[NEPS
National Educational Panel Study

Thorsten Schnapp

Samples, Weights and Nonresponse

NEPS Starting Cohort 3 — Grade 5

Paths Through Lower Secondary School — Educational Pathways of Students in Grade 5 and Higher

Wave 11

LIfBi

**LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES**

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 11)

Thorsten Schnapp, Leibniz Institute for Educational Trajectories

E-mail address of lead author:

statistik@lifbi.de

1. Prequel

This report complements NEPS Survey Paper No. 63 (Steinhauer & Zinn, 2016a) and gives details on wave 11 of Starting Cohort 3 (SC3) of the National Educational Panel Study (NEPS). It refers to the Scientific Use File (SUF; DOI:10.5157/NEPS:SC3:11.0.0). SC3 focuses on students in Grade 5 and their pathway through lower secondary education. The original sample consists of a main sample of Grade 5 students in regular schools and special-needs schools with a supplement covering students with a migration background from Turkey and the former Soviet Union. Due to the Federal-State-specific timing in transition in lower secondary education in regular schools a refreshment sample was drawn for students attending Grade 7.

To provide weights for the original samples as well as for the refreshment sample the different processes leading to the participation decision in a certain wave have to be considered. These decision processes include the schools initial decision to participate in the survey, the students initial decision to participate in the survey, and the students successive decisions to participate in each wave again. The schools initial decision to participate enters a nonresponse adjusted design weight on the institutional level. The students initial decision to participate enters a nonresponse adjusted design weight on the individual level. The successive decisions of a student to participate in a certain wave enter the corresponding wave-specific cross-sectional and longitudinal weights.

The students willingness to participate in the panel study (i.e., the panel members) are followed up over time. In the progress of the panel it is possible that students cannot be surveyed within their institutional context for several reasons. For example, the school decides to refuse further cooperation or the students switch to another school. In these cases students are surveyed in an individual context, that is, the questionnaires are sent to their home address. Surveying students in this individual context is referred to as the field of individual retracking.

Table 1 illustrates the number of students according to the sample they originally belong to and their participation status by wave. The table gives details on the size of panel cohort over time. The column "Used sample" is split up into "Participants", "Temporary dropouts", and "Final dropouts (in wave)", displaying the students status by the end of the wave. Finally, the last column presents the number of students withdrawing their panel consent between two rounds of survey waves.

For details on the sampling design and the derivation of design weights see Steinhauer, Aßmann, Zinn, Goßmann, and Rässler (2015). Details on calculating wave-specific nonresponse adjustments can be found in Steinhauer and Zinn (2016a) for waves 1 to 3, in Steinhauer and Zinn (2016b) for waves 1 to 5, in Steinhauer (2017) for waves 6 and 7, in Steinhauer (2019) for wave 8, in Hammon and Landrock (2019) for wave 9, Schnapp (2020) for wave 10 and Würbach, Landrock, Schnapp, Ziesmer, and Bergrab (2021) for the additional Corona survey.

2. Changes compared to previous version

Weights for wave 11 have been appended. Note that starting with wave 8 an AIC based backward selection is in use, adapting the initial model for estimating individual nonresponse propensities.

Table 1: Panel progress of SC3 by wave.

Wave (Time)	Study number		Total size	Not used	Used sample	Participants	Temporary dropout	Final dropout (in wave)	Final dropout (after wave)
			Panel Cohort			**Status at the end of the wave**			
1 (2010/2011)	(A28, A56, A63)	**Main**	6112	0	6112	5778	334	0	13
2 (2011/2012)	(A29, A57)	**Main**	6099	0	6099	5538	560	1	8
3 (2012/2013)	(A30, A30A, A58)	**All**	8295	0	8295	7277	989	29	10
		Main	6090	0	6090	5131	930	29	10
		Refr.	2205	0	2205	2146	59	0	0
4 (2013/2014)	(A31, A59)	**All**	8256	0	8256	6718	1505	33	580
		Main	6051	0	6051	4783	1249	19	[a]580
		Refr.	2205	0	2205	1935	256	14	0
5 (2014/2015)	(A94)	**All**	7643	0	7643	5778	1625	240	0
		Main	5452	0	5452	4001	1273	178	0
		Refr.	2191	0	2191	1777	352	62	0
6 (Spring 2015)	(A98)	**All**	7403	0	7403	5586	1740	77	2
		Main	5274	0	5274	3920	1293	61	2
		Refr.	2129	0	2129	1666	447	16	0
7 (2015/2016)	(A99, B106)	**All**	7324	244	7080	5492	1543	45	29
		Main	5211	153	5058	3925	1104	29	21
		Refr.	2113	91	2022	1567	439	16	8
8 (2016/2017)	(A100, B107)	**All**	7250	65	7185	5263	1562	360	244
		Main	5161	42	5119	3767	1095	257	188
		Refr.	2089	23	2066	1496	467	103	56
9 (2017/2018)	(A101, B108)	**All**	6646	1	6645	4988	1184	473	512
		Main	4716	0	4716	3590	812	314	346
		Refr.	1930	1	1929	1398	372	159	166
10 (2018/2019)	(B132)	**All**	5661	1	5660	3846	1516	298	238
		Main	4056	0	4056	2774	1081	201	172
		Refr.	1605	1	1604	1072	435	97	66

11 (2019/2020)	(B133)	**All**	5125	1	5124	3292	1629	203	799
		Main	3683	0	3683	2393	1161	129	567
		Refr.	1442	1	1441	899	468	74	232

[a] special-need students are excluded from the panel cohort after Wave 4.

3. Participation in Wave 11

To account for the wave-specific participation decision of students response propensity re-weighting is used to provide corresponding weights. To model binary participation decisions a model with a probit link function is used and adapted via a stepwise selection for wave 8 and subsequent waves, see Steinhauer and Zinn (2016a) and Steinhauer and Zinn (2016b) for origins and details. By Wave 11 the panel cohort has reduced to 5,125 students, see Table 1. Like in NEPS Starting Cohort 4 these students leave their schools and thus are surveyed individually. The significant coefficients for the estimated models are displayed in Table 2.

Thus, we can see that having participated in previous waves significantly influences the participation decision in wave 11 positively for both groups, except with respect to students that participated in wave 5. The latter decision of taking part in the survey's wave 5 has only a significant influence for students of the main sample. Additionally and again for both groups, students have a higher participation propensity when being part of the younger half of the age group compared to the older half.

Table 2: Models estimating the individual participation propensities for students in Wave 11 of SC3 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights.

| | Wave 11 | |
	Main Sample	Refreshment Sample
(Intercept)	-1.327^{***}	-1.492^{***}
	(0.135)	(0.168)
Age group younger half	0.111^{*}	0.150^{*}
	(0.046)	(0.073)
Student participated in Wave 5	0.139^{*}	
	(0.069)	
Student participated in Wave 8	0.295^{***}	0.469^{***}
	(0.075)	(0.113)
Student participated in Wave 9	0.242^{**}	0.414^{**}
	(0.088)	(0.126)
Student participated in Wave 10	1.247^{***}	1.303^{***}
	(0.053)	(0.083)

Notes: Reference categories are: Explicit stratum (SC3: Grade 5 and Grade 7), Age group (older half), Native language (German), Student participated in Wave t (no), Student in individual retracking (no). Students who have left the general school system prematurely and are also retracked by infas institute. To model individual participation, the `glm` function with a probit link provided in R (R Core Team, 2020) was used. ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parentheses. AIC based backward selection was used and only signficant coefficients are reported.

4. Summary of Weights

Various kinds of weights for students together with design information are provided by NEPS. Table 3 summarizes the design information and the different weights provided by SUF release version DOI:10.5157/NEPS:SC3:11.0.0. Besides individual/target (ID_t) and institutional (ID_i) identifiers, design information for the entire cohort is made available.[1] This information covers the study number corresponding to the first survey in which a student had been surveyed, the explicit sampling strata (stratum_exp) as well as the implicit sampling strata. Variables used for implicit stratification are "school type" (stratum_imp1), "federal state" (stratum_imp2_R), "regional classification" (stratum_imp3_R) and "funding" (stratum_imp4_R). Since release version 6.0.0 additional information has been added to the design data, namely the total number of students (tx80113_R) and classes (tx80114_R) in grade 8 in school year 2013/2014 as reported by official statistics.

Nonresponse adjusted design weights on the institutional (w_i) and the individual (w_t) level are given for the entire cohort.[2] For all participants in a particular wave, cross-sectional weights are provided. With respect to panel progress longitudinal weights are also available. Data from Official Statistics (Statistisches Bundesamt, Fachserie 11, Reihe 1, 2010/11) regarding the gender ratio in different school types of different Federal States has been used for raking.

The general overview of variables contained in the weighting data set can be found in Table 3. It is accompanied with summarizing statistics of respective general, cross-sectional and longitudinal weights presented in Table 4.

Table 3: Variables included in the weighting data of SC3 SUF version 11.0.0

Variable	Applies to	Content
ID_t	8317	Identifier for target person
ID_i	8317	Identifier for the institution
Design information		
tstud_st	8317	Study number the target person was first surveyed in
sample	8317	Part of the sample the target person belongs to
stratum_exp	8317	Explicit stratum referring to school
stratum_imp1	8317	Implicit stratum (school type according to sampling frame)
stratum_imp2_R	8317	Implicit stratum (federal state according to sampling frame)
stratum_imp3_R	8317	Implicit stratum (regional classification according to sampling frame)
stratum_imp4_R	8317	Implicit stratum (funding according to sampling frame)
tx80113_R	7670	Total number of classes in grade 8 as reported by official statistics
tx80114_R	7670	Total number of students in grade 8 as reported by official statistics
Design weights adjusted for initial nonresponse		
w_i	8317	Design weight for institution
w_t	8317	Design weight for target
w_t_cal	5283	Design weight for target, calibrated
w_t3_cal	8054	Design weight for target in Wave 3, calibrated

---

[1]Due to data protection, this information is not available in the download version of the SUF.

[2]The institutional weight as well as the explicit and implicit stratification variables belong to the institution and thus are equal for all cases within the institution.

Table 3: Variables included in the weighting data of SC3 SUF version 11.0.0 (continued)

Variable	Applies to	Content
Weights adjusted for wave-specific nonresponse, standardized		
w_t1	5559	Cross-sectional weight for targets participating in Wave 1
w_t2	5330	Cross-sectional weight for targets participating in Wave 2
w_t3	7111	Cross-sectional weight for targets participating in Wave 3
w_t4	6581	Cross-sectional weight for targets participating in Wave 4
w_t5	5648	Cross-sectional weight for targets participating in Wave 5
w_t6	5465	Cross-sectional weight for targets participating in Wave 6
w_t7	5367	Cross-sectional weight for targets participating in Wave 7
w_t8	5139	Cross-sectional weight for targets participating in Wave 8
w_t9	4870	Cross-sectional weight for targets participating in Wave 9
w_t10	3766	Cross-sectional weight for targets participating in Wave 10
w_t11	3231	Cross-sectional weight for targets participating in Wave 11
w_t1to2	5070	Longitudinal weight for targets participating in Wave 1 to 2
w_t1to3	4514	Longitudinal weight for targets participating in Wave 1 to 3
w_t1to4	4027	Longitudinal weight for targets participating in Wave 1 to 4
w_t1to5	3203	Longitudinal weight for targets participating in Wave 1 to 5
w_t1to6	2919	Longitudinal weight for targets participating in Wave 1 to 6
w_t1to7	2604	Longitudinal weight for targets participating in Wave 1 to 7
w_t1to8	2228	Longitudinal weight for targets participating in Wave 1 to 8
w_t1to9	1947	Longitudinal weight for targets participating in Wave 1 to 9
w_t1to10	1483	Longitudinal weight for targets participating in Wave 1 to 10
w_t1to11	1185	Longitudinal weight for targets participating in Wave 1 to 11
w_t3to4	6288	Longitudinal weight for targets participating in Wave 3 to 4
w_t3to5	5119	Longitudinal weight for targets participating in Wave 3 to 5
w_t3to6	4601	Longitudinal weight for targets participating in Wave 3 to 6
w_t3to7	4027	Longitudinal weight for targets participating in Wave 3 to 7
w_t3to8	3361	Longitudinal weight for targets participating in Wave 3 to 8
w_t3to9	2889	Longitudinal weight for targets participating in Wave 3 to 9
w_t3to10	2185	Longitudinal weight for targets participating in Wave 3 to 10
w_t3to11	1740	Longitudinal weight for targets participating in Wave 3 to 11
Weights for targets and parents adjusted for wave-specific nonresponse, standardized		
w_tp1	3550	Cross-sectional weight for joint participation in Wave 1
w_tp2	3307	Cross-sectional weight for joint participation in Wave 2
w_tp3	4248	Cross-sectional weight for joint participation in Wave 3
w_tp4	2621	Cross-sectional weight for joint participation in Wave 4
w_tp6	2776	Cross-sectional weight for joint participation in Wave 6
w_tp1to2	3042	Longitudinal weight for joint participation in Wave 1 to 2
w_tp1to3	2544	Longitudinal weight for joint participation in Wave 1 to 3
w_tp1to4	2107	Longitudinal weight for joint participation in Wave 1 to 4
w_tp1to6	1511	Longitudinal weight for joint participation in Wave 1 to 6 (without 5)
w_tp3to4	3404	Longitudinal weight for joint participation in Wave 3 to 4
w_tp3to6	2298	Longitudinal weight for joint participation in Wave 3 to 6 (without 5)
Weights for targets of the Corona-CAWI (C) adjusted for wave-specific nonresponse, standardized		
w_tC_cal	1021	Calibrated cross-sectional weight for targets participating in C
w_tC	1021	Cross-sectional weight for targets participating in C
w_t1toC	398	Longitudinal weight for targets participating in Wave 1 to C
w_t3toC	574	Longitudinal weight for targets participating in Wave 3 to C

Table 4: Summary statistics for all weights provided.

Label of weight	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.	Missings
w_i	0.943	71.559	94.820	133.864	126.427	17545.312	NA
w_t	0.949	115.382	163.353	349.874	244.940	432548.321	NA
w_t_cal	23.294	80.329	109.836	143.624	165.158	14717.814	3034
w_t3_cal	2.371	46.997	73.654	99.440	112.424	17234.585	263
w_t1	0.032	0.564	0.800	1.000	1.232	3.729	2758
w_t2	0.028	0.492	0.698	1.000	1.144	4.457	2987
w_t3	0.022	0.403	0.639	1.000	1.090	4.869	1206
w_t4	0.021	0.416	0.650	1.000	1.082	4.778	1736
w_t5	0.036	0.312	0.479	1.000	0.878	5.379	2669
w_t6	0.020	0.183	0.298	1.000	0.661	5.709	2852
w_t7	0.011	0.110	0.197	1.000	0.848	5.725	2950
w_t8	0.007	0.072	0.158	1.000	1.064	5.711	3178
w_t9	0.005	0.059	0.160	1.000	1.073	5.702	3447
w_t10	0.004	0.052	0.153	1.000	1.049	5.719	4551
w_t11	0.003	0.043	0.151	1.000	0.931	5.770	5086
w_t1to2	0.033	0.559	0.791	1.000	1.248	3.709	3247
w_t1to3	0.032	0.546	0.782	1.000	1.250	3.734	3803
w_t1to4	0.032	0.510	0.747	1.000	1.229	3.933	4290
w_t1to5	0.165	0.492	0.727	1.000	1.210	3.946	5114
w_t1to6	0.160	0.478	0.717	1.000	1.204	4.016	5398
w_t1to7	0.164	0.461	0.698	1.000	1.180	4.168	5713
w_t1to8	0.151	0.436	0.679	1.000	1.171	4.296	6089
w_t1to9	0.141	0.420	0.664	1.000	1.163	4.377	6370
w_t1to10	0.131	0.403	0.667	1.000	1.161	4.409	6834
w_t1to11	0.127	0.399	0.667	1.000	1.164	4.432	7132
w_t3to4	0.026	0.492	0.766	1.000	1.202	4.116	2029
w_t3to5	0.060	0.492	0.740	1.000	1.191	4.156	3198
w_t3to6	0.058	0.472	0.721	1.000	1.183	4.270	3716
w_t3to7	0.055	0.460	0.703	1.000	1.171	4.374	4290
w_t3to8	0.050	0.425	0.663	1.000	1.154	4.573	4956
w_t3to9	0.048	0.414	0.655	1.000	1.135	4.571	5428
w_t3to10	0.048	0.407	0.648	1.000	1.150	4.598	6132
w_t3to11	0.047	0.406	0.641	1.000	1.153	4.603	6577
w_tp1	0.185	0.630	0.832	1.000	1.217	3.053	4767
w_tp2	0.128	0.413	0.570	1.000	0.946	5.228	5010
w_tp3	0.027	0.320	0.488	1.000	0.808	5.476	4069
w_tp4	0.031	0.089	0.142	1.000	0.298	5.917	5696
w_tp6	0.017	0.116	0.206	1.000	0.560	5.830	5541
w_tp1to2	0.196	0.616	0.823	1.000	1.240	3.070	5275
w_tp1to3	0.221	0.586	0.806	1.000	1.236	3.214	5773
w_tp1to4	0.212	0.561	0.782	1.000	1.255	3.416	6210
w_tp1to6	0.188	0.517	0.752	1.000	1.274	3.672	6806
w_tp3to4	0.025	0.457	0.694	1.000	1.139	4.420	4913
w_tp3to6	0.070	0.425	0.674	1.000	1.108	4.556	6019
w_tC	0.004	0.044	0.158	1.000	0.929	5.779	7296
w_tC_cal	0.001	0.040	0.143	1.000	0.868	21.966	7296
w_t1toC	0.145	0.375	0.655	1.000	1.184	4.578	7296
w_t3toC	0.049	0.354	0.579	1.000	1.209	4.822	7296

References

Hammon, A., & Landrock, U. (2019). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 9)* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/ Datenzentrum/Forschungsdaten/SC3/9-0-0/SC3_9-0-0_W9.pdf

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Schnapp, T. (2020). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 10)* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/ Forschungsdaten/SC3/10-0-0/SC3_10-0-0_W10.pdf

Statistisches Bundesamt. (2011). *Bildung und Kultur - Allgemeinbildende Schulen: Schuljahr 2010/2011* (Fachserie 11 No. Reihe 1). Wiesbaden: Statistisches Bundesamt. Retrieved 15.03.2017, from https://www.destatis.de/DE/Publikationen/Thematisch/ BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100137004 .pdf

Steinhauer, H. W. (2017). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 7)* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/ Datenzentrum/Forschungsdaten/SC3/7-0-1/SC3_7-0-1_W.7pdf

Steinhauer, H. W. (2019). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 8)* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/ Datenzentrum/Forschungsdaten/SC3/8-0-0/SC3_8-0-0_W8.pdf

Steinhauer, H. W., Aßmann, C., Zinn, S., Goßmann, S., & Rässler, S. (2015). Sampling and weighting cohort samples in institutional contexts. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *9*(2), 131-157. doi: 10.1007/s11943-015-0162-0

Steinhauer, H. W., & Zinn, S. (2016a). *Neps technical report for weighting: Weighting the sample of starting cohort 3 of the national educational panel study (waves 1 to 3)* (Working Paper No. 63). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIII.pdf

Steinhauer, H. W., & Zinn, S. (2016b). *Neps technical report for weighting: Weighting the sample of starting cohort 3 of the national educational panel study (waves 1 to 5)* (Technical Report). Bamberg: Leibniz Institute for Educational Trajecto-

ries. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/ Forschungsdaten/SC3/5-0-0/SC3_5-0-0_W.pdf

Würbach, A., Landrock, U., Schnapp, T., Ziesmer, J., & Bergrab, M. (2021). *NEPS Technical Report for Weighting – the Additional Corona Survey of the National Educational PAnel Study (NEPS Survey Paper No.89* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. doi: 10.5157/NEPS:SP89:1.0