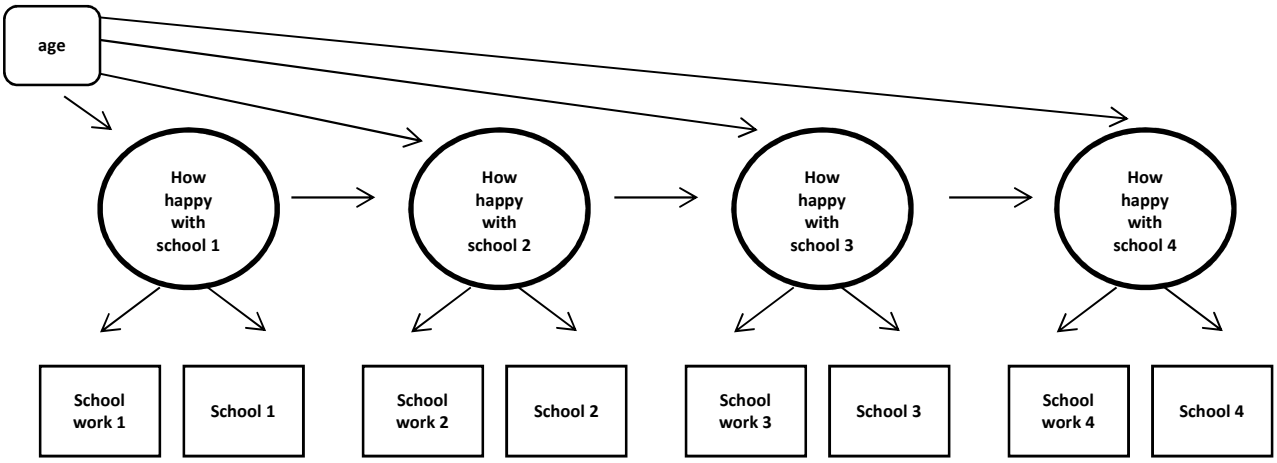


Fig.1b. Model 2: autoregressive change, correlations between residual variances

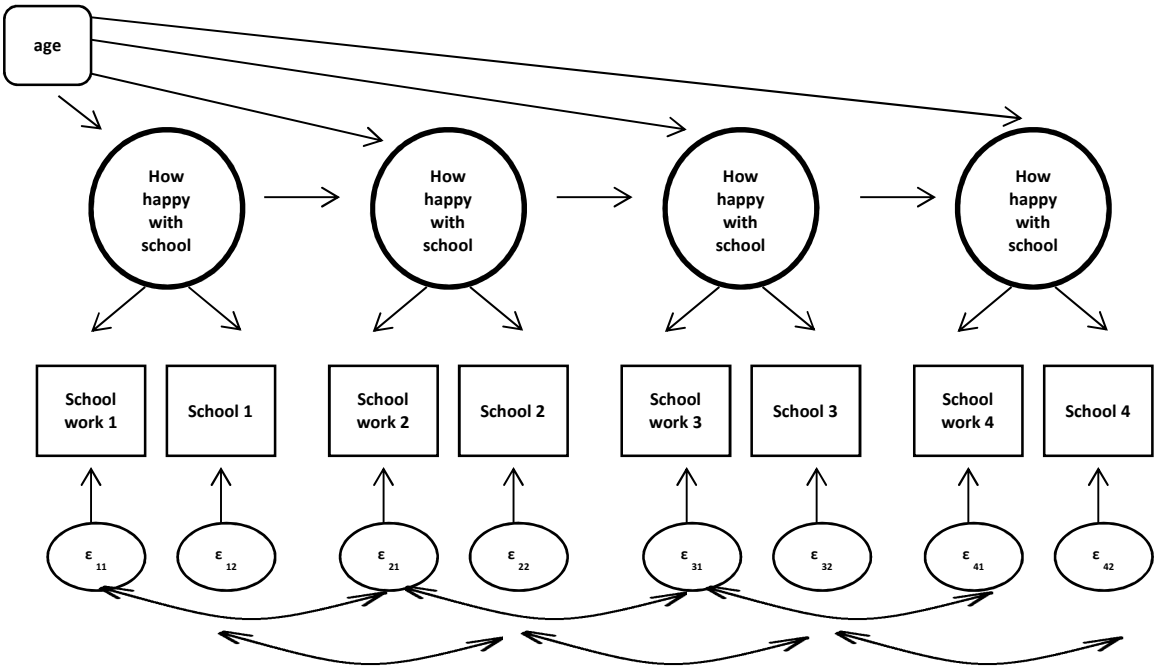
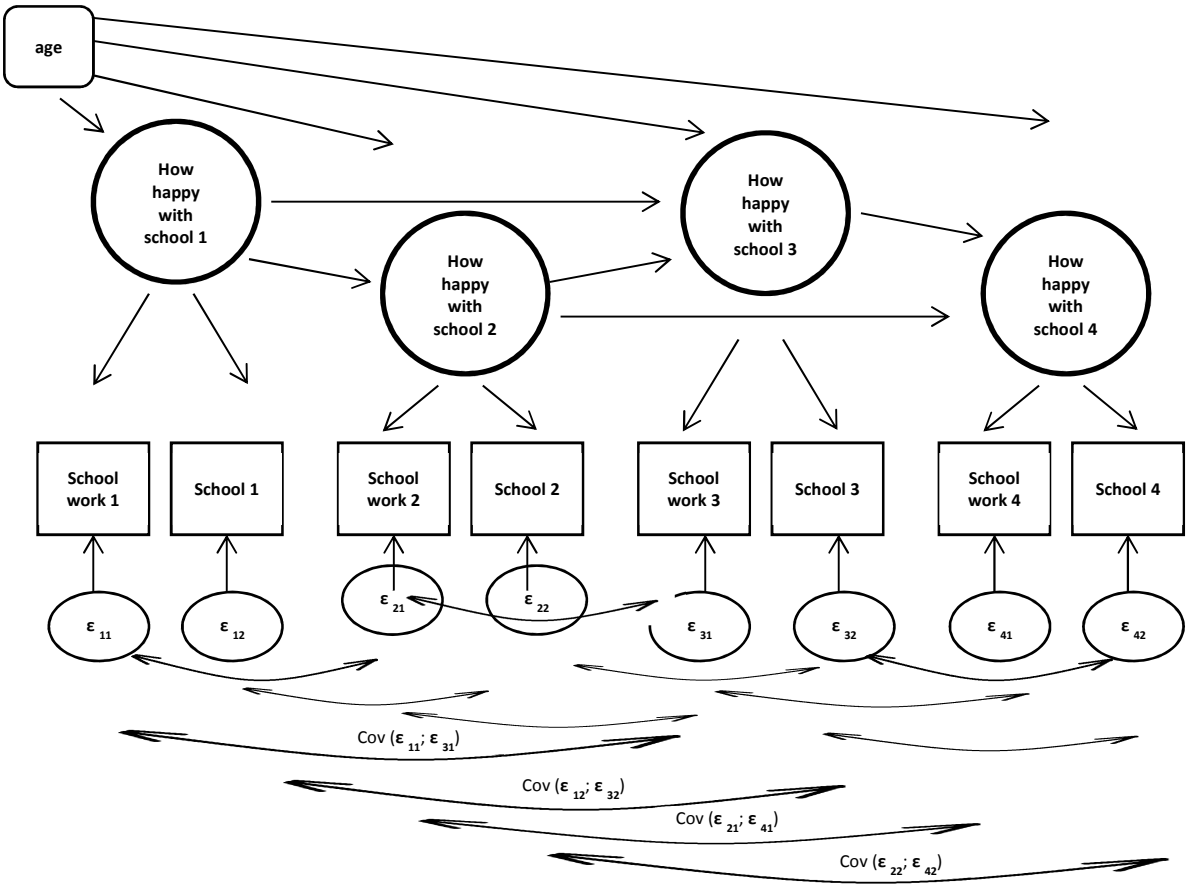


Fig.1c. Model 3, correlations between residual variances at time (i), (i-1), and (i-2)



Based on the model fit statistics, model 3 was selected for analyzing panel conditioning effect (CFI=0.992, TLI=0.969, RMSEA=0.043, SRMR= 0.019, see Table 8).

- Social desirability bias

To measure an increase in social desirability bias in the subsequent waves we ran mixed-effects logit regressions to predict the level of reporting of sensitive behavior based on three items that were asked in all four waves. Four other items on alcohol consumption and drug-taking have also been asked on the survey, but none of these were included in more than two waves. The three items included in the analysis are:

- having ever drunk alcohol;

- having ever smoked;

- playing truant in the last 12 months.

To disentangle panel conditioning effect from the attrition we included only those who completed all four waves (balanced panel analysis).

Table 8: Goodness-of-fit of the models

Model 1 [no correlations between residual variances]	Model 2 [autoregressive change, correlations between residual variances at time (i), (i-1)]	Model 3 [correlations between residual variances at time (i), (i-1), and (i-2)]
$\chi 2(df=21)=337.05$	$\chi 2(df=36)=118.14$	$\chi 2(df=9)=25.24$
CFI=0.847	CFI=0.950	CFI=0.992
TLI=0.738	TLI=0.880	TLI=0.969
RMSEA=0.125	RMSEA=0.085	RMSEA=0.043
SRMR= 0.051	SRMR= 0.034	SRMR= 0.019

Hypothesis 3:

- Reluctance

Reluctance to participate in the survey was measured by the total number of face-to-face interviewer calls made prior to completion of the household interviews. To test hypothesis 3, we fitted regression models of item nonresponse rates and rounding, as described above for hypothesis 1 but additionally including the indicator of total number of calls. A significant coefficient for this indicator would suggest support for hypothesis 3.