

NEPS

National Educational Panel Study

Research Data

Michael Bergrab

Samples, Weights and Nonresponse

NEPS Starting Cohort 4 — Grade 9

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

Wave 11

LifBi

LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES

Copyrighted Material
Leibniz Institute for Educational Trajectories (LifBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Cordula Artelt
Executive Director of Research: Dr. Jutta von Maurice
Executive Director of Administration: N. N.
Bamberg; November 19, 2020

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 11)

Michael Bergrab, Leibniz Institute for Educational Trajectories

Technical Report referring to [DOI:10.5157/NEPS:SC4:11.0.0](https://doi.org/10.5157/NEPS:SC4:11.0.0)

E-mail address of lead author:

statistik@lifbi.de

1. Prequel

This report complements NEPS Survey Paper No. 2 (Steinhauer, Aßmann, Zinn, Goßmann, & Rässler, 2015) and gives details on wave 11 of Starting Cohort 4 (SC 4) of the National Educational Panel Study (NEPS). It refers to the Scientific Use File (SUF; DOI:10.5157/NEPS:SC4:11.0.0). SC4 focuses on the educational pathways of Grade 9 students initially educated in different types of regular schools and special-need school. 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.

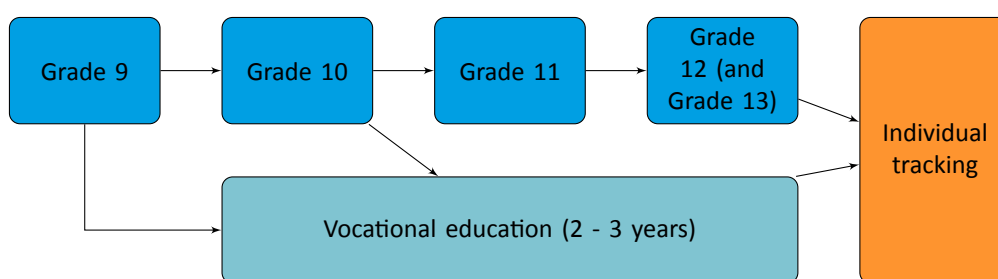


Figure 1: Ideal pathways through upper secondary and vocational education.

Table 1 complements the illustration with the number of students in the two different tracks of education in SC 4. 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 of students in the academic track (ACA) report students who stay in their schools together with those who left school but remain 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 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 including students declared as final drop-outs because of not having participated for a period of two years.

This report builds upon Steinhauer et al. (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 adjustment for Waves 1 to 6, Steinhauer and Zinn (2018) for Waves 7 to 9 and Steinhauer (2019) for Wave 10, respectively.

Table 1: Panel progress of SC 4 by wave.

Wave (Time)	Study number		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, B83	All	16425	0	16425	16106	319	0	0
2 (Summer 2011)	A47, A61, A68, A84, A87	All	16425	0	16425	15215	1210	0	61
3 (2011/2012)	A48, A62, A69, A85, A88, B37, B84	All	16364	8	16356	14011	2234	111	0
		ACA	-	0	13815	11951	1842	22	0
		VOC	-	8	2541	2060	392	89	0
4 (Spring 2012)	B38, B85	All	16253	14440	1813	1351	455	7	5
		ACA	-	13793	-	-	-	-	3
		VOC	-	647	1813	1351	455	7	2
5 (2012/2013)	A49, B39, B86	All	16241	132	16109	12982	2644	483	4
		ACA	-	0	6305	5768	522	15	1
		VOC	-	132	9804	7214	2122	468	3
6 (Spring 2013)	B40, B87	All	15754	9635	6119	5392	667	60	2
		ACA	-	6289	-	-	-	-	1
		VOC	-	3346	6119	5392	667	60	1
7 (2013/2014)	A50, B41, B88	All	15692	185	15507	11830	3121	556	45
		ACA	-	0	5333	4736	592	5	26
		VOC	-	185	10174	7094	2529	551	19
8 (2014/2015)	A96, B93	All	15091	^a 1310	13781	9871	3400	510	^b 1543
		ACA	-	0	688	610	75	3	16
		VOC	-	1310	13093	9261	3325	507	1527
9 (2015/2016)	B109	All	13038	0	13038	9044	3262	732	^c 1264
10 (2016/2017)	B110	All	11042	0	11042	7986	2382	674	^d 795
11 (2017/2018)	B135	All	9573	0	9573	6272	2879	422	^e 1240

Notes: "-" does not apply; 'n.a.': information not yet available; ^a: including 1,067 students from special-need schools not surveyed in Wave 8. ^b: including 1,396 students declared as final drop-outs because of not having participated for a period of two years. ^c: including 1,246 students declared as final drop-outs because of not having participated for a period of two years. ^d: including 780 students declared as final drop-outs because of not having participated for a period of two years. ^e: including 618 students declared as final drop-outs because of not having participated for a period of two years.

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.

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 probit link function is used and adapted with a stepwise selection. By Wave 11 the panel cohort has reduced to 9,573 students, see Table 1. All students left their schools and thus are surveyed individually.

The significant coefficients for the estimated models are displayed in Table 2. We can see that having participated in previous waves significantly influences the participation decision in Wave 11 mostly positive, except participation in Wave 6. Additionally, the younger half of students tends to have a significantly increased participation propensity. A significantly negative effect on the participation decision in Wave 11 is found for respondents with migration background compared to those having German nationality.

4. Summary of Weights

Various kinds of weights for students together with design information are provided. Table 3 summarizes the design information and the different weights provided by SUF release version [DOI:10.5157/NEPS:SC4:11.0.0](https://doi.org/10.5157/NEPS:SC4:11.0.0). 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 "Federal State" (stratum_imp1), "regional classification" (stratum_imp2) and "funding institution" (stratum_imp3).² With release version 10.0.0 additional information has been added to the design data, namely the total 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. With respect to panel progress longitudinal weights are also available. With version 10.0.0 cross-sectional and longitudinal weights are now based on the calibrated weight w_{t_ca1} . Thus, corresponding cross-sectional and longitudinal weights are also calibrated to

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

²In 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.

³The 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 2: Models estimating the individual participation propensity for students in Wave 11 of SC 4 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights.

	Wave 11
(Intercept)	-1.126*** (0.076)
AgeGroup	0.145***
younger half	(0.028)
Migration background	-0.197
missing	(0.127)
Migration background	-0.195***
yes	(0.031)
Student participated in	0.178**
Wave 5	(0.055)
Student participated in	-0.094**
Wave 6	(0.029)
Student participated in	0.202***
Wave 8	(0.039)
Student participated in	0.581***
Wave 9	(0.044)
Student participated in	0.846***
Wave 10	(0.036)
Number of students	9,573

Notes: Reference categories are: Age group (older half), migration background (not missing), migration background (no), students participated in Wave t. 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 significant coefficients are reported.

the population in Grade 9 in school year 2010/2011. The general overview of variables contained in the weighting data set can be found in Table 3. It is accompanied by summarizing statistics of all weights provided, see Table 4.

Table 3: Variables included in the weighting data of SC 4 SUF version 11.0.0

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	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	11,830	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_t11	6,272	Cross-sectional weight for targets participating in Wave 11
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_t12356	4,677	Longitudinal weight for targets participating in Wave 1, 2, 3, 5 and 6

Table 3: Variables included in the weighting data of SC 4 SUF version 11.0.0 (continued)

Variable	Applies to	Content
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
w_t12357891011	3,674	Longitudinal weight for targets participating in Wave 1, 2, 3, 5, 7, 8, 9, 10 and 11

Table 4: Summary statistics for all weights provided.

Label of weight	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.	Missings
w_i	1.720	14.259	18.148	19.381	24.251	100.049	NA
w_t	1.860	32.976	44.287	49.050	56.164	2950.450	NA
w_t_cal	2.443	32.619	44.516	50.437	58.208	2004.553	NA
w_t1	0.052	0.676	0.918	1.000	1.200	2.858	319
w_t2	0.053	0.683	0.906	1.000	1.186	2.905	1210
w_t3	0.038	0.480	0.634	1.000	0.950	4.778	2414
w_t4	0.026	0.235	0.334	1.000	0.636	5.662	15074
w_t5	0.026	0.265	0.430	1.000	0.904	5.392	3443
w_t6	0.030	0.400	0.548	1.000	0.848	5.176	11033
w_t7	0.005	0.026	0.332	1.000	1.163	5.581	4595
w_t8	0.006	0.030	0.291	1.000	1.101	5.634	6554
w_t9	0.004	0.031	0.255	1.000	0.993	5.697	7381
w_t10	0.003	0.028	0.208	1.000	0.976	5.746	8439
w_t11	0.003	0.023	0.184	1.000	0.964	5.768	10153
w_t12	0.054	0.690	0.914	1.000	1.195	2.828	1369
w_t123	0.049	0.618	0.805	1.000	1.122	3.441	3237
w_t1234	0.041	0.355	0.499	1.000	0.773	5.425	15199
w_t1235	0.044	0.418	0.603	1.000	1.109	4.600	5388
w_t12356	0.032	0.402	0.629	1.000	1.019	4.810	11748
w_t12357	0.009	0.036	0.436	1.000	1.268	5.424	6962
w_t123578	0.008	0.036	0.410	1.000	1.231	5.461	9000
w_t1235789	0.009	0.038	0.414	1.000	1.187	5.482	10463
w_t123578910	0.009	0.037	0.399	1.000	1.135	5.521	11529
w_t12357891011	0.009	0.036	0.365	1.000	1.081	5.589	12751

Acknowledgements This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 5, [DOI:10.5157/NEPS:SC4:11.0.0](https://doi.org/10.5157/NEPS:SC4:11.0.0). 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

- R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Steinhauer, H. W. (2019). *NEPS Technical Report for Weighting: Weighting the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 10)* (Technical Report). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/10-0-0/SC4_10-0-0_W.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. (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 (Waves 7 to 9)* (Technical Report). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/9-0-0/SC4_9-0-0_W.pdf