

# NEPS Technical Report for Weighting: Weighting the Sample of Starting Cohort 4 of the National Educational Panel Study (Waves 7 to 9)

*Hans Walter Steinhauer and Sabine Zinn*

# **Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 7 to 9)**

*Hans Walter Steinhauer, Leibniz Institute for Educational Trajectories  
Sabine Zinn, Leibniz Institute for Educational Trajectories*

Technical Report referring to [DOI:10.5157/NEPS:SC4:9.0.0](https://doi.org/10.5157/NEPS:SC4:9.0.0)

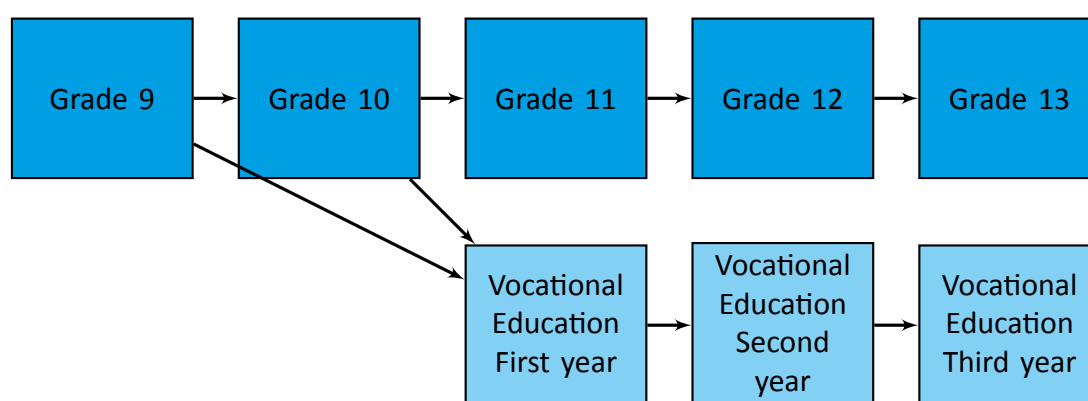
**E-mail address of lead author:**

methoden@lifbi.de

---

## 1. Prequel

This report complements NEPS Survey Paper No. 2 (Steinhauer & Zinn, 2016) and gives details on Wave 7, Wave 8 and Wave 9 of Starting Cohort 4 (SC4) of the National Educational Panel Study (NEPS). It refers to the Scientific Use File (SUF; DOI:10.5157/NEPS:SC4:9.0.0). 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: Panel progress 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	<b>All</b>	16,425	-	16,425	15,629	796	0	0
2 (Summer 2011)	A47, A61, A68, A84, A87	<b>All</b>	16,425	-	16,425	15,215	1,210	0	61
3 (2011/2012)	A48, A62, A69, A85, A88, B37	<b>All</b>	16,364	8	16,356	14,011	2,234	111	0
4 (Spring 2012)	B38	ACA	16,253	14,440	-	11,951	1,842	22	0
		VOC		13,793	8	2,541	392	89	0
5 (2012/2013)	A49, B39	<b>All</b>	16,241	132	16,109	12,982	2,644	483	4
		ACA		-	6,305	5,768	522	15	1
6 (Spring 2013)	B40	VOC		132	9,804	7,214	2,122	468	3
		<b>All</b>	15,754	9,635	-	-	-	60	2
7 (2013/2014)	A50, B41	ACA		6,289	-	-	-	-	1
		VOC		3,346	6,119	5,392	667	60	1
8 (2014/2015)	A96, B93	<b>All</b>	15,692	185	15,507	11,829	3,122	556	37
		ACA		-	5,333	4,735	593	5	22
9 (2015/2016)	B109	VOC		185	10,174	7,094	2,529	551	15
		<b>All</b>	15,099	<sup>a</sup> 1,318	13,781	9,871	3,400	510	<sup>b</sup> 1,551
		ACA		-	688	610	75	3	16
		VOC		1,318	13,093	9,261	3,325	507	1,535
		<b>All</b>	13,038	0	13,038	9,044	3,262	732	n.a.

Notes: "-" does not apply. "n.a." not available. <sup>a</sup>: including 1,067 students from special-need schools not surveyed in Wave 8. <sup>b</sup>: including 1,395 students declared as final drop-outs because of not having participated for a period of two years.

## **2. Wave and Group Specific Nonresponse Adjustments**

### **2.1 Wave 7**

By Wave 7 the panel cohort has reduced from initially 16,425 students in Wave 1 to 15,692 students. As in earlier waves, most of the students refusing further participation in the panel stem from the vocational track, see Table 1.

As before, in Wave 7 the academic and vocational track are separated, see Table 2. In the academic track being female as well as having participated in earlier waves and being in the field of individual re-tracking have a positive effect on the participation decision of students. This is because most of the students in individual re-tracking participate, whereas a larger proportion of students in schools temporary drop out in Wave 7. In the vocational track, only the previous participation in Wave 3, Wave 5 and Wave 6 have a positive effect on the participation propensities in Wave 7.

### **2.2 Wave 8**

In Wave 8, 1,352 students were not included in the survey, see Table 1. Of these 1,067 are students from special-need schools who were not surveyed by design in Wave 8. The remaining cases could not be surveyed because of insufficient contact details. By Wave 8 most students have left the academic track. After Wave 8 there are 1,395 students finally dropping out of the sample because they did not participate over a period of two years. In the NEPS such students are declared as final drop-outs, see Aßmann and Sixt (2013).

Although there are still 689 students in the academic track there is no separation of the groups into distinct models for nonresponse analysis. This is because of the small number of cases in the academic track, see Table 1, that does not allow for a meaningful model. Instead the nonresponse analysis for students in the vocational and academic track uses only one model in Wave 8 and includes a corresponding variable indicating the educational track of a student, see Table 2. In Wave 8 there is a positive impact on the participation propensity by being part of the younger half of the age group as well as by having participated in previous waves. Negative effects on the participation decision in Wave 8 are found for having a migration background and for missing information on the migration background as well as for being part of the vocational track.

### **2.3 Wave 9**

Of those students who have not been surveyed in Wave 8 the students not surveyed by design have been surveyed again. Of the remaining students the majority could be surveyed again in Wave 9 because of enormous tracking efforts, compare Table 1. By the beginning of Wave 9 there is no separation between academic and vocational track anymore, because all students have left the academic track. Most final dropouts within the wave stem from students who have not been surveyed by design in Wave 8, that is, special-need students.

As before, in Wave 8, being part of the younger half of the age group together with having participated in previous waves has a positive influence on the participation decision in Wave 9, see Table 2. Alike, having a migration background and missing information on the migration background has a negative effect on the participation propensity in Wave 9.

*Table 2: Models estimating the individual participation propensities for students in Wave 7, Wave 8, and Wave 9 of SC4 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights*

	Wave 7		Wave 8	Wave 9
	Academic education	Vocational education		
(Intercept)	-0.242 (0.128)	-0.813*** (0.040)	-0.474*** (0.081)	-0.829*** (0.054)
Gender	0.149** (0.047)			
female				
Student participated in Wave 2	0.381*** (0.097)			
Student participated in Wave 3	0.245** (0.083)	0.161*** (0.039)	0.138*** (0.038)	0.130*** (0.038)
Student participated in Wave 5	0.846*** (0.067)	1.019*** (0.039)	0.717*** (0.037)	0.326*** (0.042)
Student participated in Wave 6		0.984*** (0.035)	0.389*** (0.029)	0.087** (0.026)
Student participated in Wave 7			1.018*** (0.031)	0.290*** (0.037)
Student participated in Wave 8				0.893*** (0.027)
Individual re-tracking yes	0.418*** (0.073)			
Age group			0.135*** (0.025)	0.080** (0.025)
younger half				
Migration background missing			-0.319** (0.109)	-0.046 (0.107)
Migration background yes			-0.139*** (0.027)	-0.120*** (0.027)
Educational track in Wave 7 vocational			-0.647*** (0.068)	
Number of students	5,333	10,359	14,032	13,038

Notes: Reference categories are: Student in individual re-tracking (no), Student participated in Wave t (no), Gender (male). 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.

### 3. 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:SC4:9.0.0](https://doi.org/10.5157/NEPS:SC4:9.0.0). Besides individual/target (ID\_t) and institutional (ID\_i) identifiers, design information for the entire cohort is made available.<sup>1</sup> 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`).<sup>2</sup> 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

<sup>1</sup>Due to data protection, this information is not available in the download version of the SUF.

<sup>2</sup>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`.

cohort.<sup>3</sup> 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  $w\_t\_ca1$ . 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 weights  $w\_t7$ ,  $w\_t8$ , and  $w\_t9$  have been added. Besides the longitudinal weights  $w\_t12357$ ,  $w\_t123578$ , and  $w\_t1235789$  have been added.

Table 3: Variables included in the weighting data of SC4 SUF version 9.0.0

Variable	Applies to	Content
<i>Identifier</i>		
ID_t	16,425	Identifier for target person (students)
ID_i	16,425	Identifier for the institution (648 schools)
<i>Design information</i>		
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
h229021_d	16,425	Number of classes in grade 9 as reported by official statistics
<i>Design weights adjusted for initial nonresponse</i>		
w_i	16,425	Weight for institution
w_t	16,425	Weight for target
w_t_cal	16,425	Weight for target, calibrated
<i>Weights adjusted for wave-specific nonresponse, standardized</i>		
w_t1	15,629	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

<sup>3</sup>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 9.0.0

Variable	Applies to	Content
w_t9	9,044	Cross-sectional weight for targets participating in Wave 9
w_t12	14,579	Longitudinal weight for targets participating in Wave 1 and 2
w_t123	12,784	Longitudinal weight for targets participating in Wave 1, 2, and 3
w_t1234	1,169	Longitudinal weight for targets participating in Wave 1, 2, 3, and 4
w_t1235	10,701	Longitudinal weight for targets participating in Wave 1, 2, 3, and 5
w_t12356	4,534	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, and 6
w_t12357	9,188	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, and 7
w_t123578	7,202	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, and 8
w_t1235789	5,801	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, 8, and 9

**Acknowledgements** This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 9, DOI:10.5157/NEPS:SC4:9.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network.

## References

- Aßmann, C., & Sixt, M. (2013). *Umgang mit (temporären) Ausfällen im Nationalen Bildungspanel (NEPS) - Startkohorten-übergreifende Regelung* (Internal Discussion Paper). Bamberg: Leibniz Institute for Educational Trajectories. (Available on request from the corresponding author.)
- Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using Eigen and Eigenpack*. Retrieved 03.06.2014, from <http://CRAN.R-project.org/package=lme4>
- R Core Team. (2015). 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](https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIII.pdf)