

Item Nonresponse Rates and Panel Conditioning in a Longitudinal Survey among Youth

Aigul Mavletova

Laboratory for Comparative Social Research and Department of Sociology

National Research University Higher School of Economics

Myasnickaya street 20

Moscow, Russian Federation, 101000, Russia

Telephone: +7(495)772-95-90, 12244

e-mail: amavletova@hse.ru

Peter Lynn

Institute for Social and Economic Research

University of Essex

Wivenhoe Park

Colchester, Essex, CO4 3SQ UK

Telephone: 01206 874809

e-mail: plynn@essex.ac.uk

Online Appendix

Data

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 (1.71)	12.5 (1.67)	12.5 (1.70)	12.6 (1.69)
Males	49.8%	50.5%	50.2%	50.5%
<i>Socioeconomic status and household variables</i>				
Gross household income (month before interview, UK pounds)	3,710.9 (2,652.1)	3,926.9 (2,676.5)	4,202.6 (2,665.6)	4,266.2 (2,712.0)
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%

Standard deviation in parentheses

Questionnaires

The total number of items varied from 88 to 104 in different waves (see Table 2). Overall, the questions focus on health behavior, 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 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

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 multiple linear regression to predict the overall item nonresponse rates in each wave. In addition, we analyzed separately items that produced a high item nonresponse rate.

We included some variables from the mother's interview such as mother's expectations, sociodemographic variables (race, level of education, employment status), parent-child relationship reported by mother, and some household variables as predictors. 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.

—*Socioeconomic 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 the 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=0).

Parental involvement in the 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 sociodemographic 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

- 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 multiple 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.

- 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, $N=960$).

- Correlations between residual variances

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. Since age has an effect on responses, we controlled for age in each wave. We used lavaan package in R software

environment for estimating the model (see <http://lavaan.ugent.be>). We included only those who completed all four waves in the analysis ($N=960$).

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

Fig.1a. Model 1: autoregressive change, no correlations between residual variances

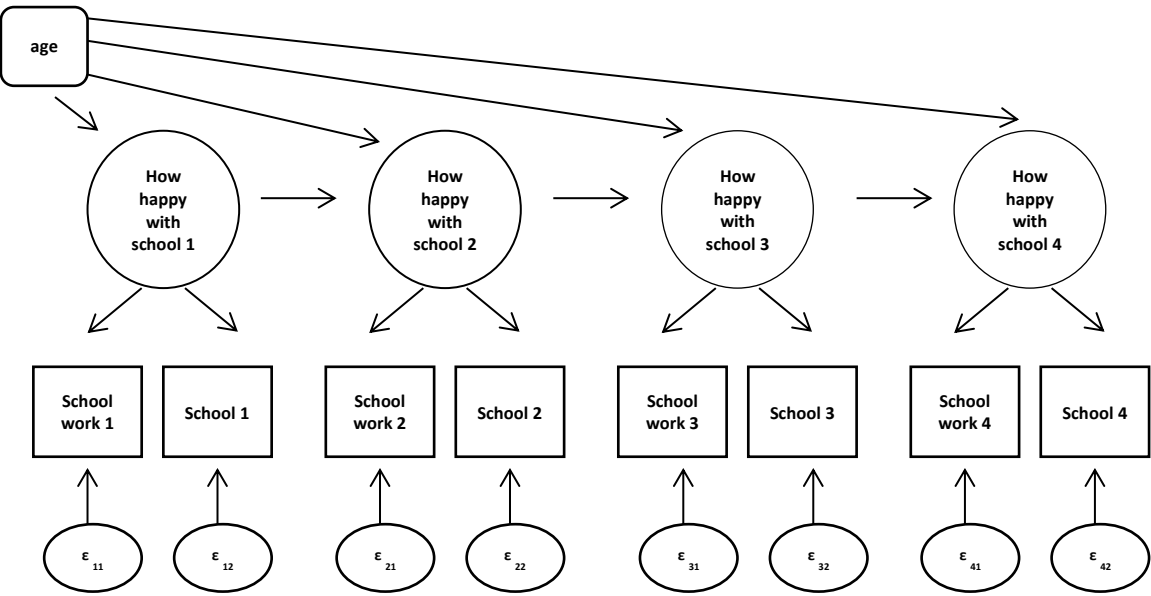


Figure 1b. Model 2: autoregressive change, correlations between residual variances at time (i) and at time (i-1)

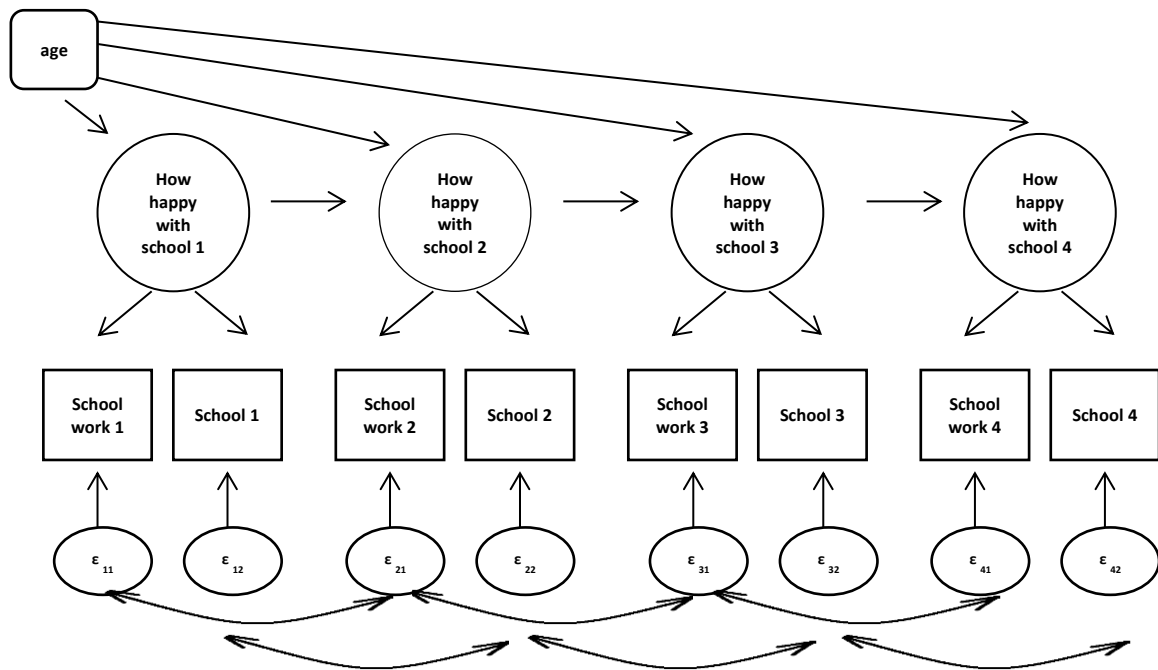
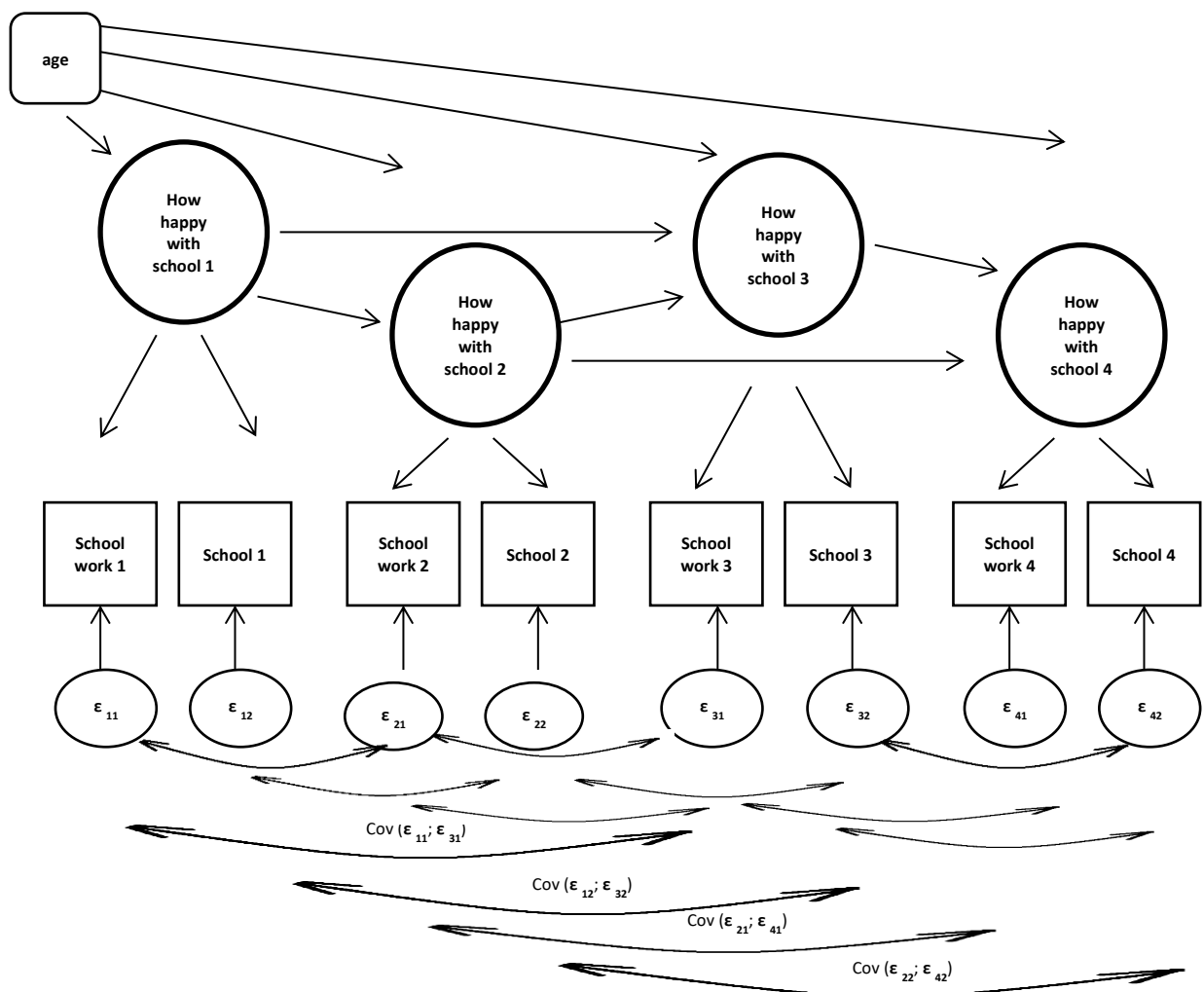


Figure 1c. Model 3: autoregressive change, correlations between residual variances at time (i), (i-1), and (i-2)



Based on the model fit statistics, model 3 was selected for analysing panel conditioning effect (CFI=0.992, TLI=0.967, RMSEA=0.044, SRMR= 0.019, see Table 3).

Table 3: 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)]
$\chi^2(df=21)=331.46$	$\chi^2(df=15)=119.56$	$\chi^2(df=9)=25.99$
CFI=0.849	CFI=0.950	CFI=0.992
TLI=0.714	TLI=0.879	TLI=0.967
RMSEA=0.131	RMSEA=0.085	RMSEA=0.044
SRMR= 0.051	SRMR= 0.034	SRMR= 0.019

Table 4: Predicting inconsistency in the questions on smoking and alcohol consumption (Odds Ratios)

	Inconsistency in the question on smoking	Inconsistency in the question on alcohol consumption
Intercept	0.00*** (0.00–01)	0.01 (0.00–0.03)
<i>Respondent's characteristics</i>		
Males	1.11 (0.87–1.42)	0.97 (0.84–1.13)
Age	1.57*** (1.40–1.76)	1.33*** (1.25–1.41)
<i>Socioeconomic status</i>		
Gross household income (month before interview)	1.00** (1.00–1.00)	1.00 (1.00–1.00)
Own home	0.64** (0.48–0.86)	0.91 (0.75–1.09)
Mother has diploma in higher education	1.21 (0.89–1.64)	0.96 (0.79–1.17)
<i>Mother expectations (waves 1 and 3)</i>		

Importance for your child to complete A level exams: very important	1.09 (0.83–1.44)	0.94 (0.80–1.11)
<i>Parental involvement in education</i>		
My parents are interested in how I do at school: always or nearly always	0.73* (0.54–0.99)	0.91 (0.74–1.11)
My parents come to school parent evenings: always or nearly always	0.63** (0.46–0.86)	0.78* (0.62–0.98)
Mother helps with homework: once a week or more often	0.84 (0.64–1.11)	0.82* (0.69–0.98)
Someone at home helps with homework	0.98 (0.75–1.27)	0.67*** (0.56–0.81)
<i>Parent–child relationship</i>		
Spending time together on leisure activities with mother: several times a week or almost every day	0.81 (0.59–1.13)	1.02 (0.85–1.23)
Quarrel with child: less than once a week	0.80 (0.62–1.02)	0.96 (0.82–1.12)
The child talks with mother about things that matter to him/her: most days	1.04 (0.81–1.35)	0.97 (0.83–1.13)
Discussing books at home, discussing TV programmes, buying books as gifts, and the like	0.95** (0.91–0.98)	0.99 (0.97–1.01)
<i>Household sociodemographic variables</i>		
Parents are married and live together	0.76 (0.58–0.99)	0.96 (0.81–1.14)
Number of children under 15 in the household	0.93 (0.82–1.06)	1.06 (0.98–1.14)
Mother's employment status: employed/self-employed	0.78 (0.59–1.05)	1.16 (0.96–1.39)
Mother's race: white	1.69** (1.17–2.44)	1.81*** (1.45–2.26)
Urban area	0.92 (0.68–1.23)	1.26* (1.05–1.50)
N	3,489	3,489
R ²	Cox & Snell R Square=0.055 Nagelkerke R Square=0.123	Cox & Snell R Square=0.058 Nagelkerke R Square=0.082

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 95% confidence interval indicated in parentheses

References

- Alwin, D. F. 2007. *Margins of error: A study of reliability in survey measurement*. Hoboken, NJ: Wiley Blackwell.
- Cernat, A. The impact of mixing modes on reliability in longitudinal studies. 2015. *Sociological Methods & Research*, 44:427–57.