NEPS
National Educational Panel Study

Hans Walter Steinhauer

Samples, Weights and Nonresponse

NEPS Starting Cohort 4 — Grade 9

School and Vocational Training — Educational Pathways of Students in Grade 9 and Higher

Wave 10

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 10)

Hans Walter Steinhauer, Leibniz Institute for Educational Trajectories

Technical Report referring to

E-mail address of lead author:

statistik@lifbi.de

1 Prequel

This report complements NEPS Survey Paper No. 2 (Steinhauer & Zinn, 2016) and Steinhauer and Zinn (2018) and gives details on Wave 10 of Starting Cohort 4 (SC4) of the National Educational Panel Study (NEPS). It refers to the Scientific Use File (SUF; DOI:10.5157/NEPS:SC4:10.0.0). SC4 focuses on the educational pathways of Grade 9 students initially educated in different types of regular schools and special-need schools. The students willing to participate in the panel study (i.e., the panel members) are followed up over time. In a typical pathway, students in Germany decide after Grade 10 to enter either the academic track or the vocational track, see Figure 1. The students entering the academic track usually remain within their school context. In contrast, students entering the vocational education leave school for a vocational training. Most students enter the vocational track after Grade 10, but some students enter the vocational track earlier or later in their educational career. Figure 1 illustrates this transition pattern.
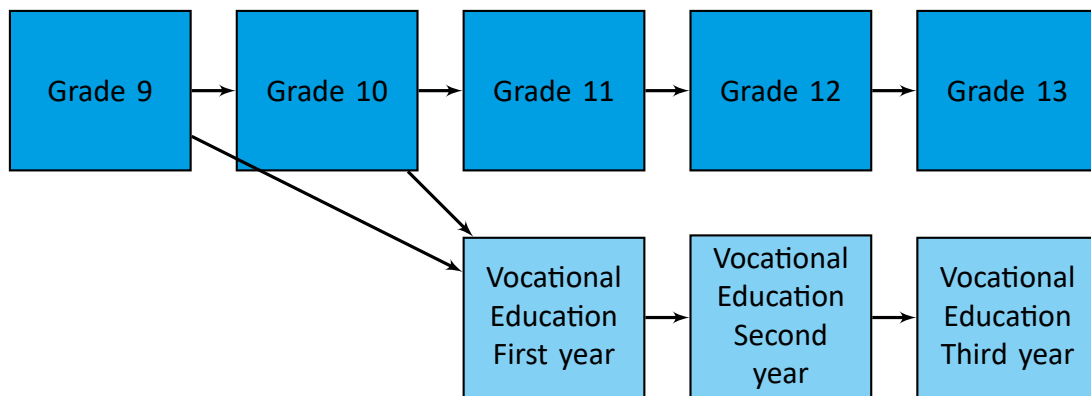

Figure 1: Ideal pathways through upper secondary and vocational education.

Table 1 complements the illustration with the number of students in the two different tracks of education in SC4. Here, the vocational track (VOC) not only consists of students in vocational training but also includes students leaving school and entering the transition system. The numbers for students in the academic track (ACA) report students who stay in their schools together with those who left school but remain in academic education. The table gives details on the size of panel cohort over time. The column "Not used" gives the number of students who have not been surveyed either by design, for example in Wave 4 or Wave 6, or who could not be surveyed because of insufficient contact details. For the latter group, the number increases over time, but most cases could be retracked by Wave 9 and are surveyed again. The "Used sample" is then 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.

This report builds upon Steinhauer, Aßmann, Zinn, Goßmann, and Rässler (2015) giving details on the sample design together with the nonresponse adjustment of design weights and Steinhauer and Zinn (2016) giving details on wave-specific nonresponse adjustments for waves 1 to 6.

Table 1: Panelprogress of SC 4 by wave.

Wave (Time)	Study number	Sample	Panel Cohort			Status at the end of the wave			
			Total size	Not used	Used sample	Participants	Temporary dropout	Final dropout (in wave)	Final dropout (after wave)
1 (Fall 2010)	A46, A60, A67, A83, A86	**All**	16,425	-	16,425	16,106	319	0	0
2 (Summer 2011)	A47, A61, A68, A84, A87	**All**	16,425	-	16,425	15,215	1,210	0	61
3 (2011/2012)	A48, A62, A69, A85, A88, B37	**All**	16,364	8	16,356	14,011	2,234	111	0
		ACA		-	13,815	11,951	1,842	22	0
		VOC		8	2,541	2,060	392	89	0
4 (Spring 2012)	B38	**All**	16,253	14,440	-	-	-	7	5
		ACA		13,793	-	-	-	-	3
		VOC		647	1,813	1,351	455	7	2
5 (2012/2013)	A49, B39	**All**	16,241	132	16,109	12,982	2,644	483	4
		ACA		-	6,305	5,768	522	15	1
		VOC		132	9,804	7,214	2,122	468	3
6 (Spring 2013)	B40	**All**	15,754	9,635	-	-	-	60	2
		ACA		6,289	-	-	-	-	1
		VOC		3,346	6,119	5,392	667	60	1
7 (2013/2014)	A50, B41	**All**	15,692	185	15,507	11,830	3,121	556	37
		ACA		-	5,333	4,735	593	5	22
		VOC		185	10,174	7,094	2,529	551	15
8 (2014/2015)	A96, B93	**All**	15,099	[a]1,318	13,781	9,871	3,400	510	[b]1,551
		ACA		-	688	610	75	3	16
		VOC		1,318	13,093	9,261	3,325	507	1,535
9 (2015/2016)	B109	**All**	13,038	0	13,038	9,044	3,262	732	[c]1,264
10 (2016/2017)	B110	**All**	11,042	0	11,042	7,986	2,382	674	n.a.

Notes: "-" does not apply. [a]: including 1,067 students from special-need schools not surveyed in Wave 8. [b] ([c]): including 1,395 (1,246) students declared as final drop-outs because of not having participated for a period of two years. "n.a." information not yet available.

2 Changes compared to previous version

The number of participants in Wave 1 (Wave 7) has been corrected and increased by 477 (1) students. Thus, the number of weighted cases for the corresponding cross-sectional and longitudinal weights have changed. Weights for Wave 10 have been appended.

3 Nonresponse Adjustments in Wave 10

By Wave 10 the panel cohort has reduced from initially 16,425 students in Wave 1 to 11,042 students. The large number of students dropping out between Wave 9 and Wave 10 originates from 1,246 students, who were declared as final drop-outs because of not having participated for a period of two years. As in earlier waves, a smaller amount of students explicitly refuse further participation in the wave., see Table 1. Being part of the younger half of the age group positively influences the participation decision in Wave 10, compared to belonging to the older half, see Table 2. Also speaking German as primary language, compared to another language than German, increases the probability for participating in Wave 10. Most of the previous wave's participation positively influences the participation decision in Wave 10. This holds for Waves 5, 7, 8, and 9, where the entire cohort was surveyed. However, this finding does not for Waves 4 and 6. Here, the participation decision negatively influences the participation in Wave 10. This is because in Wave 4 and 6 only students in the vocational track were surveyed. Students in the vocational track have been found to be more likely to drop out of the panel in later waves.

Table 2: Models estimating the individual participation propensities for students in Wave 10 of SC4 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights.

	Wave 10
(Intercept)	-1.117^{***}
	(0.075)
Age group younger half	0.100^{***}
	(0.027)
Primary language: German	0.241^{***}
	(0.045)
Primary language: missing	0.091
	(0.147)
Student participated in Wave 4	-0.168^{***}
	(0.046)
Student participated in Wave 5	0.211^{***}
	(0.050)
Student participated in Wave 6	-0.095^{**}
	(0.029)
Student participated in Wave 7	0.234^{***}
	(0.044)
Student participated in Wave 8	0.372^{***}
	(0.037)
Student participated in Wave 9	0.999^{***}
	(0.032)
Number of students	11042

Notes: Reference categories are: Primary language (other than German), Student participated in Wave t (no), Age group (older half). To model individual participation, the `glm` function with a probit link provided in R (R Core Team, 2018) was used.

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

4 Summary of Weights

Various kinds of weights for students are provided together with design information. Table 3 summarizes the design information given and the different weights provided; compare SUF release version DOI:10.5157/NEPS:SC4:10.0.0. 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 "Federal States" ($stratum\_imp1$), "regional classification" ($stratum\_imp2$) and "funding institution" ($stratum\_imp3$).[2] In this release version additional information has been added to the design data, namely the number of students ($h227102\_d$) and classes ($h229021\_d$) in Grade 9 in school year 2010/2011 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.[3] For all participants in a particular wave, cross-sectional weights are provided. These apply to all participants in that wave.

From this version on cross-sectional and longitudinal weights are based on the calibrated weight w_t_cal. Thus, corresponding cross-sectional and longitudinal weights are also calibrated to the population of students in Grade 9 in school year 2010/2011.

For the latest release version the cross-sectional weight w_t10 have been added. Besides that, the longitudinal weight $w\_t123578910$ has been added.

Table 3: Variables included in the weighting data of SC4 SUF version 10.0.0

Variable	Applies to	Content
Identifier		
ID_t	16,425	Identifier for target person (students)
ID_i	16,425	Identifier for the institution (648 schools)
Design information		
tstud_st	16,425	Study number the target person was first surveyed in (A46, A60, A67, A83, A86)
stratum_exp	16,425	Explicit sampling stratum referring to the school (school type according to sampling frame)
stratum_imp1	16,425	Implicit sampling stratum (Federal State the school is located in according to sampling frame)
stratum_imp2	16,425	Implicit sampling stratum (regional classification according to sampling frame)
stratum_imp3	16,425	Implicit sampling stratum (funding according to sampling frame)
h227102_d	16,425	Number of students in grade 9 as reported by official statistics

---

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

[2] In the SUF, these design variables are named differently, because of an error in data preparation. Here, variables $stratum\_exp$, $stratum\_imp1$, $stratum\_imp2$, and $stratum\_imp3$ are named $stratum\_imp1$, $stratum\_imp2$, $stratum\_imp3$, and $stratum\_imp4$.

[3] 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 SC4 SUF version 10.0.0

Variable	Applies to	Content
h229021_d	16,425	Number of classes in grade 9 as reported by official statistics
Design weights adjusted for initial nonresponse		
w_i	16,425	Weight for institution
w_t	16,425	Weight for target
w_t_cal	16,425	Weight for target, calibrated
Weights adjusted for wave-specific nonresponse, standardized		
w_t1	16,106	Cross-sectional weight for targets participating in Wave 1
w_t2	15,215	Cross-sectional weight for targets participating in Wave 2
w_t3	14,011	Cross-sectional weight for targets participating in Wave 3
w_t4	1,351	Cross-sectional weight for targets participating in Wave 4
w_t5	12,982	Cross-sectional weight for targets participating in Wave 5
w_t6	5,392	Cross-sectional weight for targets participating in Wave 6
w_t7	11,829	Cross-sectional weight for targets participating in Wave 7
w_t8	9,871	Cross-sectional weight for targets participating in Wave 8
w_t9	9,044	Cross-sectional weight for targets participating in Wave 9
w_t10	7,986	Cross-sectional weight for targets participating in Wave 10
w_t12	15,056	Longitudinal weight for targets participating in Wave 1 and 2
w_t123	13,188	Longitudinal weight for targets participating in Wave 1, 2, and 3
w_t1234	1,226	Longitudinal weight for targets participating in Wave 1, 2, 3, and 4
w_t1235	11,037	Longitudinal weight for targets participating in Wave 1, 2, 3, and 5
w_t12356	4,677	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, and 6
w_t12357	9,463	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, and 7
w_t123578	7,425	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, and 8
w_t1235789	5,962	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, 8, and 9
w_t123578910	4,896	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, 8, 9, and 10

of Bamberg in cooperation with a nationwide network.

References

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

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. (2016). *Neps technical report for weighting: Weighting the sample of starting cohort 4 of the national educational panel study (wave 1 to 6)* (NEPS Survey Paper No. 2). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIII.pdf

Steinhauer, H. W., & Zinn, S. (2018). *Neps technical report for weighting: Weighting the sample of starting cohort 4 of the national educational panel study (wave 7 to 9)* (Tech. Rep.). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.