Research Data

Samples, Weights and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 7)

Steinhauer, H.W.

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 3 of the National Educational Panel Study (Wave 7)

Hans Walter Steinhauer, Leibniz Institute for Educational Trajectories

Technical Report referring to DOI:10.5157/NEPS:SC3:7.0.1

E-mail address of lead author:

methoden@lifbi.de

1. Prequel

This report complements NEPS Survey Paper No. 63 (Steinhauer & Zinn, 2016a) and gives details on Wave 7 of Starting Cohort 3 (SC3) of the National Educational Panel Study (NEPS). It refers to the Scientific Use File (SUF; DOI:10.5157/NEPS:SC3:7.0.1). 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 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 willing 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, because they switch to another school, or because the school decides to refuse further cooperation. 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)", giving the students status by the end of the wave. Finally, the last column gives 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 and Steinhauer and Zinn (2016b) for waves 1 to 5 respectively.

Table 1: Panel progress of SC3 by wave.

Wave (Time)	Study number	Panel Cohort				Status at the end of the wave			
		Sample	Total size	Not used	Used sample	Participants	Temporary dropout	Final dropout (in wave)	Final dropout (after wave)
1 (2010/2011)	A28, A56, A63	**Main**	6,112	0	6,112	5,778	334	0	13
2 (2011/2012)	A29, A57	**Main**	6,099	0	6,099	5,537	561	1	8
3 (2012/2013)	A30, A30A, A58	**All**	8,295	0	8,295	7,277	989	29	10
		Main	6,090	0	6,090	5,131	930	29	10
		Refr.	2,205	0	2,205	2,146	59	0	0
4 (2013/2014)	A31, A59	**All**	8,256	0	8,256	6,718	1,505	33	580
		Main	6,051	0	6,051	4,783	1,249	19	[a]580
		Refr.	2,205	0	2,205	1,935	256	14	0
5 (2014/2015)	A94	**All**	7,643	0	7,643	5,778	1,625	240	0
		Main	5,452	0	5,452	4,001	1,273	178	0
		Refr.	2,191	0	2,191	1,777	352	62	0
6 (Spring 2015)	A98	**All**	7,403	0	7,403	5,586	1,739	78	0
		Main	5,274	0	5,274	3,920	1,292	62	0
		Refr.	2,129	0	2,129	1,666	447	16	0
7 (2015/2016)	A99, B106	**All**	7,319	241	7,078	5,491	1,542	45	0
		Main	5,207	151	5,056	3,924	1,104	28	0
		Refr.	2,112	90	2,022	1,567	438	17	0

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

2. Changes compared to previous version

Compared to version 6.0.0 of the SC3 SUF the weights for children and parents jointly partic-ipating are provided again. To analyze the data in the bivariate probit setting the R package GJRM has been used (Marra & Radice, 2017). For results see Section 4. together with Table 3 and Table 4. The variables giving the number of classes in Grade 8 (tx80113_R) and the number of students in Grade 8 (tx80114_R) have been corrected.

3. Participation in Wave 7

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 probit link function is used, see Steinhauer and Zinn (2016a) and Steinhauer and Zinn (2016b) for details. The coefficients for the estimated models are displayed in Table 2. By Wave 7 the panel cohort has reduced to 7,232 students, see Table 1. Like in Starting Cohort 4 of the NEPS students leave their schools and thus are surveyed individually. These students in in-dividual retracking are less likely to participate compared to those who are surveyed in schools. Furthermore having participated in previous waves significantly influences the participation de-cision in Wave 7 positively for both groups. Students being part of the refreshment sample in states other than Berlin and Brandenburg have a lower propensity to participate compared to students of the refreshment sample from Berlin and Brandenburg. Students from the main sample have a higher propensity to participate, when their school was sampled form the set of schools educating students in grade 5 but none in grade 7 (compared to schools educating stu-dents in both grades). Finally, students in the refreshment sample have a lower participation propensity when having a native language other than German (compared to German).

4. Participation of Students and Parents

The results of the models estimating the individual participation propensities for students and parents in the main sample of SC3 are displayed in Table 3. In Wave 4 and Wave 6 the previous waves participation status of students and parents has a positive effect on the participation propensity of students and parents. A negative effect on the probability to participate is found for students and parents if the student is in the field of individual retracking, compared to stu-dents surveyed in a school. In Wave 4 students in special-needs schools have a lower participa-tion propensity compared to students sampled in the explicit stratum where schools educate students in Grade 5 and Grade 7. Because special-needs students are dismissed from the panel after Wave 4 they do not occur in Wave 6. In Wave 6 female students are more likely to par-ticipate compared to male students. Finally, the residual correlation in the bivariate probit is weakly positive.

Results of the models estimating the individual participation propensities for students and par-ents in the refreshment sample of SC3 are displayed in Table 4. Also in the refreshment sam-ple the students and parents previous waves participation status has a positive effect on the propensity of students and parents participating in an actual wave. The effect of being in the field of individual retracking on the participation decision in Wave 6, compared to being sur-veyed in the school, is negative. In Wave 4 speaking another native language than German low-ers the participation probability of parents, compared to having German as native language. In Wave 6 there is a negative effect on the participation decision of having been sampled in other

Table 2: Models estimating the individual participation propensities for students in Wave 7 of
SC3 used to derive adjustment factors for adjusted wave-specific cross-sectional and
longitudinal weights.

	Wave 7	
	Main sample	Refreshment sample
(Intercept)	−0.222*	0.151
	(0.101)	(0.144)
Explicit stratum	0.251***	
Grade 5 but not Grade 7	(0.075)	
Explicit stratum		−0.362***
Special-needs schools		(0.089)
Gender	−0.108*	
female	(0.045)	
Student participated in	0.222**	
Wave 2	(0.082)	
Student participated in	0.259***	
Wave 3	(0.071)	
Student participated in	0.264***	0.550***
Wave 4	(0.068)	(0.101)
Student participated in	0.426***	0.525***
Wave 5	(0.065)	(0.093)
Student participated in	0.586***	0.326***
Wave 6	(0.061)	(0.085)
Student in individual retracking	−0.683***	−0.689***
yes	(0.052)	(0.070)
Native language		−0.187*
Other than German		(0.088)
Native language		0.615
Missing		(0.379)
Number of students	4,957	2,106

Notes: Reference categories are: Explicit stratum (SC3: Grade 5 and Grade 7), Age group (older half), Native language (German), Gender (male), 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, 2015) was used.

***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parentheses.

states than Berlin or Brandenburg (remaining states). Finally, the residual correlation in the bivariate probit is weakly positive.

Table 3: Models estimating the individual participation propensities for students and parents in the main sample (K5) of SC3 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 6	
	Student	Parent	Student	Parent	Student	Parent	Student	Parent	Student	Parent
(Intercept)	1.851***	0.542***	1.482***	−1.208***	0.660***	−2.147***	−0.073	−2.322***	−0.414***	−2.567***
	(0.037)	(0.026)	(0.048)	(0.044)	(0.078)	(0.104)	(0.088)	(0.098)	(0.107)	(0.104)
Explicit stratum Grade 5 but not Grade 7	0.023	0.032	−0.065	−0.023	0.035	0.110	0.202**	0.003	0.019	0.265**
	(0.121)	(0.066)	(0.093)	(0.083)	(0.076)	(0.094)	(0.077)	(0.091)	(0.075)	(0.084)
Explicit stratum Special-needs schools	0.484**	−0.463***	−0.394***	−0.514***	−0.313***	−0.385***	−0.322***	−0.344***	/[a]	/[a]
	(0.149)	(0.056)	(0.073)	(0.072)	(0.073)	(0.079)	(0.073)	(0.083)	/[a]	/[a]
Gender female									0.103*	
									(0.046)	
Native language Other than German	−0.197*	−0.446***	0.112	−0.162*	−0.070	−0.257***				
	(0.095)	(0.054)	(0.086)	(0.072)	(0.074)	(0.075)				
Native language Missing	−3.977***	−0.795***	−0.629***	−0.187	−0.534***	−0.304				
	(0.258)	(0.111)	(0.132)	(0.151)	(0.135)	(0.170)				
Age group younger half		0.132***							0.107*	0.109*
		(0.035)							(0.047)	(0.045)
Student participated in Wave 2					0.844***	0.423***	0.460***		0.269**	
					(0.069)	(0.088)	(0.077)		(0.087)	
Student participated in Wave 3							1.148***	0.369***	0.554***	0.181*
							(0.061)	(0.075)	(0.072)	(0.082)
Student participated in Wave 4									0.896***	0.412***
									(0.063)	(0.070)
Student in individual retracking yes, in Wave 2			−1.594***	−0.510***						
			(0.068)	(0.081)						
Student in individual retracking yes, in Wave 3					−1.324***	−0.334***				
					(0.053)	(0.063)				
Student in individual retracking yes, in Wave 4							−1.327***	−0.271***		
							(0.050)	(0.058)		
Student in individual retracking yes, in Wave 6									−1.111***	
									(0.050)	
Parent participated in Wave 1			0.268***	2.537***	−0.156*	1.209***		0.642***		0.518***
			(0.053)	(0.049)	(0.072)	(0.069)		(0.079)		(0.086)
Parent participated in Wave 2					0.445***	1.873***		0.960***		0.471***
					(0.070)	(0.061)		(0.075)		(0.086)
Parent participated in Wave 3							0.305***	1.657***		0.537***
							(0.047)	(0.061)		(0.075)
Parent participated in Wave 4									0.265***	1.275***
									(0.048)	(0.064)
Correlation	0.069		0.314***		0.244***		0.255***		0.331***	
	(0.039)		(0.033)		(0.031)		(0.041)		(0.032)	
Number of students and parents	5,870		5,857		5,849		5,814		5,048	

Notes: [a] Students with special educational needs have been dismissed from the panel after Wave 4. Reference categories are: Explicit stratum (SC3: Grade 5 and Grade 7), Age group (older half), Native language (German), Gender (male), Student / Parent 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 `gjrm` function with a probit link provided in R (Marra & Radice, 2017) was used.

***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parentheses.

5. Summary of Weights

Various kinds of weights for students together with design information are provided by NEPS. Table 5 summarizes the design information and the different weights provided by SUF release version DOI:10.5157/NEPS:SC3:7.0.1. 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. The implicit stratification variables are "school type" (`stratum_imp1`), "Federal States" (`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`)

---

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

Table 4: Models estimating the individual participation propensities for students and parents in the refreshment sample (K7) of SC3 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights

	Wave 3		Wave 4		Wave 6	
	Student	Parent	Student	Parent	Student	Parent
(Intercept)	1.834***	0.196**	1.061***	−1.444***	0.721***	−2.452***
	(0.069)	(0.067)	(0.095)	(0.096)	(0.123)	(0.151)
Explicit stratum Remaining states		−0.214**	−0.160	−0.232**	−0.278**	
		(0.067)	(0.090)	(0.089)	(0.089)	
Gender female			0.198**			
			(0.069)			
Native language Other than German		−0.416***		−0.319**		
		(0.077)		(0.110)		
Native language Missing		−0.579		−0.196		
		(0.426)		(0.634)		
Age group younger half	0.248*	0.270***				
	(0.117)	(0.055)				
Student participated in Wave 4					0.625***	0.380**
					(0.097)	(0.134)
Student in individual retracking yes, in Wave 6					−1.007***	
					(0.071)	
Parent participated in Wave 3			0.290***	2.701***		1.434***
			(0.070)	(0.082)		(0.119)
Parent participated in Wave 4					0.174**	1.460***
					(0.067)	(0.104)
Correlation	0.182*		0.399***		0.129**	
	(0.085)		(0.037)		(0.048)	
Number of students and parents	2,205		2,205		2,129	

Notes: Reference categories are: Explicit stratum (SC3: Berlin and Brandenburg), Gender (male), Native language (German), Age group (older half), Student / Parent 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 `gjrm` function with a probit link provided in R (Marra & Radice, 2017) was used.

***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parentheses.

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. 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.

---

[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 5: Variables included in the weighting data of SC3 SUF version 7.0.1

Variable	Applies to	Content
Identifier		
ID_t	8,317	Identifier for target person (students)
ID_i	8,317	Identifier for the institution (377 schools)
Design information		
tstud_st	8,317	Study number the target person was first surveyed in (A28, A56, A63, A30A)
sample	8,317	Part of the sample the target person belongs to
stratum_exp	8,317	Explicit sampling stratum referring to the school
stratum_imp1	8,317	Implicit sampling stratum (school type according to sampling frame)
stratum_imp2_R	8,317	Implicit sampling stratum (Federal State the school is located in according to sampling frame)
stratum_imp3_R	8,317	Implicit sampling stratum (regional classification according to sampling frame)
stratum_imp4_R	8,317	Implicit sampling stratum (funding according to sampling frame)
tx80113_R	8,317	Total number of students in grade 8 as reported by official statistics
tx80114_R	8,317	Total number of classes in grade 8 as reported by official statistics
Design weights adjusted for initial nonresponse		
w_i	8,317	Weight for institution
w_t	8,317	Weight for target
w_t_cal	5,283	Weight for target, calibrated
w_t3_cal	8,054	Weight for targets in Wave 3, calibrated
Weights adjusted for wave-specific nonresponse, standardized		
w_t1	5,559	Cross-sectional weight for targets participating in Wave 1
w_t2	5,331	Cross-sectional weight for targets participating in Wave 2
w_t3	7,114	Cross-sectional weight for targets participating in Wave 3
w_t4	6,581	Cross-sectional weight for targets participating in Wave 4
w_t5	5,648	Cross-sectional weight for targets participating in Wave 5
w_t6	5,465	Cross-sectional weight for targets participating in Wave 6
w_t7	5,352	Cross-sectional weight for targets participating in Wave 7
w_t12	5,071	Longitudinal weight for targets participating in Wave 1 and 2
w_t123	4,516	Longitudinal weight for targets participating in Wave 1, 2, and 3
w_t1234	4,029	Longitudinal weight for targets participating in Wave 1, 2, 3, and 4
w_t12345	3,203	Longitudinal weight for targets participating in Wave 1, 2, 3, 4, and 5

Table 5: Variables included in the weighting data of SC3 SUF version 7.0.1

Variable	Applies to	Content
w_t123456	2,919	Longitudinal weight for targets participating in Wave 1, 2, 3, 4, 5, and 6
w_t1234567	2,597	Longitudinal weight for targets participating in Wave 1, 2, 3, 4, 5, 6, and 7
w_t34	6,291	Longitudinal weight for targets participating in Wave 3 and 4
w_t345	5,119	Longitudinal weight for targets participating in Wave 3, 4, and 5
w_t3456	4,601	Longitudinal weight for targets participating in Wave 3, 4, 5, and 6
w_t34567	4,017	Longitudinal weight for targets participating in Wave 3, 4, 5, 6, and 7
Weights for targets and parents adjusted for wave-specific nonresponse, standardized		
w_tp1	3,550	Cross-sectional weight for joint participation in Wave 1
w_tp2	3,307	Cross-sectional weight for joint participation in Wave 2
w_tp3	4,249	Cross-sectional weight for joint participation in Wave 3
w_tp4	2,621	Cross-sectional weight for joint participation in Wave 4
w_tp6	2,776	Cross-sectional weight for joint participation in Wave 6
w_tp12	3,042	Longitudinal weight for joint participation in Wave 1 and 2
w_tp123	2,544	Longitudinal weight for joint participation in Wave 1, 2, and 3
w_tp1234	2,107	Longitudinal weight for joint participation in Wave 1, 2, 3, and 4
w_tp12346	1,511	Longitudinal weight for joint participation in Wave 1, 2, 3, and 6
w_tp34	3,404	Longitudinal weight for joint participation in Wave 3 and 4
w_tp346	2,298	Longitudinal weight for joint participation in Wave 3, 4, and 6

References

Marra, G., & Radice, R. (2017). Gjrm: Generalised joint regression modelling [Computer software manual]. (R package version 0.1-4)

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

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