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 15

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

Leibniz Institute for Educational Trajectories (LIfBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Cordula Artelt
Administrative Director: Dr. Stefan Echinger
Bamberg; February 27, 2025

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 15)

Michael Bergrab

Leibniz Institute for Educational Trajectories

Technical Report referring to DOI:10.5157/NEPS:SC4:15.0.0

E-mail address:

statistik@lifbi.de

Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 15)

1 Prequel

This report complements NEPS Survey Paper No. 2 (Steinhauer & Zinn, 2016a) and gives details on wave 15 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:SC4:15.0.0).[1] 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 9 and 10, respectively, 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.

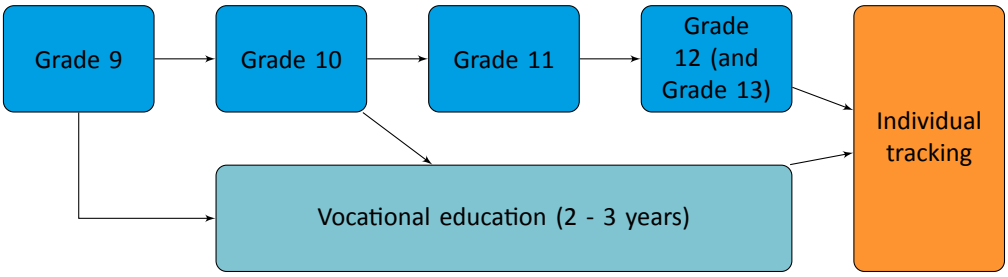

Figure 1: Ideal pathways through upper secondary and vocational education.

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.

[1] For general information on the NEPS, see Blossfeld et al., 2011. More detailed information is available in the documentation section on the homepage.

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 (2016b) giving details on wave-specific nonresponse adjustment for Waves 1 to 6, Steinhauer and Zinn (2018) for Waves 7 to 9, Steinhauer (2019) for Wave 10, Bergrab (2020) for Wave 11, Bergrab (2021) for Wave 12, Bergrab (2023) for Wave 13, and Bergrab (2024) for Wave 14, respectively.

Table 1: Panel progress of SC 4 by wave.

			Panel Cohort			Status at the end of the wave			
Wave (Time)	Study number		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

Table 1: Panel progress of SC 4 by wave.

			Panel Cohort			Status at the end of the wave			
Wave (Time)	Study number		Total size	Not used	Used sample	Participants	Temporary dropout	Final dropout (in wave)	Final dropout (after wave)
12 (2019/2020)	B136	**All**	7911	0	7911	4828	2840	243	[f]1325
13 (2020/2021)	B137	**All**	6343	0	6343	4511	1710	122	[g]945
14 (2021/2022)	B157	**All**	5276	0	5276	3891	1245	140	[h]31
15 (2022/2023)	B158	**All**	5105	0	5105	3324	1612	169	[i]28

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. [f]: including 1319 students declared as final drop-outs because of not having participated for a period of two years. [g]: including 937 students declared as final drop-outs because of not having participated for a period of two years. [h]: including 0 students declared as final drop-outs because of not having participated for a period of two years. [i]: including 0 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 15 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 15

To account for the wave-specific participation decision of students response propensity reweighting 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 15 the panel cohort has reduced to 5,105 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 15 mostly positive. In spite of previous reports no other covariates such as gender or migration background have an influence on the current status.

Table 2: Models estimating the individual participation propensity for students in Wave 15 of SC 4 used to derive adjustment factors for adjusted wave-specific cross-sectional and longitudinal weights.

	Wave 15
Constant	-1.940^{***}
	(0.109)
Student participated in wave 11	0.213^{***}
	(0.071)
Student participated in wave 12	0.497^{***}
	(0.065)
Student participated in wave 13	0.763^{***}
	(0.063)
Student participated in wave 14	1.410^{***}
	(0.050)
Observations	5,105

Notes: Reference categories are: 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. Standard errors are given in parentheses. ***, **, 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.

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:SC4:15.0.0.

Besides individual/target (ID_t) and institutional (ID_i) identifiers, design information for the entire cohort is made available.[2] 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).[3] 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.[4] 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_cal. Thus, corresponding cross-sectional and longitudinal weights are also calibrated to 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 set for SC4 SUF version 15.0.0.

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)

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

[3] In the SUF, these design variables are named differently, because of an error in data preperation. Here, variables stratum_exp, stratum_imp1, stratum_imp2 and stratum_imp3 are namend stratum_imp1, stratum_imp2, stratum_imp3 and stratum_imp4.

[4] 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 set for SC4 SUF version 15.0.0. (continued)

Variable	Applies to	Content
Identifier		
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
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	11,830	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	4,828	Cross-sectional weight for targets participating in Wave 12
w_t13	4,511	Cross-sectional weight for targets participating in Wave 13
w_t14	3,891	Cross-sectional weight for targets participating in Wave 14
w_t15	3,324	Cross-sectional weight for targets participating in Wave 15
w_t1to2	15,056	Longitudinal weight for targets participating in Wave 1 to 2
w_t1to3	13,188	Longitudinal weight for targets participating in Wave 1 to 3
w_t1to4	1,226	Longitudinal weight for targets participating in Wave 1 to 4
w_t1to6	4,677	Longitudinal weight for targets participating in Wave 1 to 6
w_t1to7	9,463	Longitudinal weight for targets participating in Wave 1 to 7
w_t1to8	7,425	Longitudinal weight for targets participating in Wave 1 to 8
w_t1to9	5,962	Longitudinal weight for targets participating in Wave 1 to 9
w_t1to10	4,896	Longitudinal weight for targets participating in Wave 1 to 10
w_t1to11	3,674	Longitudinal weight for targets participating in Wave 1 to 11
w_t1to12	2,808	Longitudinal weight for targets participating in Wave 1 to 12
w_t1to13	2,402	Longitudinal weight for targets participating in Wave 1 to 13
w_t1to14	2,021	Longitudinal weight for targets participating in Wave 1 to 14
w_t1to15	1,684	Longitudinal weight for targets participating in Wave 1 to 15

Table 4: Summary statistics for all weights provided.

Label of weights	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA.s
w_i	1.7195	14.2587	18.1477	19.3806	24.2512	100.0490	NA
w_t	1.8604	32.9759	44.2869	49.0495	56.1641	2950.4497	NA
w_t_cal	2.4428	32.6194	44.5160	50.4371	58.2075	2004.5527	NA
w_t1	0.0520	0.6761	0.9175	1.0000	1.2004	2.8581	319
w_t2	0.0530	0.6833	0.9057	1.0000	1.1864	2.9052	1210
w_t3	0.0377	0.4791	0.6329	1.0000	0.9488	4.7910	2414
w_t4	0.0260	0.2349	0.3335	1.0000	0.6362	5.6620	15074
w_t5	0.0262	0.2673	0.4312	1.0000	0.9004	5.3847	3443
w_t6	0.0308	0.4017	0.5525	1.0000	0.8544	5.1603	11033
w_t7	0.0105	0.0487	0.3787	1.0000	1.0171	5.5809	4595
w_t8	0.0103	0.0539	0.3508	1.0000	0.9475	5.6280	6554
w_t9	0.0081	0.0559	0.3018	1.0000	0.8625	5.6818	7381
w_t10	0.0059	0.0480	0.2422	1.0000	0.8374	5.7270	8439
w_t11	0.0051	0.0399	0.2158	1.0000	0.8744	5.7435	10153
w_t12	0.0048	0.0321	0.1887	1.0000	0.8448	5.7861	11597
w_t13	0.0039	0.0321	0.1708	1.0000	0.8728	5.7863	11914
w_t14	0.0034	0.0260	0.1581	1.0000	0.9290	5.7862	12534
w_t15	0.0026	0.0197	0.1292	1.0000	0.9157	5.8291	13101
w_t1to2	0.0537	0.6898	0.9145	1.0000	1.1947	2.8281	1369
w_t1to3	0.0494	0.6176	0.8046	1.0000	1.1217	3.4410	3237
w_t1to4	0.0409	0.3552	0.4994	1.0000	0.7734	5.4247	15199
w_t1to5	0.0435	0.4215	0.6048	1.0000	1.1095	4.5934	5388
w_t1to6	0.0323	0.4065	0.6329	1.0000	1.0206	4.7682	11748
w_t1to7	0.0183	0.0716	0.5188	1.0000	1.2570	5.3248	6962
w_t1to8	0.0173	0.0715	0.4981	1.0000	1.2147	5.3715	9000
w_t1to9	0.0180	0.0739	0.5013	1.0000	1.2208	5.3732	10463
w_t1to10	0.0178	0.0718	0.4759	1.0000	1.2023	5.4095	11529
w_t1to11	0.0186	0.0700	0.4432	1.0000	1.1799	5.4847	12751
w_t1to12	0.0199	0.0687	0.4229	1.0000	1.1898	5.4806	13617
w_t1to13	0.0200	0.0680	0.3906	1.0000	1.1846	5.5069	14023
w_t1to14	0.0203	0.0680	0.2938	1.0000	1.1881	5.5343	14404
w_t1to15	0.0215	0.0712	0.2185	1.0000	1.1881	5.5413	14741

For further information on weighting please contact statistik@lifbi.de.

Acknowledgements

This paper uses data from the National Educational Panel Study (NEPS; see Blossfeld and Roßbach, 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network.

5 References

Bergrab, M. (2020). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 11)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/11-0-0/SC4_11-0-0_W.pdf

Bergrab, M. (2021). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 12)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_W.pdf

Bergrab, M. (2023). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 13)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/13-0-0/SC4_13-0-0_W.pdf

Bergrab, M. (2024). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 14)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/14-0-0/SC4_14-0-0_W.pdf

Blossfeld, H.-P., & Roßbach, H. G. (Eds.). (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS). Edition ZfE* (2nd). Springer VS.

Blossfeld, H.-P., Roßbach, H. G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (Vol. 14). VS Verlag für Sozialwissenschaften.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Steinhauer, H. W. (2019). *Samples, Weights, and Nonresponse: the Sample of Starting Cohort 4 of the National Educational Panel Study (Wave 10)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. 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. https://doi.org/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). Leibniz Institute for Educational Trajectories. Bamberg. 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 4 of the National Educational Panel Study (Wave 1 to 6)* (NEPS Survey Paper No. 2). Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. 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)*. Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/9-0-0/SC4_9-0-0_W.pdf