

## Appendix

Table A1: Overview of sample sizes and response rates in the NEPS adult cohort data

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5	
	First time respondents	Panel respondents	First time respondents	Panel respondents	Panel respondents	Panel respondents	First time respondents	Panel respondents		
Gross sample (GS)	22,564	8,995	17,967	12,195	11,390	17,106	15,249			
Complete Interview (I)	10,404	6,579	5,074	9,352	8,925	5,239	11,707			
Partial Interview (P)	1	3	2	4	3	3	0			
Refusal and break-off (NR)	9,272	1,655	10,505	2,291	2,210	9,882	3,061			
Not able (NA)	253	21	541	32	18	320	21			
Non-contact (NC)	253	681	0	455	174	399	377			
Unknown eligibility (UE)	2356	0	1,677	0	0	1,087	0			
Not eligible (NE) <sup>a</sup>	25	56	168	61	60	176	83			
AAPOR RR1 <sup>b</sup>	46.2%	73.6%	28.5%	77.1%	78.8%	30.9%	77.2%			

Notes: <sup>a</sup> Persons are considered not eligible if they are deceased or if they have moved abroad permanently. Among first-time respondents, persons born before 1944 or after 1986 were coded non-eligible, as they are not part of the survey population.

<sup>b</sup> Response rates were calculated according to AAPOR standards (AAPOR 2016). We report RR1, also called “minimum response rate”, which is the most restrictive response rate proposed by AAPOR. It is defined as “the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus non-contacts plus others) plus all cases of unknown eligibility” (AAPOR 2016:61).

Table A2: Modeling random assignment to experimental groups in wave 3

Variable	Both tests		Math only		Reading only	
	Coef.	z	Coef.	z	Coef.	z
Sex: female	0.003	(0.30)	-0.002	(-0.18)	-0.001	(-0.18)
Age in years (mean centered)	-0.000	(-0.50)	-0.000	(-0.44)	0.000	(0.98)
Native language: non-German only	-0.033	(-1.45)	-0.005	(-0.26)	0.038*	(2.10)
Number of children in household	0.002	(0.37)	-0.002	(-0.45)	0.000	(0.03)
Educational attainment (ref. tertiary)						
No voc. training (max. med sec. school)	0.008	(0.37)	0.001	(0.07)	-0.009	(-0.57)
Lower secondary + voc. training	0.003	(0.18)	0.018	(1.48)	-0.020	(-1.73)
Medium secondary + voc. training	0.006	(0.46)	0.003	(0.27)	-0.009	(-0.80)
Higher secondary school	0.002	(0.18)	0.007	(0.52)	-0.010	(-0.83)
Emp. status and working time (ref. 35-44 hrs./week)						
Not working	0.008	(0.64)	-0.001	(-0.12)	-0.007	(-0.60)
Up to 18 hrs./week	-0.018	(-0.82)	0.027	(1.39)	-0.009	(-0.47)
19-34 hrs./week	0.011	(0.73)	0.011	(0.87)	-0.022	(-1.58)
45 hrs./week and more	0.016	(1.25)	-0.008	(-0.69)	-0.008	(-0.70)
Interviewer age (ref. 50 years and over)						
Up to 29 years	0.011	(0.44)	-0.019	(-0.85)	0.007	(0.42)
30 to 49 years	0.003	(0.25)	-0.001	(-0.05)	-0.002	(-0.23)
Interviewer workload (ln, mean centered)	-0.016*	(-2.08)	0.003	(0.48)	0.013	(1.96)
Interviewer success rate (ln, mean centered)	0.040	(1.46)	-0.023	(-1.05)	-0.016	(-0.69)
Resp. uncooperative, prev. interview	0.007	(0.38)	0.005	(0.32)	-0.013	(-0.70)
N contacts, previous wave (ln, mean centered)	-0.002	(-0.42)	0.005	(1.26)	-0.003	(-0.70)
N refused items, previous interview (ln)	-0.001	(-0.07)	0.001	(0.09)	-0.000	(-0.00)
N don't knows, previous interview (ln)	0.002	(0.23)	0.004	(0.47)	-0.006	(-0.68)
Subsample (ref. 2007 with test)						
2007 without test	-0.007	(-0.60)	0.010	(0.95)	-0.003	(-0.26)
2009	-0.000	(-0.02)	-0.000	(-0.00)	0.000	(0.02)
Interview mode, previous wave (ref. telephone)						
Face-to-face	-0.012	(-0.84)	0.010	(0.84)	0.001	(0.09)
Temporary dropout	-0.050*	(-2.33)	0.028	(1.37)	0.022	(1.26)

N = 11,460; Pseudo R<sup>2</sup> (McFadden) = 0.001

\* p&lt;0.05, \*\* p&lt;0.01, \*\*\* p&lt;0.001

Notes: Multinomial logistic regression, average marginal effects, robust standard errors (clustered by interviewer).

Source: NEPS starting cohort 6 wave 1-3 consortium data, authors' estimations.

Table A3: Control variables and descriptive statistics

Variable	Wave 3		Wave 4		Wave 5	
	Mean	SD	Mean	SD	Mean	SD
Test result wave 3 (percent correct)*	0.57	0.22	0.57	0.22	0.57	0.22
Female	0.51	0.50	0.51	0.50	0.51	0.50
Age in years (mean centered)	0.00	11.07	0.00	11.04	0.00	10.93
Native language Non-German only	0.05	0.22	0.05	0.21	0.04	0.21
Number of children in household	0.85	1.01	0.83	1.00	0.82	1.00
Educational attainment						
No voc. training (max. med sec. school)	0.06	0.23	0.05	0.23	0.05	0.22
Lower secondary + voc. training	0.19	0.39	0.19	0.39	0.18	0.39
Medium secondary + voc. training	0.30	0.46	0.30	0.46	0.30	0.46
Higher secondary school	0.17	0.38	0.17	0.37	0.17	0.37
Tertiary education (ref.)	0.27	0.45	0.29	0.45	0.29	0.45
Employment status and working time						
Not working	0.27	0.44	0.27	0.44	0.26	0.44
Up to 18 hrs./week	0.06	0.23	0.05	0.22	0.05	0.23
19-34 hrs./week	0.15	0.36	0.15	0.35	0.15	0.36
35-44 hrs./week (ref.)	0.31	0.46	0.31	0.46	0.29	0.45
45 hrs./week and more	0.22	0.41	0.22	0.42	0.24	0.43
Interviewer age						
Up to 29 years	0.05	0.21	0.26	0.44	0.04	0.20
30 to 49 years	0.26	0.44	0.40	0.49	0.21	0.41
50 years and over (ref.)	0.69	0.46	0.34	0.47	0.75	0.43
Interviewer workload (ln, mean centered)	0.00	0.72	0.00	0.82	0.00	0.55
Interviewer success rate (ln, mean centered)	0.00	0.18	0.00	0.19	0.00	0.20
Respondent uncooperative, previous interview	0.06	0.23	0.09	0.29	0.05	0.22
N contacts, previous wave (ln, mean centered)	0.00	1.08	0.00	0.78	0.00	0.95
N refused items, previous interview (ln)	0.09	0.31	0.12	0.34	0.09	0.31
N don't knows, previous interview (ln)	0.32	0.53	0.12	0.33	0.17	0.39
Subsample						
2007 with test (ref.)	0.28	0.45	0.29	0.45	0.30	0.46
2007 without test	0.32	0.47	0.31	0.46	0.31	0.46
2009	0.39	0.49	0.40	0.49	0.39	0.49
Interview mode, previous wave**						
Telephone (ref.)	0.79	0.41	---	---	0.82	0.38
Face-to-face	0.16	0.36	---	---	0.05	0.22
Temporary dropout	0.05	0.23	0.03	0.17	0.12	0.33
N	11,460		10,621		9,623	

\* N wave 4 = 7,096; N wave 5 = 6,846.

\*\* We did not include control variables for interview mode in the previous wave in the wave 4 models as participation mode in wave 3 (our main variable of interest) and interview mode in the previous wave (for wave 4 also wave 3) are closely related so that including both in the same model would not have been feasible. Information on the distribution of participation mode in wave 3 for wave 4 can be found in Table A4.

Source: NEPS starting cohort 6 wave 1-5 consortium data, authors' estimations.

Table A4: Test assignment and participation mode in wave 4 (N and %)

Part. mode (w3)	Test group (w3)	Participation (w4): N			Participation (w4): %		
		Interview	Nonresp.	Total	Interview	Nonresp.	Total
Nonresponse	One test	323	423	746	43.3%	56.7%	100.0%
	Two tests	310	518	828	37.4%	62.6%	100.0%
Telephone	One test	695	123	818	85.0%	15.0%	100.0%
	Two tests	845	136	981	86.1%	13.9%	100.0%
FtF w/o test	One test	68	18	86	79.1%	20.9%	100.0%
	Two tests	139	31	170	81.8%	18.2%	100.0%
FtF w test	One test	3,250	374	3,624	89.7%	10.3%	100.0%
	Two tests	3,012	356	3,368	89.4%	10.6%	100.0%
<b>Total</b>		8,642	1,979	10,621	81.4%	18.6%	100.0%

Source: NEPS starting cohort 6 wave 4 consortium data, authors' estimations.

Table A5: Test assignment and participation mode in wave 5 (N and %)

Part. mode (w3)	Test group (w3)	Participation mode (w5)				Total
		Nonresponse	Telephone conversion	FtF, test aborted/refused	FtF with test	
Nonresponse	One test	221	56	9	166	452
		48.9%	12.4%	2.0%	36.7%	100.0%
	Two tests	250	59	12	156	477
		52.4%	12.4%	2.5%	32.7%	100.0%
Telephone conversion	One test	228	145	26	369	768
		29.7%	18.9%	3.4%	48.0%	100.0%
	Two tests	253	216	26	440	935
		27.1%	23.1%	2.8%	47.1%	100.0%
FtF, test aborted/refused	One test	23	10	4	43	80
		28.7%	12.5%	5.0%	53.8%	100.0%
	Two tests	38	20	8	93	159
		23.9%	12.6%	5.0%	58.5%	100.0%
FtF with test	One test	428	244	108	2,719	3,499
		12.2%	7.0%	3.1%	77.7%	100.0%
	Two tests	420	218	79	2,536	3,253
		12.9%	6.7%	2.4%	78.0%	100.0%
<b>N total</b>		1,861	968	272	6,522	9,623
<b>% total</b>		19.3%	10.1%	2.8%	67.8%	100.0%

Source: NEPS starting cohort 6 wave 5 consortium data, authors' estimations.

Table A6: Effect of test assignment and response mode in NEPS wave 3 on response mode in NEPS wave 5, AME of response mode in the two test groups and Wald test of difference

	One test	Two tests	Difference
<i>Outcome w5: Nonresponse</i>			
Nonresponse	0.200***	0.202***	0.002
Telephone conversion	0.136***	0.104***	-0.032
FtF, test aborted/refused	0.098*	0.063 <sup>+</sup>	-0.035
<i>Outcome w5: Telephone conversion</i>			
Nonresponse	0.047**	0.068***	0.021
Telephone conversion	0.079***	0.108***	0.029
FtF, test aborted/refused	0.043	0.053*	0.011
<i>Outcome w5: FtF, test aborted/refused</i>			
Nonresponse	-0.006	0.009	0.014
Telephone conversion	0.006	0.006	0.000
FtF, test aborted/refused	0.017	0.024	0.007
<i>Outcome w5: FtF with test</i>			
Nonresponse	-0.241***	-0.278***	-0.037
Telephone conversion	-0.221***	-0.218***	0.003
FtF, test aborted/refused	-0.158**	-0.141***	0.017

N = 9,623

+p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

Notes: Multinomial logistic regression, average marginal effects (AME), robust standard errors (clustered by interviewer); including interaction effect of response mode and test assignment; additional control variables (effects not shown): age, sex, education, native language, number of children in the household, employment status and working time, subsample, whether respondents were temporary dropouts in the previous wave.