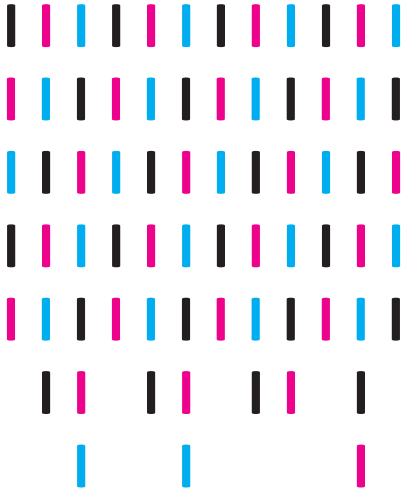

NEPS SURVEY PAPERS

Sabine Zinn, Hans Walter Steinhauer, and Christian Aßmann

SAMPLES, WEIGHTS, AND NONRESPONSE: THE STUDENT SAMPLE OF THE NATIONAL EDUCATIONAL PANEL STUDY (WAVE 1 TO 8)

NEPS Survey Paper No. 18
Bamberg, March 2017

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

Survey Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Survey Paper Series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

The NEPS Survey Papers are edited by a review board consisting of the scientific management of LIfBi and NEPS.

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS Survey Papers are available at https://www.neps-data.de (see section “Publications“).

Editor-in-Chief: Corinna Kleinert, LIfBi/University of Bamberg/IAB Nuremberg

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – contact@lifbi.de

Samples, Weights, and Nonresponse: the Student Sample of the National Educational Panel Study (Wave 1 to 8)

Zinn, S., Steinhauer, H.W., & Aßmann, C.

Leibniz Institute for Educational Trajectories

Technical Report referring to DOI:10.5157/NEPS:SC5:8.0.0

E-mail address of lead author:

methoden@lifbi.de

Bibliographic data:

Zinn, S., Steinhauer, H.W. & Aßmann, C. (2017). *Samples, Weights, and Nonresponse: the Student Sample of the National Educational Panel Study (Wave 1 to 8)* (NEPS Survey Paper No. 18). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study. doi:10.5157/NEPS:SP18:1.0

Samples, Weights, and Nonresponse: the Student Sample of the National Educational Panel Study (Wave 1 to 8)

Abstract

This report documents the target population, the sampling, the sample sizes, and the weighting procedures of the Waves 1 to 8 of the NEPS Starting Cohort 5 (SC5, first-year undergraduate students in higher education). It introduces the target population of the Starting Cohort and the sampling design applied. Furthermore, the composition of the gross and the net samples of the different waves are detailed. The derivation of the sampling weights is described. This includes the computation of the design weights and the accordant nonresponse adjustments. In this context, the selectivity due to nonresponse and attrition is inquired into. This report concludes with a summary of the design variables and sampling weights as well as some comments regarding the usage of sampling weights in statistical analysis.

1. Prequel

This report refers to the Scientific Use File (SUF) doi:10.5157/NEPS:SC5:8.0.0 of the survey "first-year undergraduate students in higher education in 2011" (Starting Cohort 5, SC5) conducted within the National Educational Panel Study (NEPS).The SC5 survey is part of the main cohort samples of the NEPS and focuses on central issues such as educational choices, competence development, the benefits of higher education, and entry into the job market. On the basis of a short review of the survey and the sampling design applied, this report presents information on the initial sample and results of the weighting procedures applied. Weighting for these students involves a step-by-step process. First, a correction of design weights was performed in order to adequately reflect the current numbers of students based on data from the Federal Statistical Office of Germany for the winter semester 2010/2011. Second, weights for participating students were calculated for eight studies and survey waves, respectively, see Table 1. The studies B52 (Wave 1), B55 (Wave 3), B59 (Wave 5), and B94 (Wave 7) were conducted via computer-assisted telephone interviews (CATIs). The studies B54 (Wave 2), B56 (Wave 4), B58 (Wave 6), and B95 (Wave 8) are online surveys. The study B53 (Wave 1 Test) involves competence tests that have been conducted in parallel to the telephone interviews of the B52 study.[1] Table 6 in Appendix A depicts the wave-specific number of participants, temporary dropouts, and final drop-outs in and after the survey.

The remainder of this paper is structured as follows: Section 2 presents the target population of SC5 and details the sampling design of the initial SC5 sample. The following Section 3 describes the derivation of design weights and different sets of nonresponse adjusted design weights. Section 4 gives the procedure applied for trimming and standardizing the weights. Finally, Section 5 summarizes the survey weights provided in SUF 8.0.0 and gives advices regarding their usage.

2. Population and Sampling Design

The target population is defined as all first-year students (German and non-German) enrolled for the first time in public or state-approved institutions of higher education in Germany who are aiming at a Bachelor's degree, a state examination (*Staatsexamen*) in medicine, law, pharmacy, and teaching, a diploma or Master's degree in Roman Catholic or Protestant theology or specific art and design degrees in the academic year of 2010/2011. Students attending universities, technical universities or universities of applied sciences run by Federal Ministries or Federal States for members of their public services are excluded.[2]

[1] Because of methodological issues, no sampling weights are provided for students attending the competence tests of Wave 5 (B57) and Wave 7 (B90). In Wave 5, different test modes had been used to measure competence (online, distinct computer based assessment modes, and paper based assessment). To not create the impression that competence measures measured by distinct modes are comparable per se, no survey weights are provided. In Wave 7, only students studying *BWL* had been tested. In sum, only 338 students of the 17910 panel members attended the test. For this pre-selected group, survey weights have not been computed as well.

[2] In the beginning, the plan was to conduct a census among the students with a non-traditional admission certificate. However, difficulties during the recruiting process hindered this project. In detail this means that even though students with a non-traditional admission certificate were contacted separately, namely by conventional mail, a significant part of them was additionally recruited in the same way as students with traditional

Table 1: Attribution of studies to panel waves.

Wave	Study	Survey Time
Wave 1	B52 CATI	Winter 2010/11
Wave 1 Test	B53 Test	Winter 2010/11
Wave 2	B54 CAWI	Autumn 2011
Wave 3	B55 CATI	Spring 2012
Wave 4	B56 CAWI	Autumn 2012
Wave 5	B59 CATI	Spring 2013/Summer 2013
Wave 5 Test	B57 Test	Spring 2013/Summer 2013
Wave 6	B58 CAWI	Autumn 2013
Wave 7	B94 CATI	Summer 2014
Wave 7 Test	B90 Test	Winter/Spring 2014
Wave 8	B95 CAWI	Autumn 2014

A stratified cluster sample was drawn from the defined population of first-year students at corresponding higher education institutions, see also Aßmann et al. (2011). We define a cluster as all students enrolled in a certain subject (of the sixty officially listed fields, see Table 2) [3] at a particular higher education institution. For example, all students studying social sciences (*Sozialwissenschaften*) at the (*public*) University of Bamberg form one cluster. Within each cluster, all students are to be surveyed. The student cohort has been set up to incorporate an oversampling of teacher education students and students attending private higher education institutions, that is, private universities and private universities of applied sciences. This objective is addressed by setting up a first stratification level according to educational institution. This first stratification level defines four strata: Stratum h_1 comprises the clusters linked to teacher education at public universities. Stratum h_2 is set up to include all fields of study (except of teacher education) at public universities, whereas stratum h_3 summarizes all fields of study offered by public universities of applied sciences. Finally, stratum h_4 comprises all degree programs offered by private universities or private universities of applied sciences. This level of stratification allows us to carry out an oversampling of teacher education students and students at private higher education institutions by using different sampling rates of clusters in the different strata. Overall, the plan was to establish a gross sample of 66,450 students[4]– 15,950 students in stratum h_1, 26,500 students in stratum h_2, 16,800 students in stratum h_3, and 7,200 students in stratum h_4.

Given the heterogeneous distribution of students across the officially listed fields of study, sampling within the defined strata would result in a large sampling variation concerning the coverage of the range of subjects within the sample. Hence, a further level of stratification was

admission certificate, namely in courses targeted at first-year students. As a consequence, in the end it was impossible to disentangle both groups of students completely. Therefore, in the sampling process students with traditional and students with non-traditional admission certificate were not further differentiated.

[3] In contrast to the definition provided by the Federal Statistical Office of Germany we separated three clusters of teacher training programmes from the fields of subjects and added them to the list.

[4] Assuming that a quarter of the sampled students participates, this yields approximately the intended net sample size of 16,500 students, see, for example, Aßmann et al. (2011).

Table 2: Allocation of the sixty listed fields of study to the two stratification levels h_i and s_j, with $i = 1, \cdots, 4, j = 1, \cdots, 29$.

Code	Officially listed subject	h_1	h_2	h_3	h_4
1	*Sprach- und Kulturwissenschaften allgemein*	–	s_5	s_{20}	s_{27}
2	*Evangelische Theologie, Religionslehre*	–	s_4	s_{20}	s_{27}
3	*Katholische Theologie, Religionslehre*	–	s_4	s_{20}	s_{27}
4	*Philosophie*	–	s_4	–	s_{27}
5	*Geschichte*	–	s_4	–	s_{27}
6	*Bibliothekswissenschaft, Dokumentation, Publizistik*	–	s_5	s_{20}	s_{27}
7	*Allgemeine und vergleichende Literatur- und Sprachwissenschaft*	–	s_5	s_{20}	s_{27}
8	*Altphilologie (klassische Philologie), Neugriechisch*	–	s_4	–	s_{27}
9	*Germanistik (Deutsch, germanische Sprachen ohne Anglistik)*	–	s_6	–	s_{27}
10	*Anglistik, Amerikanistik*	–	s_7	–	s_{27}
11	*Romanistik*	–	s_7	–	s_{27}
12	*Slawistik, Baltistik, Finno-Ugristik*	–	s_7	–	s_{27}
13	*Außereuropäische Sprach- und Kulturwissenschaften*	–	s_7	s_{20}	s_{27}
14	*Kulturwissenschaften i.e.S.*	–	s_7	s_{20}	s_{27}
15	*Psychologie*	–	s_8	–	s_{27}
16	*Erziehungswissenschaften*	–	s_8	s_{21}	s_{27}
17	*Sonderpädagogik*	–	s_8	s_{21}	s_{27}
18	*Sport, Sportwissenschaft*	–	s_8	s_{20}	s_{27}
19	*Wirtschafts- und Gesellschaftslehre allgemein*	–	s_9	s_{20}	s_{27}
20	*Regionalwissenschaften*	–	s_9	s_{20}	s_{27}
21	*Politikwissenschaften*	–	s_9	s_{20}	s_{27}
22	*Sozialwissenschaften*	–	s_9	s_{20}	s_{27}
23	*Sozialwesen*	–	s_8	s_{21}	s_{27}
24	*Rechtswissenschaft*	–	s_{10}	s_{20}	s_{27}
25	*Verwaltungswissenschaft*	–	s_{10}	s_{20}	s_{27}
26	*Wirtschaftswissenschaften*	–	s_{11}	s_{22}	s_{28}
27	*Wirtschaftsingenieurwesen*	–	s_{11}	s_{22}	s_{28}
28	*Mathematik, Naturwissenschaften allgemein*	–	s_{14}	s_{23}	s_{29}
29	*Mathematik*	–	s_{12}	s_{23}	s_{29}
30	*Informatik*	–	s_{12}	s_{23}	s_{29}
31	*Physik, Astronomie*	–	s_{12}	–	s_{29}
32	*Chemie*	–	s_{13}	s_{23}	s_{29}
33	*Pharmazie*	–	s_{13}	s_{23}	s_{29}
34	*Biologie*	–	s_{14}	–	s_{29}
35	*Geowissenschaften (ohne Geographie)*	–	s_{14}	s_{23}	s_{29}
36	*Geographie*	–	s_{14}	–	s_{29}
37	*Gesundheitswissenschaften allgemein*	–	s_{15}	s_{23}	s_{29}
38a	*Humanmedizin ohne Zahnmedizin (ohne Approbation)*	–	s_{15}	–	s_{29}
38b	*Humanmedizin ohne Zahnmedizin (mit Approbation)*	–	s_{19}	–	s_{29}
39	*Zahnmedizin*	–	s_{15}	–	s_{29}
40	*Veterinärmedizin*	–	s_{15}	–	s_{29}
41	*Landespflege, Umweltgestaltung*	–	s_{15}	s_{23}	s_{29}
42	*Agrarwissenschaften, Lebensmittel- und Getränketechnologie*	–	s_{15}	s_{23}	s_{29}
43	*Forstwissenschaft, Holzwirtschaft*	–	s_{15}	s_{23}	s_{29}
44	*Ernährungs- und Haushaltswissenschaften*	–	s_{15}	s_{23}	s_{29}
45	*Ingenieurwesen allgemein*	–	s_{17}	–	s_{29}
46	*Bergbau, Hüttenwesen*	–	s_{17}	s_{26}	s_{29}
47	*Maschinenbau/Verfahrenstechnik*	–	s_{16}	s_{24}	s_{29}
48	*Elektrotechnik*	–	s_{17}	s_{25}	s_{29}
49	*Verkehrstechnik, Nautik*	–	s_{17}	s_{26}	s_{29}
50	*Architektur, Innenarchitektur*	–	s_{17}	s_{26}	s_{29}
51	*Raumplanung*	–	s_{17}	s_{26}	s_{29}
52	*Bauingenieurwesen*	–	s_{17}	s_{26}	s_{29}
53	*Vermessungswesen*	–	–	s_{26}	s_{29}
54	*Kunst, Kunstwissenschaft allgemein*	–	s_{18}	s_{20}	s_{27}
55	*Bildende Kunst*	–	s_{18}	s_{20}	s_{27}
56	*Gestaltung*	–	s_{18}	s_{20}	s_{27}
57	*Darstellende Kunst, Film und Fernsehen, Theaterwissenschaft*	–	s_{18}	s_{20}	s_{27}
58	*Musik, Musikwissenschaft*	–	s_{18}	s_{20}	s_{27}
59	*Außerhalb der Studienbereichsgliederung/Sonstige Fächer*	–	s_{18}	–	s_{27}
60a	*Lehramt: LA Grund+Haupt/LA Grund/LA Haupt/BA Sek I+Primar/ LA+BA Grundschule+SekI/LA Real/LA Real+BA Real+Haupt/ LA+BA Sonder+Förder*	s_1	–	–	–
60b	*Lehramt: LA Gym/BA Gym/BA allg./LA Oberstufe+Sek II/ LA+BA Berufl./LA Ober+Sek II+berufl.*	s_2	–	–	–
60c	*Lehramt: BA Lehramt allg.*	s_3	–	–	–

Table 3: Number of clusters sampled and realized in each stratum.

Stratum		**Number of clusters**	
1^{st} level	2^{nd} level	sampled	realized
h_1	s_1	21	18
	s_2	26	25
	s_3	7	9
h_2	s_4	10	11
	s_5	9	9
	s_6	8	9
	s_7	16	10
	s_8	18	20
	s_9	17	18
	s_{10}	8	8
	s_{11}	18	21
	s_{12}	24	23
	s_{13}	11	12
	s_{14}	17	15
	s_{15}	10	8
	s_{16}	5	9
	s_{17}	14	12
	s_{18}	12	9
	s_{19}	6	7
h_3	s_{20}	15	14
	s_{21}	12	13
	s_{22}	35	35
	s_{23}	31	28
	s_{24}	15	20
	s_{25}	13	9
	s_{26}	24	23
h_4	s_{27}	21	13
	s_{28}	29	19
	s_{29}	21	17

Note: Discrepancies between the number of sampled and realized clusters are caused by (i) whole clusters dropping out and (ii) incorrect information of students about their main subject. We use poststratification to correct for these deficiencies.

introduced where strata are defined by groups of related subjects. This stratification was accompanied by an exclusion of clusters with less than thirty enrolled students in the academic year of 2008/2009. In summary, the sixty fields of study are pooled in several subject groups within the first-level stratum, see Table 2. Thus, the strata s_1 to s_3 pool fields of study in the stratum h_1, the strata s_4 to s_{19} correspond to the first-level stratum h_2, and the strata s_{20} to s_{26} comprise fields of study within the stratum h_3. Finally, pooling in the stratum h_4 is achieved by means of the second-level strata s_{27} to s_{29}.

The number of clusters to be drawn within each stratum h_1 to h_4 was determined such that the sample distribution of students across the fields of study resembled the one in the population. At the same time, the intended oversampling could be incorporated in a straightforward way and homogeneous inclusion probabilities were probable to realize. In particular, the number of clusters m_i sampled within stratum h_i is calculated according to

$$m_i = \frac{\tilde{n}_i}{\frac{1}{K_i}\sum_{k=1}^{K_i} N_{ik}}, \tag{1}$$

namely by dividing the planned sample size $\tilde{n}_i$ in stratum h_i by the average cluster size in terms of the number of first-year students N_{ik} in the academic year of 2008/2009 for all clusters $k = 1, \ldots, K_i$ in stratum h_i. Here, K_i denotes the total number of clusters in stratum h_i. In the strata h_1 and h_4 an oversampling was carried out resulting in $m_1 = 54$ clusters to be sampled for stratum h_1 and $m_4 = 71$ clusters to be sampled for stratum h_4. For the strata h_2 and h_3, where no oversampling was carried out, a total of 348 clusters to be sampled has been found sufficient to generate the planned gross sample size. Here, clusters are allocated proportionally to the overall number of clusters in both strata, resulting in $m_2 = 203$ clusters to be sampled in stratum h_2 and $m_3 = 145$ clusters in stratum h_3. For each substratum the number of clusters m_{ij} to be sampled from the stratum h_i, $i = 1, \ldots, 4$ are calculated according to

$$m_{ij} = m_i \frac{K_{ij}}{K_i}, \tag{2}$$

where K_{ij} denotes the total number of clusters in the second-level stratum s_j embedded in the first-level stratum h_i. Table 3 gives the corresponding numbers. Within each stratum h_i and s_j the m_{ij} clusters are sampled by simple random sampling without replacement so that the inclusion probability for cluster k_{ij} is given by

$$p_{ij} = \frac{m_{ij}}{K_{ij}}. \tag{3}$$

Inserting equation (2) yields

$$p_{ij} = \frac{m_i}{K_i} \tag{4}$$

and the corresponding design weight d_i is given by the inverse of that inclusion probability

$$d_i = \frac{K_i}{m_i} = \begin{cases} \frac{90}{54} & = 1.667 \quad \text{for} \quad i = 1 \\ \frac{1276}{203} & = 6.286 \quad \text{for} \quad i = 2 \\ \frac{923}{145} & = 6.366 \quad \text{for} \quad i = 3 \\ \frac{134}{71} & = 1.887 \quad \text{for} \quad i = 4. \end{cases} \tag{5}$$

To handle institutional nonparticipation, the following replacement strategy was implemented. If a university refuses to participate, all fields of study sampled at this specific university are lost. Hence, only those institutions are eligible for replacement that maintain the original sample composition with regard to the sampled departments and subjects. For each combination of sampled subjects at a particular higher education institution, all institutions offering the same combination of subjects within the frame are listed, irrespective of whether the institutions have already been sampled or not. Institutions not sampled are given preferential consideration in the choice of replacement candidates. Given that several replacement institutions offer the same combination of subjects to be replaced, the replacement institution is defined as the one with the smallest difference in numbers of enrolled students compared to the nonparticipating institution.

These steps were carried out on the basis of information on first-year students from the winter semester 2008/2009 (provided by the Federal Statistical Office of Germany). At the point of planning the sampling and recruitment procedures, these were the most current data available for the population of students. As (during the planning process) the absolute number of first-year students had risen from 2008/2009 to 2009/2010 by about 6.5%, a further rise in 2010/2011 seemed probable. This fact was taken into account by incrementing the 2008/2009 data by 10% in order to have a good estimate of the actual number of students for the sampling process in 2010.

In order to achieve high response rates, two different contact modes were employed to approach the sampled students: First, all students were informed about the NEPS and invited to participate in den panel study via conventional mail. Besides this, several institutions facilitated a second way of contact by the personal information and recruitment in courses targeted at or mandatory for first-year students. In a pilot study, this twofold recruitment process yielded higher participation rates, as well as a higher panel attendance. In total, 31,082 first-year students could be contacted via this procedure. The following section outlines the performed weighting adjustments.

3. Derivation of Survey Weights

To mirror the recruitment and participation process within the weighting adjustments, consecutive modeling of the decision and participation process is performed, see Figure 1. The first modeling step involves the correction of the stratum-specific design weights d_i in relation to the nonresponse occurring from the gross sample of students (in the clusters previously determined) to the set of students who provided (any kind of) contact information. The second modeling step corrects for nonresponse occurring from the sample of persons with contact information (of any kind) to the sample of persons with valid contact information–that is, to

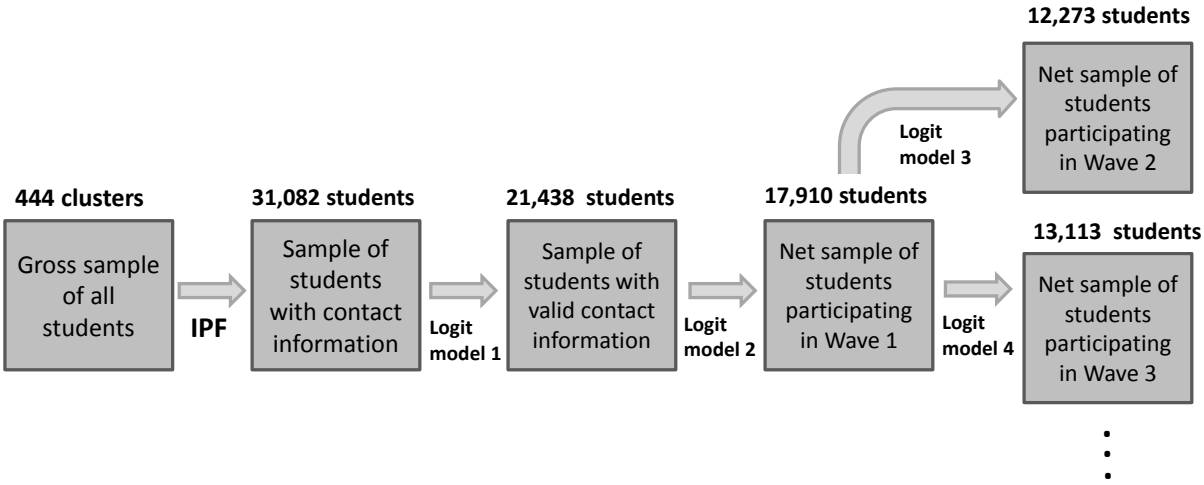

Figure 1: Steps of consecutive modeling of the decision and participation process.

the gross sample of Wave 1 (corresponding to the CATI of the study B52). All further modeling steps correct for the nonresponse among the recruited students in the distinct survey waves (i.e., in the studies B52, B53, B54, B55, B56, B58, B59, B94, and B95). The participation in the first telephone interview (i.e., in the study B52) forms the indispensable backbone of the panel study. Thus, the panel cohort is defined as the set of students who participated in Wave 1. In total, the panel cohort comprises 17,910 students. Consequently, all computations related to nonresponse adjustments in further waves refer to this set of students. With regard to the first step, an iterative proportional fitting (IPF) mechanism originally described by Deming and Stephan (1940) was implemented. The IPF uses mathematical scaling to ensure that a multidimensional table of data is adjusted so that its row and column totals correspond to constrained row and column totals obtained from alternative sources.[5] We apply the procedure to determine weighting factors for the 31,082 students who provided contact information, on the basis of current frame information on student numbers and attributes from the winter semester 2010/2011–when sampling took place. The respective variables were gender, German versus non-German students, public versus private higher education institutions, universities versus universities of applied sciences as well as an indicator variable for the subject.[6] The weighting factors derived that way are multiplied to the design weights d_i referring to the first-level strata h_1 to h_4, yielding sampling weights w^0_{ijs} for all students s in the first-level stratum h_i and in the second-level stratum s_j who have provided their contact information.

The second modeling step (logit model 1 and 2 in Figure 1) determines the propensity of students to actually participate in Wave 1. Therefore, first the loss occurring from the sample of students with contact information (i.e, the recruited sample) to the sample of students with valid contact information (i.e., the gross sample of Wave 1) is modeled. Thereafter, the decision of all contacted students to actually participate is specified. The variables considered here are gender, nationality (German, foreign, unknown), type of institution (university, *Fachhochschule*, abroad/not specified), year of birth, intended university degree (Bachelor, *Staat-*

[5]To this end, values of the original table are gradually adjusted through repeated calculations to fit row and column constraints.

[6]The corresponding data were taken from the Federal Statistical Office of Germany (Statistisches Bundesamt, 2011).

sexamen Lehramt, other) and type of contact (personal or postal). Note that for only 26,913 of the 31,082 students who provided any kind of contact information enough (valid) data were available to include them into the analysis. Only 18,030 of the 21,438 students who were asked to participate in the first wave could be taken into account in the second model because they provided sufficient information on the considered variables. The corresponding empirical analysis is performed under the assumption that data are missing completely at random. Table 7 and 8 (given in the Appendix B) document the results of the corresponding models.[7] We find that, of those students who gave valid contact information, females, German students, students aiming for a teacher training programme or a *Staatsexamen*, and students who were contacted by mail, could be assigned to the Wave 1 gross sample with a significantly higher probability than their counterparts. With regard to the Wave 1 gross sample, we find significantly higher participation propensities of females, German students, students studying at university, and students who were contacted by mail. On the basis of the outcome of the two logit models presented, adjustment factors for all students participating in Wave 1 are computed. Multiplying these by the weights w^0_{ijs} yields the (cross-sectional) weights w^1_{ijs} of students to attend Wave 1. We correct for potential deviation of the weights distribution from the distribution of first-year students in winter semester 2010/2011 in the distinct fields of study[8] and in the first strata by poststratification, and align the weights w^1_{ijs} accordingly.

Participation modeling of Wave 2 and all further waves (i.e., studies B54, B53, B55, B56, B57, B58, B59, B94, and B95) is based on the panel cohort (i.e., the sample of Wave 1), see Figure 1. In the response/nonresponse models the following variables are considered (with values given in parentheses):

- participation in previous waves (always, often, seldom)[9],
- type of institution (university or university of applied science),
- funding of institution (public or private),
- gender (female and male),
- educational degree of parents (measured by CASMIN categories),
- migration background (measured by generation status),
- household size (one person, two persons, more than two persons),
- kids in household (yes or no),
- region (Eastern and Western Germany)
- year of birth (before 1989, in 1989 and 1990, later than 1990),
- reading ability (quantiles, measured by NEPS tests in the study B53),
- teacher education (yes or no),

[7] The estimation of these two models and the related data preparation were conducted by Martin Kleudgen and Reiner Gilberg from *infas - Institut für angewandte Sozialwissenschaften GmbH*.

[8] The following ten categories were considered: *Spach-/Kulturwissenschaften, Rechts-/Wirtschafts-/Sozialwissenschaften, Mathematik/Naturwissenschaften, Humanmedizin, Agrar-/Forst-/Ernährungswissenschaften, Ingenieurwissenschaften, Kunst, Lehramt*.

[9] The definition of the participation frequency depends on the number of Waves preceding the current Wave. Concretely, we have defined this variable as follows: always (permanent participation in all preceding waves), often (no permanent participation but participation in more than 0.6 percent of all preceding waves), rare (otherwise).

- traditional higher education entrance qualification (yes and no),
- school leaving qualification (German university entrance qualification (*Abitur*), Nongerman university entrance qualification, no general university entrance qualification) and
- the field of study (nine categories, see Table 9 in the Appendix B).

The variables 'kids in household', 'household size', 'field of study', and 'teacher education' are modeled to be time-dependent and updated every time when more recent data is available. Missing values are considered by defining accordant missing categories. The results of the corresponding logit models are given in Tables 10 to 17 (see Appendix B).

As expected we find that the participation in previous waves is a very strong indicator for the propensity to participate in future waves. Furthermore, as already noted before, usually more women and students in teacher education take part in the survey than men and students who are not in teacher education. A further impediment to participation is low reading ability, having a migration background, and being older than the average (i.e., being born before 1989). Furthermore, household size effects participation propensity. There is strong evidence that one person households take more often part in online surveys than households with more persons. The corresponding results for telephone interviews are inconclusive but for two studies (i.e., B59 and B94) we find households with more than one person overrepresented among the participants. Finally, studying in Western Germany has a negative impact on the participation propensity in online studies, but partly a positive effect for participating in telephone interviews (see Table 14 showing the results for B59 (CATI)).

On the basis of all estimated models participation probabilities are predicted and adjustment factors are derived.[10] By means of these adjustment factors, cross-sectional sampling weights w_{ijs}^{c} for participating in the single survey waves $c = 2, \ldots, 8$ are computed. Likewise, distinct sets of longitudinal sampling weights w_{ijs}^{l}, $l \subseteq \{2, \ldots, 8\}$, (e.g., for always participating or for participating in all CATI interviews) can be derived. However, as the set of possible participation patterns becomes highly complex with an increasing number of survey waves conducted, the set of longitudinal weights provided is restricted to only successive waves and/or to the survey mode–that is, CATI or online, see Table 4.

4. Trimming and Standardizing Weights

To possibly increase the statistical efficiency of weighted analysis, the adjusted weights were trimmed. The general goal of weight trimming is to reduce sampling variance and, at the same time, to compensate for potential increase in bias. Trimming was performed using the so-called "Weight Distribution" approach Potter (1990). Here, design weights are assumed to follow an inverse beta distribution with a cumulative distribution function F_w. Parameters of the sampling weight distribution are estimated using the sampling weights, and a trimming level u is computed, whose occurrence probability is 1%, that is, $1 - F_w(u) = 0.01$. Sampling weights in excess of u are trimmed to this level and the excess is distributed among the untrimmed weights. The parameters for the sampling weight distribution are then again estimated using the trimmed adjusted weights, and a revised trimming level $\tilde{u}$ is computed. The trimmed adjusted weights are compared to the revised level $\tilde{u}$. If any weights are in excess of $\tilde{u}$, they are

[10] Adjustment factors are defined as the inverse participation probabilities.

trimmed to this level, and the excess is distributed among the untrimmed weights. This procedure is iteratively repeated until no weights are in excess of a newly revised trimming level. To ease statistical analysis, the trimmed sampling weights are standardized with mean 1.

5. Summary of Weights and Advice Regarding the Usage of Weights

All weights are provided in a trimmed and standardized form. For Wave 1, additionally a set of extrapolated cross-sectional weights is given allowing extrapolating sample distributions to the population level of first-year students in the winter semester 2010/2011 according to field of study, type of institution, sex, nationality, and kind of funding. Table 4 lists the types of weights provided for SUF release version 8-0-0 and Table 5 gives some summary statistics of the weights provided.

Table 4: Types of weights provided.

Type of weight	Label
Weights of strata	`w_h`
Weights of students participating in B52	`w_t1`
Weights (extrapolated) of students participating in B52	`w_t1ext`
Weights of students participating in B53	`w_t1comp`
Weights of students participating in B54	`w_t2`
Weights of students participating in B55	`w_t3`
Weights of students participating in B56	`w_t4`
Weights of students participating in B59	`w_t5`
Weights of students participating in B58	`w_t6`
Weights of students participating in B94	`w_t7`
Weights of students participating in B95	`w_t8`
Weights of students participating in all online studies B54, B56, B58,& B95	`w_t12468`
Weights of students participating in the online studies B54, B56, & B58	`w_t1246`
Weights of students participating in all online studies B52, B55, B59,& B94	`w_t1357`
Weights of students participating in all Waves	`w_t12345678`

No general recommendation for the usage of sampling weights can be given. Whether, and if so how, weights have to be used depends on the problem to be studied, see for example Solon, Haider, and Wooldridge (2013) for recommendations for empirical practice. It is commonly recommended to apply sampling weights when conducting descriptive statistics. For analytical analysis, models have to be tested for their dependence on the sampling design. Specifically, this means that the user has to ensure that the way of sampling has no or only negligible effect on the model results or/and that the sampling design is adequately considered in the model specification. A general description of how to test and account for the sampling design is given in, for example, Snijders and Bosker (2012). As a guideline, we recommend including the stratum information (to account for the unequal selection probabilities of clusters in the distinct strata) into the model under consideration. Furthermore, all variables that have been found to have a significant effect on the response probability of the considered sample should be included as explanatory variables.

Table 5: Summary statistics for (trimmed and standardized) weights.

Label of weight	Number of students	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.
w_t1	17,910	0.009	0.329	0.997	1.000	1.328	3.386
w_t1ext	17,910	0.174	6.020	18.270	18.470	24.330	325.300
w_t1comp	5,949	0.142	0.302	0.824	1.000	1.292	4.139
w_t2	12,273	0.009	0.347	0.923	1.000	1.333	3.678
w_t3	13,113	0.008	0.309	0.877	1.000	1.275	3.906
w_t4	11,202	0.008	0.306	0.836	1.000	1.275	4.114
w_t5	12,698	0.008	0.314	0.865	1.000	1.303	3.949
w_t6	10,183	0.021	0.318	0.796	1.000	1.271	4.260
w_t7	9,547	0.007	0.576	0.795	1.000	1.118	3.807
w_t8	8,629	0.011	0.265	0.749	1.000	1.143	4.698
w_t12468	5,853	0.026	0.333	0.825	1.000	1.323	4.052
w_t1246	5,853	0.042	0.544	1.348	1.598	2.161	5.123
w_t1357	7,645	0.008	0.055	0.807	1.000	1.175	3.723
w_t12345678	3,673	0.182	0.527	0.827	1.000	1.329	3.447

The *survey* package of Stata allows defining the survey design of the sample at hand, and thus conducting design-based inference, see for example Kreuter and Valliant (2007). The accordant command for the whole SC5 sample is

```
gen f_h = w_h^{-1}
svyset ID_cl [pweight=w_t1], strata(stratum) fpc(f_h)
```

In this command, `f_h` gives the sampling rate used as final population correction factor, `ID_cl` determines the cluster membership of a sampled student, and `w_t1` describes the corresponding survey weight (to be part of the SC5 sample). The term `stratum` is self-explanatory. All subsequent analysis has to be preceded by the prefix `svy`. Also the statistical software R provides a survey package to deal with design-based inference, see Lumley (2004). Here, the definition of a design object is similar to the one asked for in Stata.

References

Aßmann, C., & Sixt, M. (2013). *Umgang mit (temporären) Ausfällen im Nationalen Bildungspanel (NEPS) - Startkohorten-übergreifende Regelung* (Internal Discussion Paper). Leibniz Institute for Educational Trajectories. (Available on request from the corresponding author.)

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., & Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: challenges and solutions. In H.-P. Blossfeld, H. G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (Vol. 14, p. 51-65). VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0181-8

Deming, W., & Stephan, F. (1940, Dec). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, *11*(4), 427-444. doi: 10.1214/aoms/1177731829

Kreuter, F., & Valliant, R. (2007). A survey on survey statistics: What is done and can be done in stata. *Stata Journal*, *7*(1), 1-21.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(1), 1-19.

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In American Statistical Association (Ed.), *Proceedings of the survey research methods section* (p. 225-230). Retrieved 17.02.2014, from http://www.amstat.org/sections/srms/Proceedings/papers/1990_034.pdf

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Solon, G., Haider, S., & Wooldridge, J. (2013, February). *What are we weighting for?* (NBER Working Paper No. 18859). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18859

Statistisches Bundesamt. (2011). *Fachserie 11 reihe 4.1: Bildung und kultur: Studierende an hochschulen: Wintersemester 2010/2011* (Tech. Rep.). Statistisches Bundesamt. Retrieved from https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00006845/2110410117004.pdf

A. Case Numbers, Respondents, Nonrespondents and Final Dropouts

Table 6: Case numbers, respondents, nonrespondents and final drop-outs.

Wave	Sub-sample	Panel sample	Gross sample	Participants	Participation proportion	Temporary dropouts	Final dropouts (within wave)	Final dropouts (after wave)
1	**Total**	-	31082	17910	0.576	0	13172	0
	LA	-	7864	5555	0.706	0	2309	0
	UNI	-	11904	8024	0.674	0	3880	0
	FH	-	7460	3894	0.522	0	3566	0
	PR	-	3854	437	0.113	0	3417	0
1T	**Total**	17910	17910	5949	0.332	11942	20	0
	LA	5555	5555	2021	0.364	3527	7	0
	UNI	8024	8024	2715	0.338	5303	6	0
	FH	3894	3894	1115	0.286	2772	7	0
	PR	437	437	98	0.224	339	0	0
2	**Total**	17890	17910	12273	0.685	5610	27	13
	LA	5548	5555	3839	0.691	1708	8	2
	UNI	8018	8024	5609	0.699	2400	15	8
	FH	3887	3894	2510	0.645	1381	3	3
	PR	437	437	315	0.721	121	1	0
3	**Total**	17850	17855	13113	0.735	4562	180	28
	LA	5538	5540	4253	0.768	1236	51	8
	UNI	7995	7995	5841	0.731	2076	78	10
	FH	3881	3884	2701	0.696	1138	45	8
	PR	436	436	318	0.729	112	6	2
4	**Total**	17642	17635	11202	0.635	6444	17	15
	LA	5479	5478	3695	0.674	1784	4	2
	UNI	7907	7903	5003	0.632	2910	6	9

Table 6: Case numbers, respondents, nonrespondents and final drop-outs.

Wave	Sub-sample	Panel sample	Gross sample	Participants	Participation proportion	Temporary dropouts	Final dropouts (within wave)	Final dropouts (after wave)
	FH	3828	3828	2219	0.579	1607	7	4
	PR	428	428	285	0.666	143	0	0
5	**Total**	17610	17621	12694	0.721	4630	297	3
	LA	5473	5476	4186	0.764	1218	72	0
	UNI	7892	7897	5615	0.711	2153	129	0
	FH	3817	3820	2582	0.676	1148	90	3
	PR	428	428	311	0.727	111	6	0
5T	**Total**	17310	17479	8767	0.506	8707	5	61
	LA	5401	5446	2907	0.534	2538	1	17
	UNI	7763	7830	3963	0.506	3865	2	30
	FH	3724	3778	1687	0.447	2089	2	10
	PR	422	425	210	0.494	215	0	4
6	**Total**	17244	17255	10183	0.590	7049	23	6
	LA	5383	5385	3352	0.622	2029	4	1
	UNI	7731	7736	4594	0.594	3127	15	4
	FH	3712	3716	1975	0.531	1737	4	1
	PR	418	418	262	0.627	156	0	0
7	**Total**	17205	14457	9547	0.660	4488	422	2103
	LA	5378	2639	1906	0.722	670	63	564
	UNI	7712	7702	5101	0.662	2418	183	976
	FH	3707	3699	2265	0.612	1278	156	518
	PR	418	417	275	0.659	122	20	45
7T	**Total**	14690	600	338	0.563	237	25	2
	LA	4751	57	38	0.667	19	0	0
	UNI	6553	343	202	0.589	126	15	0

Table 6: Case numbers, respondents, nonrespondents and final drop-outs.

Wave	Sub-sample	Panel sample	Gross sample	Participants	Participation proportion	Temporary dropouts	Final dropouts (within wave)	Final dropouts (after wave)
	FH	3033	158	76	0.481	73	9	2
	PR	353	42	22	0.524	19	1	0
8	**Total**	14663	14665	8629	0.588	6025	11	0
	LA	4751	4751	2933	0.617	1817	1	0
	UNI	6531	6533	3945	0.603	2588	7	0
	FH	3022	3022	1546	0.512	1473	3	0
	PR	352	352	205	0.582	147	0	0

Notes: (i) *LA*: students in teacher education, *UNI*: students at public university without *LA*, *FH*: students at public universities of applied science, *PR*: students at private universities, (ii) 'T' indicates testing, (iii) Discrepancies between the sizes of the gross and the panel cohort samples are due to the short time periods available between forming the wave-specific gross samples and recording all the final drop-outs from previous waves. In some cases, the study of the previous wave was still running while the next wave-specific study already started.

B. Nonresponse Modeling: Variables and Results

Table 7: Modeling the propensity of students providing valid contact information.

Variable	Reference Category	Estimated	P-Value
Gender	female		
male		-0.152	0.000***
not specified		1.179	0.009**
Nationality	German		
foreign		-0.198	0.003**
unknown		-0.498	0.279
Type of institution	university		
Fachhochschule		0.067	0.047*
not specified/abroad		0.292	0.000***
Year of birth	1989 or earlier		
1990 - 1995		-0.057	0.049*
not specified		-1.187	0.000***
Intended degree	Bachelor		
Staatsexamen		0.154	0.004**
Lehramt		0.324	0.000***
other, unknown		-0.412	0.000
Type of contact (WS 2010/11)	personal		
postal		0.758	0.000***
Number of cases	26,913		

Notes: (i) The calculations were performed by *infas - Institut für angewandte Sozialwissenschaften GmbH*. (ii) Among the 31,082 first-year students who could be contacted, only 26,913 students provided any information on the variables considered in this model. We assume no selection bias by omitting the set of students with invalid or partial information. Nonetheless, at a later stage we use poststratification to correct for potential bias.

Table 8: Modeling participation in Wave 1 (Study B52).

Variable	Reference Category	Estimated	P-Value
Gender	female		
male		-0.109	0.040*
not specified		0.072	0.937
Nationality	German		
foreign		-0.732	0.000***
unknown		-0.826	0.413
Type of institution	university		
Fachhochschule		-0.136	0.030*
not specified/abroad		-0.580	0.393
Year of birth	1989 or earlier		
1990 - 1995		-0.007	0.896
not specified		0.171	0.724
Intended degree	Bachelor		
Staatsexamen		0.018	0.855
Lehramt		0.093	0.161
other, unknown		-0.256	0.196
Type of contact (WS 2010/11)	personal		
postal		0.382	0.000***
Instrument	CATI		
without telephone number		0.080	0.172
Attempts to contact target	1 to 3 attempts		
4 to 6 attempts		0.136	0.092
7 to 10 attempts		0.083	0.443
More than 10 attempts		-2.189	0.000***
Number of cases	18,030		

Notes: (i) The calculations were performed by *infas - Institut für angewandte Sozialwissenschaften GmbH*. (ii) Among the 21,438 first-year students who could be contacted, only 18,030 students provided valid information on the variables considered in this model. We assume no selection bias by omitting the set of students with invalid or partial information. Nonetheless, at a later stage we use post-stratification to correct for potential bias.

Table 9: Categorization of study fields (in German) used in the nonresponse models of the Waves 1 to 8.

Category	Field of Study
Field 1	*Erziehungswissenschaften, Außereuropäische Sprach- und Kulturwissenschaften, Germanistik (Deutsch, germanische Sprachen ohne Anglistik), Philosophie, Evang. Theologie, -Religionslehre, Sonderpädagogik, Anglistik, Amerikanistik, Geschichte, Romanistik, Kulturwissenschaften i.e.S., Sprach- und Kulturwissenschaften allgemein, Psychologie, Altphilologie (klass. Philologie), Neugriechisch, Slawistik, Baltistik, Finno-Ugristik, Kath. Theologie, -Religionslehre, Bibliothekswissenschaft, Dokumentation, Allgemeine und vergleichende Literatur- und Sprachwissenschaft*
Field 2	*Sport, Sportwissenschaft*
Field 3	*Sozialwesen, Wirtschaftswissenschaften, Wirtschaftsingenieurwesen mit wirtschaftswiss. Schwerpunkt, Rechts-, Wirtschafts- und Sozialwissenschaften allgemein, Sozialwissenschaften, Rechtswissenschaften, Politikwissenschaften, Regionalwissenschaften, Verwaltungswissenschaften*
Field 4	*Mathematik, Informatik, Pharmazie, Biologie, Geographie, Geowissenschaften (ohne Geographie), Mathematik, Naturwissenschaften allgemein, Physik, Astronomie, Chemie*
Field 5	*Veterinärmedizin*
Field 6	*Gesundheitswissenschaften allgemein, Humanmedizin (ohne Zahnmedizin), Zahnmedizin*
Field 7	*Ernährungs- und Haushaltswissenschaften, Landespflege, Umweltgestaltung, Agrarwissenschaften, Lebensmittel- und Getränketechnologie, Forstwissenschaft, Holzwirtschaft*
Field 8	*Maschinenbau/Verfahrenstechnik, Verkehrstechnik, Nautik, Bauingenieurwesen, Elektrotechnik, Vermessungswesen, Wirtschaftsingenieurwesen mit ingenieurwiss. Schwerpunkt, Bergbau, Hüttenwesen, Architektur, Innenarchitektur, Ingenieurwesen allgemein, Raumplanung*
Field 9	*Kunst, Kunstwissenschaft allgemein, Darstellende Kunst, Film und Fernsehen, Theaterwissenschaft, Musik, Musikwissenschaft, Gestaltung, Bildende Kunst*

Table 10: Modeling participation in Wave 1b (Study B53).

Variable	Reference Category	Estimated	P-Value
University	no		
yes		0.0816	0.3888
Gender	female		
male		-0.2648	<0.0001***
Teacher Education	no		
yes		0.0666	0.2612
Funding	private		
public		0.4587	0.1027
Field of Study (see Table 9)	Field 1		
Field 2		-0.0390	0.5917
Field 3		0.0071	0.9462
Field 4		-0.0731	0.3516
Field 5		0.3803	0.0003***
Field 6		-1.4079	<0.0001***
Field 7		0.2020	0.4012
Field 8		-0.0339	0.7833
Field 9		-0.3326	0.0102*
Nontraditional Admission	no		
yes		-0.6263	<0.0001***
missing		-0.1940	0.2260
Own Children	no		
yes		-0.8081	<0.0001***
Household Size	one person		
two persons		-0.1391	0.0083**
more than two persons		-0.2378	<0.0001***
Region	East		
West		-0.2156	0.0001***
Educational Attainment Mother (CASMIN)	1a, 1b, 2b		
1c, 2a		0.1850	0.0394*
2a		0.2081	0.0484*
3a, 3b		0.1829	0.0679.
missing		-0.0747	0.7784
Educational Attainment Father (CASMIN)	1a, 1b, 2b		
1c, 2a		0.0001	0.9991
2a		0.0435	0.6927
3a, 3b		0.0749	0.5215
missing		-0.1409	0.2722
Birth Year	<1989		
1989/90		0.1682	0.0007***
>1990		0.2232	0.0004***
School-leaving Qualification	no *Abitur*		
German *Abitur*		0.1606	0.4745
Nongerman *Abitur*		0.2790	0.2830
missing		-0.8308	0.0002***
Migration Background	Generation Status ≥ 3		
Generation Status < 3		-0.1464	0.0027**
Number of cases	17,910		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 11: Modeling participation in Wave 2 (Study B54).

Variable	Reference Category	Estimated	P-Value
University	no		
yes		0.1450	0.0016**
Gender	female		
male		-0.1873	<0.0001***
Teacher Education	no		
yes		-0.1473	<0.0001***
Funding	private		
public		-0.3590	0.0036**
Field of Study	Field 1		
Field 2		-0.1121	0.1624
Filed 3		-0.0037	0.9540
Filed 4		0.0026	0.9543
Filed 5		-0.0869	0.4394
Filed 6		-0.3134	0.0002***
Filed 7		-0.1588	0.1003
Filed 8		-0.0074	0.8697
Filed 9		0.0807	0.4336
Nontraditional Admission	no		
yes		0.2287	0.0404*
missing		-0.0764	0.6318
Own Children	no		
yes		0.0493	0.6372
Reading Competence Wave 1	low		
Lower medium		0.0980	0.2207
Upper medium		0.3521	0.0001***
high		0.3537	0.0001***
missing		-0.8993	<0.0001***
Household Size	one person		
two persons		-0.0152	0.7757
more than two persons		-0.2084	<0.0001***
Region	East		
West		-0.1235	0.0039**
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		0.0334	0.6268
2c		0.0446	0.5531
3a, 3b		0.0291	0.7500
missing		0.0919	0.2296
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		0.1067	0.1755
2c		0.0407	0.6717
3a, 3b		0.1127	0.2401
missing		-0.0353	0.7007
Birth Year	<1989		
1989/90		0.0269	0.5638
>1990		0.0975	0.0513.
School-leaving Qualification	no *Abitur*		
German *Abitur*		0.0693	0.5424
Nongerman *Abitur*		-0.1702	0.3193
missing		-0.8449	<0.0001***
Migration Background	Generation Status $\geq$ 3		
Generation Status < 3		-0.1194	0.0698.
Number of cases	17,910		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 12: Modeling participation in Wave 3 (Study B55).

Variable	Reference Category	Estimated	P-Value
Participation in B54	no		
yes		0.5761	<0.0001***
University	no		
yes		0.0811	0.1660
Gender	female		
male		0.1035	0.0062**
Teacher Education	no		
yes		0.1395	<0.0001***
Funding	private		
public		-0.0529	0.6653
Field of Study	Field 1		
Field 2		-0.0133	0.9247
Field 3		-0.0128	0.8212
Field 4		0.0630	0.2453
Field 5		0.0854	0.0669.
Field 6		0.4960	<0.0001***
Field 7		-0.0035	0.9853
Field 8		0.1137	0.1057
Field 9		-0.1677	0.1430
Nontraditional Admission	no		
yes		0.0355	0.8136
missing		-0.1242	0.4274
Own Children	no		
yes		0.2785	0.0056**
Reading Competence Wave 1	low		
Lower medium		0.0509	0.6991
Upper medium		-0.1281	0.2242
high		0.0130	0.9864
missing		-0.3675	0.0001***
Household Size	one person		
two persons		-0.0533	0.3719.
more than two persons		0.0759	0.0765.
Region	East		
West		-0.0176	0.7268
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		-0.0549	0.3422
2c		0.0050	0.9313
3a, 3b		-0.0630	0.5205
missing		0.0131	0.8755
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		0.1801	0.0720.
2c		0.1683	0.0802.
3a,3b		0.1305	0.2834
missing		0.2133	0.0827.
Birth Year	<1989		
1989/90		0.0269	0.6313
>1990		0.062	0.9294
School-leaving Qualification	no *Abitur*		
German *Abitur*		-0.2841	0.0708.
Nongerman *Abitur*		-0.6115	0.0702.
missing		-2.3560	<0.0001***
Migration Background	Generation Status $\geq$ 3		
Generation Status < 3		-0.0776	0.0944.
Number of cases	17,910		

Table 13: Modeling participation in Wave 4 (Study B56).

Variable	Reference Category	Estimated	P-Value
Participation in previous waves	always		
often		-1.5326	<0.0001***
seldom		-3.2581	<0.0001***
University	no		
yes		0.0120	0.8585
Gender	female		
male		-0.1264	0.0012**
Teacher Education	no		
yes		0.1094	0.0111*
Funding	private		
public		-0.2105	0.0212*
Field of Study	Field 1		
Field 2		-0.1889	0.0626.
Field 3		-0.1290	0.0961.
Field 4		-0.0492	0.4718
Field 5		0.0192	0.7942
Field 6		0.0346	0.6842
Field 7		-0.1377	0.1855
Field 8		-0.1059	0.1642
Field 9		0.2849	0.0321*
Nontraditional Admission	no		
yes		0.2372	0.0778.
missing		-0.1355	0.4843
Own Children	no		
yes		0.0118	0.9266
missing		0.2597	0.2839
Reading Competence Wave 1	low		
Lower medium		0.2067	0.0126*
Upper medium		0.3782	<0.0001***
high		0.4323	<0.0001***
missing		-0.1689	0.0088**
Household Size	one person		
two persons		0.0152	0.6378
more than two persons		-0.1147	0.0142*
missing		0.5592	0.0297*
Region	East		
West		-0.1231	0.0251*
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		0.0022	0.9768
2c		0.0428	0.6338
3a, 3b		-0.1384	0.2148
missing		0.0799	0.3770
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		0.2088	0.0128*
2c		0.1686	0.0388*
3a, 3b		0.2688	0.0119*
missing		0.1423	0.0508.
Birth Year	<1989		
1989/90		0.1068	0.0315*
>1990		0.1365	0.0119*
School-leaving Qualification	no *Abitur*		
German *Abitur*		0.0961	0.6314
Nongerman *Abitur*		-0.0872	0.8002
missing		-0.8960	<0.0001***
Migration Background	Generation Status $\geq$ 3		
Generation Status < 3		-0.0092	0.8859
Number of cases	17,910		

Table 14: Modeling participation in Wave 5 (Study B59).

Variable	Reference Category	Estimated	P-Value
Participation in previous waves	always		
often		-0.5395	<0.0001***
seldom		-1.3736	<0.0001***
University	no		
yes		0.0104	0.9061
Gender	female		
male		0.1475	<0.0001***
Teacher Education	no		
yes		0.2326	<0.0001***
Funding	private		
public		-0.0292	0.8752
Field of Study	Field 1		
Field 2		0.1046	0.2492
Field 3		0.0242	0.7984
Field 4		0.1870	0.0385*
Field 5		0.5175	0.0007***
Field 6		0.0454	0.7009
Field 7		-0.0667	0.5658
Field 8		0.0612	0.5493
Field 9		0.0404	0.7592
Nontraditional Admission	no		
yes		0.0152	0.8418
missing		-0.2329	0.1144
Own Children	no		
yes		0.0243	0.8647
missing		0.3348	0.1893
Reading Competence Wave 1	low		
Lower medium		0.0174	0.8539
Upper medium		0.1385	0.1706
high		-0.0154	0.9181
missing		-0.5520	<0.0001***
Household Size	one person		
two persons		0.1475	0.0427*
more than two persons		0.3240	<0.0001***
missing		-1.6443	<0.0001***
Region	East		
West		0.1665	0.0031**
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		0.1958	0.0517.
2c		0.1853	0.0815.
3a, 3b		0.2538	0.0932.
missing		0.1622	0.0351*
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		-0.0217	0.8806
2c		-0.0633	0.6699
3a, 3b		0.0205	0.8888
missing		-0.0676	0.6259
Birth Year	<1989		
1989/90		0.0808	0.1301
<1990		0.0957	0.0725.
Migration Background	Generation Status $\geq$ 3		
Generation Status < 3		-0.1649	0.0045**
Number of cases	17,910		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 15: Modeling participation in Wave 6 (Study B58).

Variable	Reference Category	Estimated	P-Value
Participation in previous waves	always		
often		-1.4644	<0.0001***
seldom		-3.0676	<0.0001***
University	no		
yes		0.0511	0.5077
Gender	female		
male		-0.0106	0.7438
Teacher Education	no		
yes		0.0011	0.9817
Funding	private		
public		-0.2627	0.0008***
Field of Study	Field 1		
Field 2		-0.2898	0.0210*
Field 3		-0.1611	0.0505.
Field 4		-0.0243	0.7432
Field 5		0.1341	0.2383
Field 6		0.1865	0.3643
Field 7		-0.1508	0.1324
Field 8		-0.0973	0.3109
Field 9		0.0206	0.8224
Nontraditional Admission	no		
yes		0.0848	0.5504
missing		-0.2419	0.1677
Own Children	no		
yes		-0.3382	0.0147*
missing		-0.1300	0.2759
Reading Competence Wave 1	low		
Lower medium		0.1833	0.0262*
Upper medium		0.1605	0.0778.
high		0.3466	<0.0001***
missing		-0.2524	0.0001***
Household Size	one person		
two persons		0.0145	0.8230
more than two persons		-0.0891	0.0561.
missing		-1.2181	0.1584
Region	East		
West		-0.1015	0.0848.
Educational Attainment Mother	1a, 1b, 2a		
1c, 2a		-0.0245	0.8045
2c		-0.0896	0.4410
3a, 3b		-0.0543	0.5921
missing		-0.0029	0.9725
Educational Attainment Father	1a, 1b, 2a		
1c, 2a		0.1090	0.3309
2c		0.0505	0.6795
3a, 3b		0.1929	0.1819
missing		0.1419	0.1769
Birth Year	<1989		
1989/90		0.1656	0.0004***
>1990		0.1619	0.0042**
School-leaving Qualification	no *Abitur*		
German *Abitur*		-0.1011	0.5906
Nongerman *Abitur*		-0.2011	0.4672
missing		1.1128	0.1757
Migrant Background	Generation Status $\geq$ 3		
Generation Status < 3		-0.1022	0.0514.
Number of cases	17,910		

Table 16: Modeling participation in Wave 7 (Study B94) .

Variable	Reference Category	Estimated	P-Value
Participation in previous waves	always		
often		-0.4083	<0.0001***
seldom		-1.2833	<0.0001***
University	no		
yes		0.1197	0.0104*
Gender	female		
male		0.0467	0.1740
Teacher Education	no		
yes		-1.7469	<0.0001***
Funding	private		
public		-0.0336	0.6587
Field of Study	Field 1		
Field 2		0.1951	0.2679
Field 3		-0.0369	0.4758
Field 4		0.0600	0.0226*
Field 5		0.2596	0.0055**
Field 6		0.2556	0.2614
Field 7		-0.0822	0.1843
Field 8		0.0518	0.3166
Field 9		-0.1280	0.1131
Nontraditional Admission	no		
yes		-0.2095	0.0604.
missing		-0.9709	<0.0001***
Kids in Household	no		
yes		0.0116	0.9294
missing		-0.1236	0.5597
Reading Competence Wave 1	low		
Lower Medium		-0.0821	0.4429
Upper Medium		0.0209	0.8371
high		0.0527	0.5626
missing		-0.1905	0.0003***
Household Size	one person		
two persons		0.1057	0.1281
more than two persons		0.1386	0.0153**
missing		0.7597	0.1765
Region	East		
West		0.0148	0.8374
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		-0.0143	0.8451
2c		-0.0051	0.9535
3a, 3b		0.0870	0.3604
missing		-0.0398	0.5903
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		0.1334	0.1262
2c		0.1413	0.1627
3a, 3b		0.1654	0.0827.
missing		0.2246	0.0058**
Birth Year	<1989		
1989/90		0.0446	0.4536
>1990		-0.0036	0.9507
School-leaving Qualification	no *Abitur*		
German *Abitur*		-0.4353	0.0003***
Nongerman *Abitur*		0.3033	0.2728
missing		-2.8734	<0.0001***
Migration Background	Generation Status ≥ 3		
Generation Status < 3		0.0269	0.6794

Table 17: Modeling participation in Wave 8 (Study B95).

Variable	Reference Category	Estimated	P-Value
Participation in previous waves	always		
often		-1.2525	<0.0001 ***
seldom		-3.3042	<0.0001***
University	no		
yes		0.1405	0.0258*
Gender	female		
male		-0.1214	0.0001***
Teacher Education	no		
yes		0.3910	<0.0001***
Funding	private		
public		-0.0334	0.7279
Field of Study	Field 1		
Field 2		0.0597	0.6639
Field 3		-0.1444	0.0374*
Field 4		-0.0175	0.6738
Field 5		0.2220	0.0170*
Field 6		0.5067	0.0022**
Field 7		-0.2829	0.0003***
Field 8		-0.1549	0.0298*
Field 9		-0.0384	0.5982
Nontraditional Admission	no		
yes		0.1297	0.3257
missing		-0.2991	0.0966.
Kids in Household	no		
yes		-0.0737	0.5583
missing		-0.0905	0.4620
Reading Competence Wave 1	low		
Lower medium		0.2552	0.0013**
Upper medium		0.2296	0.0016**
high		0.3693	<0.0001***
missing		-0.0258	0.6974
Household Size	one person		
two persons		0.0096	0.9006
more than two persons		-0.1533	0.0471*
missing		-0.1442	0.3670
Region	East		
West		-0.0504	0.2001
Educational Attainment Mother	1a, 1b, 2b		
1c, 2a		0.1841	0.0411*
2c		0.2323	0.0217*
3a, 3b		0.3883	0.0008***
missing		0.3148	0.0010**
Educational Attainment Father	1a, 1b, 2b		
1c, 2a		-0.0465	0.7261
2c		-0.0518	0.7062
3a, 3b		0.0324	0.8259
missing		0.0303	0.8261
Birth Year	<1989		
1989/90		0.1272	0.0031**
>1990		0.1622	<0.0001***
School-leaving Qualification	no *Abitur*		
German *Abitur*		0.2256	0.1535
Nongerman *Abitur*		0.3850	0.1630
missing		-0.0941	0.5777
Migration Background	Generation Status $\geq$ 3		
Generation Status <3		-0.1883	0.0001***
Number of cases	17,910		