Andreas Daniel / Björn Huß / Percy Scheller

DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

Data and Methods Report on the Surveys of School Leavers with a Higher Education Entrance Qualification from the 2008 cohort (1st to 3rd survey wave)

Data and Methods Report

February 2017

fdz.DZHW
Research data centre for higher education research and science studies

SUBSIDISED BY THE
Federal Ministry of Education and Research

Author(s):
Andreas Daniel
Björn Huß
Percy Scheller

With assistance from:
Ute Hoffstätter (Final Editing)
Dr. Anne Weber (Final Editing)

Editor:
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)
German Centre for Higher Education Research and Science Studies (DZHW)
Lange Laube 12 | 30159 Hannover | Germany | www.dzhw.eu
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | info@dzhw.eu

Management board:
Prof. Dr. Monika Jungbauer-Gans
Karen Schlüter

Chairman of the supervisory board:
Ministerialdirigent Peter Greisler

Registration court:
Amtsgericht Hannover | B 210251

Table of Contents

Table of Figures

List of Tables

I Introduction

The DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification represents a series of studies into the training and study decisions and training and study procedures for school leavers with a general higher education entrance qualification or an advanced technical college certificate.[1] They are carried out by the German Centre for Higher Education Research and Science Studies (DZHW)[2], financed by the Federal Ministry of Education and Research (BMBF) and facilitate – in addition to the official university statistics – the national educational monitoring and the response to questions posed by the higher education research and science studies. New cohorts of school leavers with a higher education entrance qualification have been surveyed every one to three years since 1976.

Within the scope of the BMBF funded project on the development of a Research Data Centre at the DZHW (RDC-DZHW), the data contained on some of the school leavers cohorts are subsequently processed and documented for the purpose of later use.[3] Using various modes of access, they are made available as *Scientific Use Files* (SUF) for secondary scientific use. Along with the survey data sets, documentation material on the data sets and the implementation of the studies are provided.[4]

This data and methods report is part of the documentation from the first to third survey waves of the 2008 cohort of school leavers with a higher education entrance qualification (doi: 10.21249/DZHW:gsl2008:1.0.0). Further documentation material on the study (data set reports, questionnaires, question flow diagrams etc.) can be downloaded from the search portal of the RDC-DZHW (https://metadata.fdz.dzhw.eu/#!/en).

Section II of this report presents an overview of the key data from the study. Key information on the use of the data follows in Section III. Chapter 1 introduces the content and structure of the DZHW Panel Series of School Leavers with a Higher Education Entrance Qualification up to 2008 in general, and the 2008 Panel Study in particular. The remaining chapters of the report orient themselves to stages of the research process. In Chapter 2, the applied survey instruments are described, and the survey implementation process (e.g. sampling procedure, survey operation, response, data preparation etc.) is detailed in Chapters 3-6. In Chapters 7 and 8, weighting and anonymisation practices used are presented.

1 Current updates on the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification can be found on the project website (http://www.dzhw.eu/projekte/pr_show?pr_id=465).

2 The German Centre for Higher Education Research and Science Studies (DZHW, http://www.dzhw.eu) was founded in August 2013 as a spin-off of HIS Hochschul-Informations-System GmbH. Throughout the following text, the term DZHW is used, even if the study was carried out before 2013.

3 At the time of data collection, no subsequent data use was planned. Some information on the survey was not documented with a focus on subsequent data use and may in part no longer be reconstructed. This is indicated at the corresponding points in the text.

4 Information on the available data sets and documentations is provided on the RDC website (https://fdz.dzhw.eu).

II Overview of the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

Survey Series	DZHW Survey Series of School Leavers
Cohorts	School Leavers Cohort 2008 (17th cohort in the School Leavers Panel Survey Series)
Surveying Institution	German Centre for Higher Education Research and Science Studies (DZHW)
Funding	Federal Ministry of Education and Research (BMBF)
Project Contributors (Project Leaders)	1st wave: Heike Spangenberg, Heiko Quast 2nd wave: Heiko Quast, Christoph Heine, Julia Willich 3rd wave: Heiko Quast, Markus Lörz, Percy Scheller
Themes	Vocational and higher education intentions Educational decisions and biographies Choice of engineering studies Financing of studies/tuition fees
Survey Design	Cohort Panel Design
Population	School leavers from general schools and vocational schools in the 2007/2008 academic year with a higher education entrance qualification from the Federal Republic of Germany
Sampling	Simple (disproportionally) stratified random sampling
Survey Method	Standardised self-administered survey
Survey Period	1st wave: December 2007 to August 2008 2nd wave: December 2008 to September 2009 3rd wave: December 2012 to December 2013
Number of Cases (Data Set)	1st wave: n = 28,182 2nd wave: n = 5,933 3rd wave: n = 3,671

Response Rate	1st wave: 48.9 % 2nd wave:33.8 % 3rd wave: 61.9 %
Data Products and Mode of Access	SUF: Remote-Desktop, On-Site
Data Set Structure	Individual data in wide-format
DOI	10.21249/DZHW:gsl2008:1.0.0
Further information	https://fdz.dzhw.eu

Project Publications*

1st wave: Heine, Christoph; Quast, Heiko (2009): Studierneigung und Berufsausbildungspläne. Studienberechtigte 2008 ein halbes Jahr vor Schulabgang. HIS. Hannover (HIS: Forum Hochschule, 4/2009)

2nd wave: Heine, Christoph; Quast, Heiko; Beuße, Mareike (2010): Studienberechtigte 2008 ein halbes Jahr nach Schulabschluss. Übergang in Studium, Beruf und Ausbildung. HIS. Hannover (HIS: Forum Hochschule, 3/2010)

3rd wave: Quast, Heiko; Scheller, Percy; Lörz, Markus (2014): Bildungsentscheidungen im nachschulischen Verlauf. Dritte Befragung der Studienberechtigten 2008 viereinhalb Jahre nach Schulabschluss. HIS. Hannover (HIS: Forum Hochschule, 9/2014)

* All project publications are available for download on the project website (http://www.dzhw.eu/projekte/pr_show?pr_id=270).

Publications using the Data Set (selected)

Lörz, Markus; Quast, Heiko; Roloff, Jan (2015): Konsequenzen der Bologna-Reform: Warum bestehen auch am Übergang vom Bachelor- ins Masterstudium soziale Ungleichheiten? In *Zeitschrift für Soziologie* 44 (2), pp. 137–155

Baier, Tina; Helbig, Marcel (2014): Much ado about € 500: do tuition fees keep German students from entering university? Evidence from a natural experiment using DiD matching methods. In *Educational Research and Evaluation: An International Journal on Theory and Practice* 20 (2), pp. 98–121. DOI: 10.1080/13803611.2014.881745

Helbig, Marcel; Baier, Tina; Kroth, Anna J. (2012): Die Auswirkung von Studiengebühren auf die Studierneigung in Deutschland. Evidenz aus einem natürlichen Experiment auf Basis der HIS-Studienberechtigtenbefragung. In *Zeitschrift für Soziologie* 41 (3), pp. 227–246

Lörz, Markus (2012): Mechanismen sozialer Ungleichheit beim Übergang ins Studium: Prozesse der Status- und Kulturreproduktion. In Rolf Becker, Heike Solga (Eds.): Soziologische Bildungsforschung. Wiesbaden: VS Verlag für Sozialwissenschaften (Kölner Zeitschrift für Soziologie und Sozialpsychologie: Sonderheft, 52), pp. 302–324

III Data Use Instructions

[Data Use Requirements] Data from the DZHW Panel Study of School Leavers are anonymised and made available by the RDC-DZHW in accordance with Federal Data Protection Law (cf. § 40 paras. 1 and 2 BDSG) exclusively for scientific research purposes.[5] The RDC provides a *Scientific Use File* (SUF) for scientific secondary use.

Requirements for the use of a SUF are an employment at a scientific institution and the conclusion of a data use agreement. Before the conclusion of a data use agreement, the RDC verifies the presence of a scientific use purpose. Students or doctoral students without an employment at a scientific institution must be able to prove cooperation with a supervisory employee of a scientific institution. A form for the data use contract can be downloaded from the RDC website.

[Data Access] The SUF is provided using two modes of access, which differ in their restrictions with respect to storage location, the opportunity for autonomous access to external data and RDC control options for restrictive data. These methods include:

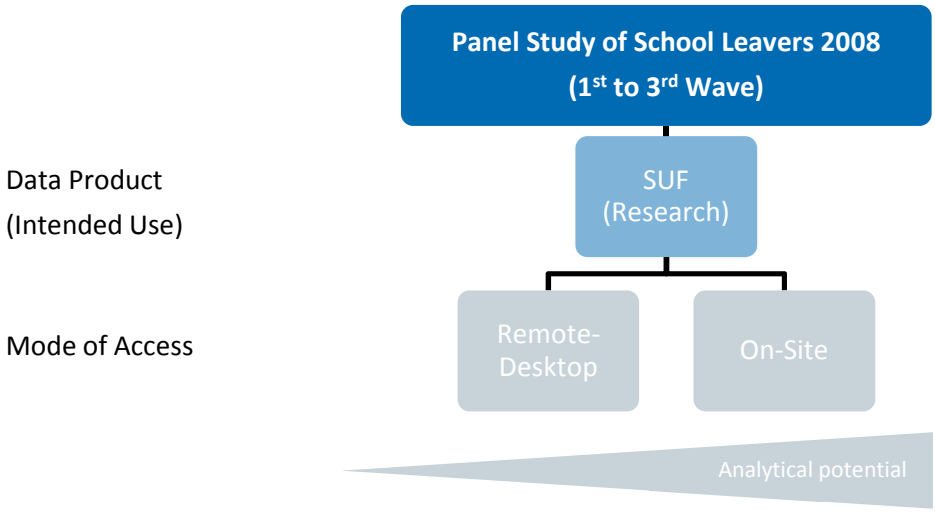
- **Remote Desktop:** Data are available on a RDC terminal server. Using a secure connection between the user's local computer and the RDC terminal server, the data can be analysed using the software on the terminal server. The transfer of data to the local computer is not possible. Analysis results are made available only after a data protection clearance test by the RDC.
- **On-Site:** Data are made available for analysis at a secure computer on RDC premises and in a controlled environment. As with remote desktop access the analysis results are made available only after a data protection clearance test by the RDC.

The extent of information access from the data made available differs according to the mode of access, which further impacts analytical potential (cf. Figure 1). More detailed information is made available for data users in accordance with the degree of restrictions governing the user's data access through technical and organisational measures.[6] Such procedures ensure the highest degree of usability, and simultaneously, the best possible data protection.

[5] The RDC's data protection policy is based on the portfolio approach of Lane et al. 2008, 6ff., on upon which the Leibniz Institute for Educational Trajectories (LIfBi) (cf. Koberg 2016, 699ff.) and the RDC of the Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner et al. 2012, 9f.) have oriented themselves. The RDC-DZHW has adapted the portfolio approach to the requirements of its own data files and uses four categories of measures in securing data protection, which are combined in various ways: legal-institutional measures, informational measures, technical measures and statistical measures.

[6] Cf. Chapter 8 on the various levels of anonymisation and analytical potential of the differing SUF variants.

Figure 1: Modes of Access and Analytical Potential



[Data Products] With the *Digital Object Identifier* (DOI) (10.21249/DZHW:gsl2008:1.0.0) central information on the study, further documentation materials and an overview of available data products from the study can be found on the website.

The data from the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008 are available for on-site or remote desktop work with access-specific analytical potential in each case (see Figure 1). The data are provided in wide format (see Chapter 6.5). The data set is issued in Stata format as standard.

[Charges for Data Access] Currently SUF and CUF are available free of charge (effective June 2017). The present fees regulation can be found on the RDC website (https://fdz.dzhw.eu).

[Responsibilities of Data Users] Data users are obliged to observe the following rules[7]:

- **Scientific Use:** Data must be used exclusively for scientific research purposes. Commercial use is forbidden.
- **De-anonymisation forbidden:** Any attempt of re-identification for the units of analysis (e.g. persons, households, institutions) is prohibited.
- **Duty to report security loopholes:** If data users become aware of security loopholes with respect to data protection or data security, the RDC should be informed immediately.
- **No data disclosure:** SUF may only be used by persons who have made a data use contract.
- **Duty to delete:** SUF downloads must be deleted after expiry of the agreed period of use (as a rule three years) from all computers, servers and data storage devices. Likewise all backup copies, modified data sets (e.g. work-, excerpt- or help-data) as well as print-outs must be destroyed.

[7] The data use contract regulates terms and conditions of use in detail.

- **Notification/Provision of Publications:** The RDC has to be notified of all types of publications that are produced using data of the RDC. An electronic version of the publication shall be provided immediately. A list of existing publications based on the data can be found in the Metadata Search Portal.[8]
- **Citation rules**: The data set used must be cited according to the following requirements in publications, other essays (e.g. final dissertations) and presentations.

[Citation]

- **Data Set:**
  Heine, C., Quast, H., Spangenberg, H., Lörz, M., Scheller, P. & Willich, J. (2014). *DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008*. Version 1.0.0. doi: 10.21249/DZHW:gsl2008:1.0.0., DATA SET NAME[9], Hannover: RDC-DZHW.
- **Data and Methods Report:**
  Daniel, A., Huß, B., Scheller, P. (2016). *DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008: Data and methodological report on the surveys of school leavers with a higher education entrance qualification from the 2008 cohort (1st to 3rd survey wave).* Version 1.0.0. doi: 10.21249/DZHW:gsl2008:1.0.0. Hannover: RDC-DZHW.

In addition, the following formulation should be used in the text to make reference to the data used:

> *"This scientific work uses data of the Panel Study of School Leavers 2008, conducted by the German Centre for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the Research Data Centre of the DZHW, doi: doi:10.21249/DZHW:gsl2008:1.0.0."*

[8] https://metadata.fdz.dzhw.eu/#!/en

[9] Please provide the exact name of the dataset version used at this point, e.g. gls2008_r_1_0_0 the remote desktop SUF from the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008.

1 Content and Design of the Study

[Survey Series] The DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008 is part of the survey series on school leavers conducted by the DZHW, in which information is gathered on the post-school career paths of school leavers with higher education entrance qualifications by means of standardised panel surveys.[10] The survey series begins with the survey on the 1976 cohort of people with higher education entrance qualifications in the former BRD; since then, cohorts have been surveyed every two or three years. The school leavers in the new German federal states were included for the first time in 1990. The target populations of all the DZHW panel studies of school leavers with a higher education entrance qualification are school leavers from general schools and vocational schools with a 'Hochschulreife' (higher education entrance qualification) from the Federal Republic of Germany.

As a rule[11], multiple survey waves are conducted for each cohort of school leavers at various times before and after the obtaining of their higher education entrance qualification, i.e. in a combined cohort/panel design. Table 1 highlights the conceptual changes. Up to 1986, the DZHW pursued the aim of surveying every cohort of school leavers with higher education entrance qualifications shortly after leaving school and every 2½, 4½ and 12½ years thereafter. From 1990 onwards, the survey on belated or amended educational and study decisions two and a half years after leaving school was dropped.[12] Since the 2005 cohort of school leavers with a higher education entrance qualification, the cohorts have also been surveyed six months *before* leaving school and thus four survey waves have been carried out.

[10] Current information on the Panel Study of School Leavers with a Higher Education Entrance Qualification and the available Scientific user Files from the series of surveys can be found at https://fdz.dzhw.eu.

[11] The exceptions to this were the cohorts of 1991, 1993, 1996 and 2004, who were only surveyed once.

[12] Another exception is the additional survey wave from the 1990 cohort of people with a higher education entrance qualification, which was carried out in cooperation with the Federal Institute for Vocational Education and Training (BIBB) and which took place five and a half years after the respondents had left school.

Table 1: Structure and Conceptual Changes in the DZHW Panel Study of School Leavers from 1976 to 2008

Time of Survey	Thematic Focus	School Leavers Cohort		
		up to 1986[b]	1990 to 2004	since 2005
6 months *before* leaving school	Educational and study *intentions*	-	-	1st wave
6 months *after* leaving school	Educational and study *decisions*	1st wave	1st wave	2nd wave
2½ years after leaving school	Belated or amended educational and study *decisions*	2nd wave	-	-
3½ or 4½ years after leaving school	Educational and study *biographies*	3rd wave	2nd wave	3rd wave
10½ to 20½ years after leaving school	Entering the labour market and career paths	4th wave	3rd wave	4th wave

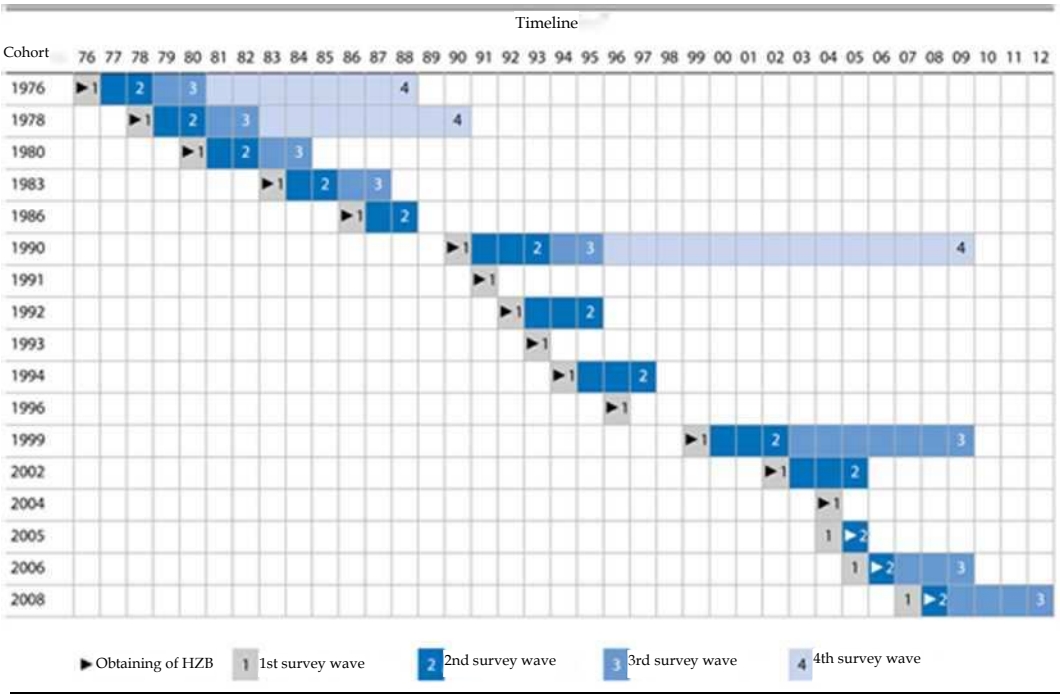
[a] All of the survey waves could not be realised for some of the cohorts (see Figure 2).

[b] The surveys for the cohorts from 1976 to 1986 focussed exclusively on people with higher education entrance qualifications from the states of the former Federal Republic of Germany.

Source: Quast, Spangenberg (2008), own adaptation

The various surveys were conducted as written/postal paper-and-pencil questionnaires. The survey tools for all cohorts contained questions on educational aspirations, post-school career path, degree course, the transition into professional life and sociodemographic and educational biographic characteristics. The thematic focus of each survey wave focuses on the present educational, professional and life stage of the respondents at the time of the survey. Spanning a period of almost 40 years up to the present survey, the result was a time series of 17 cohorts with a total of 41 survey waves (see Figure 2).

Figure 2: Cohort Panel Design of the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification



Note: The surveys for the cohorts from 1976 to 1986 focussed exclusively on people with higher education entrance qualifications from the states of the former Federal Republic of Germany.

Source: Schneider, Franke (2014); own adaptation

[Analytical Potential] In all cohorts, an identical core set of information is collected per survey wave. On this basis, long-term trends can be considered in the transition from school to the further education system and the workplace using time series or cohort comparisons. It should be stressed that in all of the surveys of school leaver cohorts which took place more than six months after the respondents obtained their higher education entrance qualification, continuous historical data on the individual's activities since leaving school was gathered on a half-yearly or monthly basis, which is suitable for event data analyses and sequence pattern analyses. In principle, the thematic focus varies from wave to wave within a survey cohort. Specific questions were repeatedly asked, however, so that the responses from the same respondents from different waves could be directly linked with one another and intra-individual changes over the course of time taken into account. Therefore, on the basis of the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification, cross-sectional, trend, event data, sequence pattern and limited panel causality analyses can be performed. The data from the panel study of school leavers is normally weighted using the characteristics of sex, federal state, type of school and higher education entrance qualification and adapted for the population. Panel-typical default processes are also taken into account during weighting.

[Research Field] The design of the random sample and the survey, as well as the associated analysis possibilities distinguish the DZHW Panel Study from other surveys carried out in Germany on people with higher education entrance qualifications. No other survey enables nationwide

analyses. Furthermore, other surveys in this field have either significantly shorter time series or none at all. Other school leaver surveys include, for example, the Saxony School Leaver Survey[13] (conducted by the Centre of Excellence for Education and Higher Education Planning at TU Dresden), die TOSCA Study[14] (now conducted by the Hector Institute for Empirical Research in Education at the University of Tübingen,) and the Berliner-Studienberechtigtenpanel (Berlin School Leavers Panel Survey) *Best Up*[15] (conducted by the German Institute for Economic Research [DIW] and the Berlin Social Science Center [WZB]).

[Particularities of the Panel Study of School Leavers 2008] Along with the general characteristics of the survey series, the survey of the 2008 cohort exhibits the following specifics. As in the pilot study with the 2005 cohort or the survey of the 2006 cohort, the educational and study aspirations of the respondents were already taken into account in the first survey wave six months before they were due to leave school. As a new feature of the 2008 cohort of people with higher education entrance qualifications the survey took place in the class for the first time and during lesson time where possible.

Furthermore, the study phase of the 2008 cohort is characterised by the introduction of tiered courses of study (bachelor and master degree) agreed on in 2002. This change was accounted for by using a detailed collection of the study-related educational decisions. The shifting of the survey time in the third wave from three and a half to four and a half years after the school leaving date also made it possible to survey a larger proportion of respondents with higher education entrance qualifications after they had been accepted onto a master's degree course (34 per cent of all persons entitled to further study in 2008 with admission to a bachelor's programme had already taken up a master's degree at this time (cf. Quast et al. 2014, p. 115). Additionally, the decision by the Federal Constitutional Court in 2005 enabled seven federal states to charge general tuition fees or contributions (from the beginning of the first degree) at the time of the first survey of the 2008 cohort of university entrants. Moreover, the survey took place against the background of the educational policy discussion on an anticipated shortfall in the number of university graduates from science and engineering subjects; this was also taken into account by way of a more detailed recording of the related motives.

[13] see also: http://www.kfbh.de/projekte/abiturientenstudie.html
[14] see also: http://www.wiso.uni-tuebingen.de/faecher/hector-institut-fuer-empirische-bildungsforschung/forschung/laufende-studien/tosca.html
[15] see also: http://www.best-up.eu

2 Survey Instruments

A standardised paper questionnaire in German was used as a survey instrument for the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008.[16] Chapter 2.1 introduces the main content from the three survey instruments.[17] Chapter 2.2 describes the pre-test carried out to review and improve the questionnaire.

2.1 Contents of the Survey Instruments

[Characteristics of the Survey Series] Just as with the other cohorts in the survey series, the description and explanation of educational decisions is a focal point of the 2008 Panel Study. The content of the instruments used for the survey of the educational processes as well as its explanation is based on – analogous to a classification from Mertens (1976) – the expected chronological occurrence of the various life stages, from the *post-school transitional phase* (1) and the *qualification phase* (2) through to the *job entry phase* (3). Along with the collection of information on educational intentions and decisions, all of the survey instruments also contain questions about attitudes and convictions to enable, in theory, well-founded explanations for the educational decisions.

The surveying of school leavers before they have left school[18] and obtained a higher education entrance qualification (1st wave) focusses on their educational and study aspirations. To this end, information was collected on personality traits, convictions regarding schooling, future plans for the period after leaving school and the relevant decision-making, and on the importance of different information channels for study and educational planning and student financing.

The educational and study decisions were once more the focus of the first post-school survey (2nd wave). Other information is collected in the process, mostly in relation to schooling as well as the person, their current activity, detailed information on the (possibly envisaged) educational steps, the reasons for their post-school career path, employment prospects and their career and life ambitions in general.

The other post-school survey approximately three to four years after leaving school (3rd wave) focusses retrospectively on all the cohorts, in particular on the realisation of their educational aims. Consideration is also given to the educational and study biographies. The time frames[19], type of activity and explanatory notes are collated using an activity table (question 2.1 in the 3rd wave of the 2008 cohort panel study). For the degree courses, questions are asked about the main study subjects, the final qualification being targeted and the name and location of the university/vocational academy. The professional title and, where appropriate, the type of training are also recorded for vocational training or a professional activity. The survey and reconstruction of the educational pathways is supplemented by the reflection on the decision-making

[16] The questionnaire can be downloaded from the RDC website. There is a also a question flow diagram showing the filter procedure for the three questionnaires.

[17] There is no information available on the origin and design of the scales used.

[18] The first survey to take place before the school leaving date was the survey of the 2005 cohort of school leavers and this has been a standard part of the panel study of school leavers ever since. The previous cohorts (1976 to 2004) were not contacted and interviewed until after leaving school (see chapter 1).

[19] The time frames were surveyed every six months at the beginning of the series of studies and on a monthly basis since the 1999 cohort.

and the survey of other educational aims, the repeated measurement of the general career and life goals and evaluations of the educational decision (e.g. the study subject).

[Particularities of the 2008 Panel Study of School Leavers] Along with the fact that the 2008 school leavers are only the third cohort in the serial survey to be interviewed before leaving school, the specific educational policy context should be highlighted in relation to the survey instruments. In conjunction with the introduction of general tuition fees and contributions, and the switch to the two-tier study system of bachelor's and master's degrees, additional questions were included for the second and third survey waves in particular. Tools for the analysis of lessons (questions 4 and 5 in wave 1 and question 1.3 in wave 3) and of personal perceptions in relation to the decision-making regarding the post-school career path (questions 16 to 19 in wave 1 and question 1.2 in wave 3) were added for the first time in the first and third waves of the 2008 panel study (in comparison with the previous cohorts of 2005 and 2006). In the second survey wave, specific questions were asked about the admission to a science or engineering degree course (questions 13 and 14 in wave 2). In the second and third survey waves, the intentions of beginning a master's course (questions 27 and 28 in wave 2) and comprehensive information on the subject of tuition fees (questions 35 to 38 in wave 2) and aborting studies (questions 6.1 in wave 3) as well as additional information on social background (questions 8.4 and 8.5) was collected. Furthermore, in the third wave, more precise questions were asked than in earlier cohorts about the retrospective alternative choice of educational and career path (question 1.5), the reasons against taking up a master's degree course (question 5.2) as well as information on gainful employment (questions 7.2 and 7.4 to 7.6). Information was also gathered for the first time about student work activities and student financing, such as casual jobs and internships (questions 4.12 to 4.14). The third survey wave of this cohort also included the use of new tools for measuring personal factors, such as satisfaction (questions 1.1 and 1.8), self-perceived strengths and weaknesses (question 1.4) and decision-related personality traits (questions 1.6 to 1.7).

2.2 Pre-tests

[Goal and Procedure] The survey instruments for the three survey waves were tested in advance of the survey using pre-tests. The aim of this was firstly to test whether the question and response guidelines already used in previous cohorts would be perceived the same way by the students of 2008 as was the case for previous cohorts.[20] Secondly, the newly adopted survey instruments, which were used in all three waves of the 2008 panel study, needed to be tested for their comprehensibility, answerability, theoretical significance, reliability and validity. Thirdly, the inclusion of the new questions also led to basic changes in the structure and layout of the questionnaire as well as in the duration of the survey which needed to be evaluated.

In all three waves, an initial version of the survey instruments was first examined within the context of *expert reviews* (cf. Häder 2015, pp. 406–407) . Along with these evaluations, target-oriented development and graduation pre-tests[21] also took place in each of the respective waves. These were performed as so-called *pre-tests in the field* (cf. Häder 2015, p. 396), thus under as

[20] The survey instruments used in the 2008 panel study of school leavers are based largely on the design established for the survey series and the written questionnaires which were tried and tested in the surveys of the 2005 and 2006 school leavers cohorts.

[21] In contrast to development pre-tests, final pre-tests are normally used for smaller residual corrections to the interview tool, such as abbreviations or adjustments, as well as the checking of new filter procedures or changes in the print layout (cf. Schnell et al. 2005, pp. 348.

similar conditions as possible to those in the actual surveys. In the first two waves, iterative pre-tests (cf. Prüfer, Rexroth 2000) were carried out, which included three phases (wave 1) or two phases (wave 2) in total. [22] A pre-test was performed without iterations to test the survey tools used in the third wave.

[Test Subjects] Five to ten employees from the DZHW were brought in from the field of university research for all three survey waves as well as 50 to 100 external test staff, who were the same age, or rather, in a similar life situation as the respondents.

[Implementation] The pre-tests for the first survey wave took place at the start of the 2007/2008 academic year, i.e. six months before the respective field start. They were carried out at various types of school in order to keep the selection of respondents as structurally close to the final sample as possible.

A total of 58 pupils took part in the pre-tests for the first wave. In an initial pre-test, the survey was completed by pupils from the final year of a general education grammar school. In a second phase, the revised[23] survey instrument was presented to the final year group of a vocational school. A third phase saw the final year group of a comprehensive school being given the newly revised survey instrument. All versions of the instrument were completed during lessons in the presence of DZHW staff. They then discussed with the students their understanding of the questions and multiple-choice questions, and any difficulties with the survey instrument.

Two pre-tests were carried out for the second wave in summer 2008. In the first pre-test, 30 school leavers from the 2008 cohort[24] were sent the provisional survey instrument by post. The completion of the questionnaires was done with the request that any difficulties in completion, filter errors, unintelligible aspects such as incomplete multiple-choice questions were noted down and returned (postage free) to the DZHW. A second pre-test was geared towards the strongly represented university entrants. 65 test subjects completed the survey as part of an introductory event on empirical social research at the University of Hannover. Their suggestions for the improvement of the questions, multiple-choice questions and completion instructions, particularly on the focus topic of tuition fees and reasons for aborting studies, were incorporated into the final survey instrument.

The third wave in October 2012 was also preceded by a pre-test from students from the University of Hannover. The pre-test with 50 participants led to an improvement of the survey instrument, particularly in the area of newly developed questions on the subject of ‘Casual jobs and work placements during your studies’.

[22] In each phase the questionnaires, which each contained revisions from the previous pre-tests, were presented to the next survey groups.

[23] There is no information available on the changes made to the survey instrument after the pre-tests.

[24] These pre-test participants were not part of the sample, but they had taken part in the first survey wave and therefore provided their addresses (see section 6.3 Data Checking and Data Cleansing).

3 Population and Sample Procedure

[Population] The population for the 2008 panel survey 2008 included all students in the Federal Republic of Germany, who obtained their (subject-related) higher education entrance qualification in the 2007/2008 academic year at a general education or vocational school.

[Sample Procedure] Due to missing or inaccessible student lists at an individual level, it was impossible to extract a basic random sample. The individuals therefore had to be recruited via the schools. A disproportional cluster sample process was chosen as the random sample design, as a special case in a multistage random selection (cf. Häder 2015, p. 169) . The schools and school branches[25] thereby represent the 'primary sampling units' (clusters), and the school leavers from the 2008 cohort the 'secondary sampling units' within this cluster.[26] The selected population comprised of a list containing all relevant schools and school branches. The stratification took place using the characteristics of federal state and school type. Within the layers, the sample quotas for the clusters *and* individuals were defined so that, on the one hand, they were based on the distribution of the 2004/2005 cohort of school leavers according to official statistics, and on the other hand, they take into account the response ratios of past surveys. In the course of this, persons with higher education entrance qualifications from Bremen, Hamburg and the Saarland, as well as pupils from evening schools and colleges, were given disproportionate selection probabilities in order to generate sufficient cases within these groups.

As part of the practical implementation of the sample drawing procedure, the selection of clusters (schools) took place first in every layer via a basic random sampling. Following this, the chosen schools confirmed their school branches and pupil numbers so that a second selection step could then be carried out, which aimed at fulfilling the sample quotas at an individual level. In this second step, as many schools/school branches were accidentally drawn in each *successive* layer, that at least 97 per cent of the layer-specific sample quota for the school leavers was achieved.[27] Overall, five different samples were generated after this process, from which the most variant deemed most suitable was selected.[28] After this, the drawn schools (or their school branches) were fully collated.

[25] Some vocational schools consist of various school branches, from which, however, only those whose attendance led to obtaining a higher education entrance qualification were relevant for the sample. Therefore, only certain school branches were considered in some cases.

[26] Example of the group of persons in a cluster: all persons with higher university entrance qualifications from the 2008 examination year from XY general education grammar schools in YZ federal state.

[27] The examination to establish whether the sample quota was achieved in each layer took place after each stage of the sampling. If, for example, the cumulative number of pupils from five schools fulfilled the sample quota for the number of pupils within the specific layer, no further schools were drawn after the fifth school had been selected.

[28] However, since the drawing in the second step took place without any replacement, but the draw sequence was not recorded and does not allow for the reconstruction of the selection criteria used for the selection of the final sample variant, the case-specific selection probabilities can only be estimated (see Chapter 7).

4 Implementation of the Surveys

[Maintenance of Contacts and Addresses] The drawn schools were contacted by the DZHW and invited to participate. The DZHW also communicated the criteria to the schools, which they could use to identify the target persons for the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008 (see Chapter 3).[29] As the schools are not permitted to give out the contact details of their pupils due to data protection reasons, they only communicated their respective pupil numbers (where appropriate for each school branch) to the DZHW. The DZHW then sent the appropriate number of survey documents for the first survey wave to each school.[30] Due to the direct field access, the survey was carried out via the schools within the class groups and, where possible, during lesson time.[31]

In order to enable direct contact through the DZHW with the persons who were willing to participate in the subsequent waves, their address details were collected in the questionnaire from the first wave. Upon receipt of a questionnaire by the DZHW, the address details on the submitted questionnaires were provided with a serial number in addition to the imprinted identification number. From this, a reference list with identification and serial numbers as well as the associated addresses was created.[32] In the second and third waves, all persons who had provided their address in the first wave and who had participated in the previous wave were contacted. The identification numbers and addresses were printed in the questionnaires for subsequent survey waves, so that the survey documents could be sent directly to the participants' postal addresses.[33] A reconciliation of addresses was performed in all three survey waves before the survey documents were sent out, in order to take account of persons who had changed their address in the meantime. The DZHW also carried out further address checks whenever any survey documents could not be delivered.[34]

[Survey Documents] The survey documents in all survey waves consisted of one cover letter (incl. data protection information) and one questionnaire per respondent. A postage-free return of the completed questionnaire was ensured through their collective return in small packs by the schools using a response envelope addressed to the DZHW.

[Fieldwork Phase] The survey period of the first survey wave lasted from December 2007 until the end of August 2008.[35] Due to the method of contact used via the schools, the DZHW did not have any direct influence on the exact survey date.

---

29 At this point it was occasionally possible to identify persons through the schools as belonging to the random sample, who were not part of the target population (overcoverage). This occurred, for example, whenever the school leaver had only obtained the school-based part of the 'Fachhochschulreife' (advanced technical college entrance qualification) without subsequently aiming for the full Fachhochschulreife qualification, or the schools had distributed the questionnaire to pupils from the wrong age group.

30 The school correspondence contained a prepaid response postcard, a flyer with selected results from previous panel studies on school leavers and the notice regarding participation awards. An explicit letter of recommendation was also sent to the schools from the educational authorities in 14 of the 16 federal states.

31 The exception to the rule were the schools, teachers or classes who did not take part in the survey (about which, however, no further information is available).

32 In order to ensure data protection, the address section was detached from the questionnaire and the reference list was stored separately from the survey data on a protected server.

33 It was possible for participants to update their address details in the address section of the questionnaires.

34 Along with the address update service from Deutsche Post and the information from the European civil register, enquiries were also made at the local residents' registration offices.

35 The fieldwork period was as extended for as long as possible, to run parallel with the other project stages, and every questionnaire received up to August 2008 was considered.

The survey periods of the second and third survey waves lasted from December 2008 until early September 2009 or respectively December 2012 to December 2013.[36] Specific delivery times could be determined for the survey documents from the available address list at the DZHW.[37] It was also possible to purposely send the reminder letters approximately two months after the start of the survey only to the specific individuals who had not yet taken part in the survey.

[Measures to Increase Response] The measures for increasing the response rate were aimed on the one hand at inducing organisational support from the schools for the survey, and at the individual motivation of respondents on the other. For the initial contact, the organisational support of the schools (and responsible contact persons) consisted of a willingness to fill in their lesson times, to collect the questionnaires at the end of lessons and, where appropriate, to help with the return of the questionnaires for particular classes. In return for this, the schools were provided with school-specific ratings. Prizes were also awarded to the schools that showed a strong willingness to participate (x3 prizes of 500 euros). Also of use were the reference in the covering letter to the overall social importance for the survey, ministerial requests to support the survey and letters and telephone contact with schools, from which few or none of the questionnaires had been returned. The response-inducing measures, which were aimed at creating individual motivation among respondents, consisted of a reference to the overall social and age group-specific use of the survey in the covering letter to respondents, a flyer enclosed with the letter, a project website[38] (with information on the project and the resulting publications) and the sending of a reminder letter (in waves 2 and 3). As an individual material incentive, respondents were given the opportunity in the first wave to win 15 book vouchers valued at 100 euros each. The use of response-inducing material incentives was dispensed within the second wave. In the third wave, all survey participants took part in a prize draw with the following prizes: one Apple iPad4, two Apple iPod nanos and twenty Amazon vouchers with a value of 50 euros each.

[36] Only the occasional questionnaire was returned to the DZHW after June 2013. The last questionnaire in the dataset was received in December 2013.
[37] The mailing dates can no longer be exactly reconstructed.
[38] Over the course of the survey, the project website (http://www.panel2008.de/) was updated for each wave.

5 Response Rate

[Response Rate] Based on the random sample and survey design (see Chapter 3), letters were sent to 1,363 schools for the survey in the first wave. 959 (70 per cent) of these schools replied with information on the number of anticipated school leavers with higher education entrance qualifications. This corresponds to a figure of 91,311 pupils. The adjusted gross sample resulting from the first survey wave included 57,622 school leavers who were contacted (13% der Grundgesamtheit, cf. Heine et al. 2010, p. 10).[39] The response rate was 48.9 per cent, meaning that a net sample of 28,182 evaluable questionnaires were recorded. 17,558 persons, i.e. around 62 per cent of the 28,182 participants in the first wave, agreed to be contacted for further surveys.

Consequently in the second survey, which was conducted by post, 17,558 persons could be included in the gross sample. The response rate of 33.8 per cent for the second survey was significantly lower than that of the first survey, which led to a net sample of 5,933 persons.

These 5,933 respondents, who had taken part in both the first and second surveys, formed the gross sample for the third survey wave. The response rate for this survey wave was 61.9 per cent (3,671 persons).

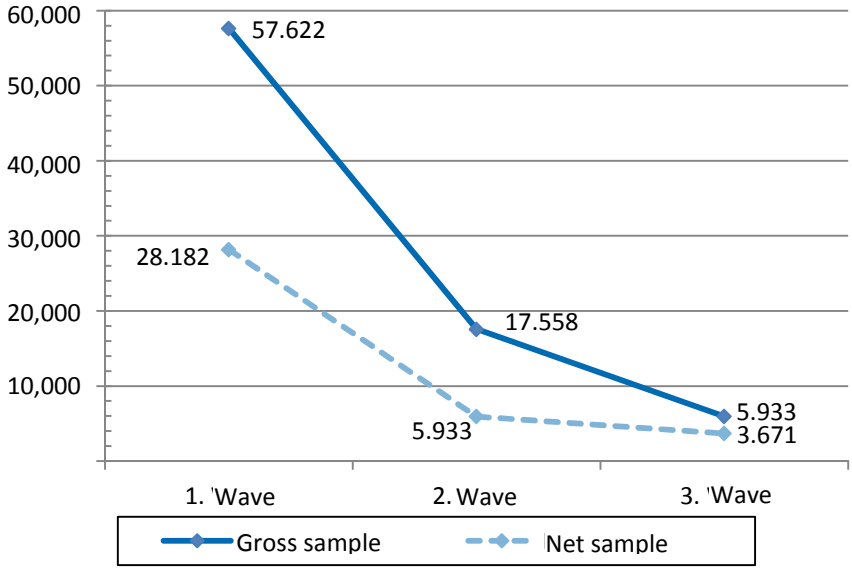
[Panel Attrition] The Panel Study of School Leavers with a Higher Education Entrance Qualification 2008 was affected by attrition processes.[40] These were manifested in the basic refusal to participate in subsequent surveys (non-provision of address details for contacting in the second wave) or in the form of non-participation in the second survey wave after (attempted) contact was made. The outlined development of the gross and net samples over time is illustrated in Figure 3 and Table 2.

As only those persons were contacted in the second and third waves who had taken part on previous occasions, there is a clear reduction in the gross sample from one survey date to the next. Thus only 6.4 per cent remain in the three-wave-panel accounted for in the initial sample.

[39] In the course of the responses from the schools, persons were also classified retrospectively as not belonging to the random sample as they had only completed the school-based part of the Fachhochschulreife in the relevant examination year of 2008. This type of *overcoverage* occurred because some schools had several school branches, which were not all part of the sample, but for which the schools had reported back information. Both the gross and net samples have already been adjusted for these figures.

[40] cf. Schnell et al. 2005, pp. 421.

Figure 3: Development of Gross and Net Samples in Time Sequence



Note on gross and net sample: only people entitled to take part in the sample

The clear discrepancy between the response rate in the second wave of the 2008 survey and that of the second wave in previous school leaver age groups can partly be explained by the form of survey used in the first wave: By conducting the survey in class groups, it is presumed that people were also motivated to take part who would otherwise not have done so. This basic unwillingness to participate possibly contributed to a higher rate of refusal in the second wave (sent by post) and was also manifested in the rather high number of forms without addresses. At the same time, no material incentives were given to respondents in the second wave. These could also have contributed to the lower willingness to participate.

Table 2: Gross and Net Samples and Response Rates from the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

	1st wave	2nd wave	3rd wave
Gross sample	57,622	17,558	5,933
Net sample	28,182	5,933	3,671
Response rate	48.9%	33.8%	61.9%
Net sample share at the previous net sample	--	21.1%	61.9%
Net sample share at the gross sample t1	48.9%	10.3%	6.4%
Net sample share at the previous gross sample	--	30.5%	33.8%
Net sample share at the gross sample t1	--	30.5%	10.3%

6 Data Preparation

In the following sections, various steps in data preparation are described. These took place analogously in the three survey waves. The procedures described in Chapters 6.1 to 6.3 had already been conducted by the primary research project. The generation of variables (Chapter 6.4) was carried out by the primary research project as well as the RDC during data preparation. Procedures described in Chapters 6.5 to 6.7 were carried out by the RDC building on the work of the primary research project. Additional procedures (e.g. weighting and anonymization) are explained separately in Chapters 7 and 8.

6.1 Data Transfer

For further processing, respondent data retrieved from the paper questionnaires was transferred to a computer readable format using a code plan. Prior to this, numerical codes for the open responses were recorded on each of the paper questionnaires (see Chapter 6.2) and preliminary manual corrections were undertaken to facilitate data transfer (see Chapter 6.3).

[Production of a Code Plan] A code plan was produced based on the survey questionnaire. It recorded to which question or sub-question a variable is assigned, the name of the variable and the numerical codes used for the standardised answers of the respondents. To establish the order of data entry, the variables were additionally numbered.[41]

[Data Entry] For data transfer, the code plan, further instructions on data entry and the prepared paper questionnaires were given to an external service provider. Their typists manually performed the compilation of the data.

6.2 Coding of Open Responses

Before the data transfer, the (semi-) open responses were coded. Using a coding list, numerical codes were assigned to them. For each variable, various code lists were used. This was done using classification keys for official statistics (e.g. German Classification of Occupations, key lists of student and examination statistics etc.) or keys already used in prior studies For some variables, new code lists were developed on the basis of the entries from the 2008 cohort of school leavers. For some semi-open questions, no new variables with numerical codes were created. Instead, entries were only assigned to the existing (closed) response categories. Some open questions were not coded at all because they were primarily created as context information for coding other open responses.[42]

Coded topics and respective code lists are presented Table 3. The data set contains exclusively the coded numerical variables. The open entries themselves are not contained in the data set.

[41] Data were generated in a simple, column-oriented text format without a heading containing the variable names. The code plan therefore established in which order the data were to be generated so that the data belonging to a variable could be entered in the correct column.

[42] This refers to the open questions about the next steps after leaving school in the first wave (question 11) and the open follow-up questions in the second and third waves. These data were gathered for the purposes of obtaining additional information in case of uncertainties in the activity table.

The values of the variables are documented in the Data Set Report as well as in the Metadata Search Portal[43].

Table 3: Coded Topics and Code Lists used in the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

Characteristic	Code List	Code List ID[a]
Type of school (identified)	Project's Own Coding	cl-dzhw-6
Type of school	Project's Own Coding	cl-dzhw-20
School subjects, groups	Project's Own Coding	cl-dzhw-7
(Skilled) job description	Destatis German Classification of Occupations 1992	cl-destatis-kldb-1992
Study subjects	Destatis Key List of Student and Examination Statistics Winter Semester 2007/08)	cl-destatis-studienfach-2008[b]
Graduation (2nd wave)	Project's Own Coding	cl-dzhw-12
Graduation (3rd wave)	Project's Own Coding	cl-dzhw-13
Higher education institution	German Universities: Destatis Key List of Student and Examination Statistics (Winter Semester 2007/08)	cl-destatis-hochschule-2008[c]
	Foreign Universities: Project's Own Coding	cl-dzhw-9
Type of higher education institution	Destatis Key List of Student and Examination Statistics (Winter Semester 2007/08)	cl-destatis-hochschule-2008[c]
Federal state	Destatis Federal State Codebook (corresponds to both first numbers of the Official Municipality Codebook, AGS)	cl-destatis-bundesland-1990[d]
Nationality (other)	Project's Own Coding	cl-dzhw-17
Country (foreign)	Project's Own Coding	cl-dzhw-19
Other open enquiries	Assignment to given categories or project's own coding	---

[a] A code list-ID was only provided if the categories were not derived from the actual entries in the data set, but rather derived from another classification system.
[b] supplemented by project's own codes for missing subjects
[c] supplemented by project's own codes for missing institutions of higher education
[d] supplemented by project's own codes for foreign countries

6.3 Data Checking and Data Cleansing

[Preliminary Manual Correction] Before the transfer of data, a preliminary manual correction was performed on the questionnaires, and if necessary, responses were adjusted.[44] First, this was intended to improve the clarity of the data. To this end, the form of existing data was changed. For example, difficult-to-read responses or deletions of the surveyed persons were clarified, figures were entered in the corresponding boxes, verbal responses translated from grades into digits (e.g. "good" = 2.0) and entries in the educational biography were divided across a number

[43] https://metadata.fdz.dzhw.eu/#!/en
[44] The number of corrections was only documented on the paper questionnaires and can no longer be systematically reconstructed.

of lines if the surveyed persons had written multiple entries on one line (e.g. with two parallel courses of study).

Second, the manual test aimed to cleanse mistakes or inconsistencies in the responses of the surveyed persons before the corrections backed up by software were performed (see below). To this end, consistency tests within a wave were performed, for which content- and time-related responses on the same thematic areas were aligned with each other. In the third wave, for example, a comparison was made between the data in the activity table (question 2.1) and the data on vocational training (questions 3.1 to 3.4 and 3.8), the data on study subject, degree and institution of higher education (questions 4.1 to 5.9) or the data on gainful employment (questions 7.1 and 7.2).). If possible, identified inconsistencies were cleared up through comparison with other entries in the questionnaire. Otherwise, a corresponding missing code (see Chapter 6.7) was assigned.

[Software-Assisted Correction] Following data transference, a comprehensive test and correction of the data with the help of DZHW's own software was performed. Software-Assisted Correction was used to catch mistakes in the preliminary manual correction and data transfer. Furthermore, inconsistent responses of the respondents that could not be tested in the preliminary manual correction were identified.

To this end, the compiled questionnaires data (as well as in the second [and third] survey waves, also data from the first [and second] survey waves) were read into a data bank. In addition, valid value ranges and answer combinations were defined and tested systematically. The following types of tests were carried out:

- *Test of Value Ranges:* It was tested whether the response lay in the value range defined of the respective recorded variable.
- *Test of Adherence to Filter Procedures:* Based on the defined filter procedure of the questionnaire, it was tested whether responses that would have been expected from the respondent were not (i.e. completeness test) and whether responses were made that should not have been (i.e. filter errors).[45]
- *Test of Variable Consistency:* The consistency of responses within a questionnaire as well as between survey waves was tested. In addition to feature combinations, which were already tested in the preliminary manual correction, more complex feature combinations could also be tested here. The consistency of data between the different survey waves was also checked. Particular emphasis should be made here on the comparison between the data on the educational steps in the second wave (questions 17 to 20 and question 22) and the activity table in the third wave (question 2.1).[46]

In total, hundreds of consistency rules were defined and tested. Missing, incorrect or implausible values were first tested using the paper questionnaire to determine whether the corresponding value had been falsely (or not at all) transferred. Then the correct value was inferred using other

[45] The input filter of the variables assigned to the individual questions is documented in the Data Set Report as well as in the Metadata Search Portal (https://metadata.fdz.dzhw.eu/#!/en). They define which surveyed group should answer a question for a respective variable.

[46] For the activity table and the individual questions on educational steps, overlaps were purposely devised in the activity data between the two waves.

responses in the questionnaire.[47] In case of doubt, a specific missing code was assigned (see chapter 6.7). Corrections of mistakes were documented[48] and checked by at least one further person.

[Deletion of Cases] Cases were removed from the dataset in all three waves. A case was deleted if half of the questions or core questions (e.g. on course of study) were not answered or if too many inconsistencies were present. These cases were graded as not possible to evaluate and removed. Moreover, some cases were identified after the first wave as not belonging to the target population.[49] These were likewise removed from the data set. A total of 301 cases were deleted in the first wave, 84 cases in the second wave and 8 cases in the third.

6.4 Generation of Variables

In addition to the variables containing the coded answers of the respondents, the dataset from the 2008 cohort of school leavers also contains generated variables. One the one hand, this includes variables that were numerically coded from the originally open entries (see chapter 6.2). On the other hand, variables were changed due to data protection reasons and more frequently required variables were generated from the values of one or more source variables (e.g. merging course subjects in to areas of study and subject groups or deriving the location and type of the higher education institution from the higher education institution variables). The newly generated variable is identified in the data by the suffix "_g#". An overview of all generated variables for the 2008 panel study of school leavers as well as detailed documentation of the individual variables with information on their respective characteristics and calculation rules can be found in the data set report as well as the Metadata Search Portal[50].

Where possible, generated variables were positioned in the data set according to the respective output variable. If the output variable is no longer available in the data set due to anonymisation (see Chapter 8), the generated variable takes its place in the data set If a variable was generated from various source variables, it was inserted after the variable to which it is thematically closest. If a clear assignment was not possible (as with the weighting variables), the generated variable was inserted at the end of the data set. With minor exceptions, variables generated in the context of anonymisation measures were created by the RDC.

6.5 Generation of the Data Sets

[Merging of the Waves] The data of the three survey waves were merged. The matching of data from the three survey waves were made using the identification numbers of the respondents produced in the fieldwork phase (see chapter 4).

47 Inconsistencies between the survey waves were corrected in the third wave, while they were uncorrected in the previous surveys.

48 Documentation of the correction of mistakes was performed manually on the paper questionnaires and thus cannot be systematically reconstructed.

49 This occurred, for example, if the school leaver had only obtained the school-based part of the 'Fachhochschulreife' (advanced technical college entrance qualification) without subsequently aiming for the full 'Fachhochschulreife' qualification, or the schools had distributed the questionnaire to pupils from the wrong age group.

50 https://metadata.fdz.dzhw.eu/#!/en

[Generation of the Data Set] The merged data was stored in a data set and completed by the additionally generated variables. One data row (wide format) exists for each respondent. The sequence of the variables is oriented to the sequence of related questions in the questionnaire

[Data Format] The dataset is available in Stata format as standard. Possibilities for the use of other formats and/or analysis software can be found on the website (see also Section III).

6.6 Assignment of Variable Names, Variable Labels and Value Labels

[Variable and Value Label Assignment] For variable and value label assignment, formulations from the questionnaire were used, or in some instances, concise formulations were chosen. As a rule, the variable labels are based on the corresponding question. Depending on the type of question, value label assignments are based on the response options or a combination of the question and response options. For generated variables based on definite classifications, value labels were adopted verbatim from the classification keys. The variables and value labels are stored as bilingual labels (German and English) within the same data set.

[Naming Variables] A consistent naming system was created at the RDC for the naming of variables. With the exception of the identifier variable (pid) and the wave variable (wave), variable names were formed according to a prefix-root-suffix scheme that facilitates automated processing. In addition, the variable names provide meta-information on the corresponding variable. The prefix of the variable contains the wave identification in one letter. The root of the variable contains the thematic area to which the variable is assigned and is denoted by a three-letter English abbreviation. Table 4 presents an overview of the various thematic areas of the 2008 cohort of School Leavers as well as the related abbreviations for the root of the variable name. . The suffix, separated from the root by an underscore, contains various additional information so as to identify generated variables as well as various modes of data access. For indicators that are used in multiple survey waves, the names of the associated variables were harmonised through the assignment of an identical root.

Detailed information on variable assignment for the 2008 panel study of school leavers can be found in the Data Set Report.

Table 4: Thematic Areas and Abbreviations for the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

Thematic Area Abbreviation	Thematic Area (English)	Thematic Area (German)
act	activity	Tätigkeiten
dec	decisions	Entscheidungen
dem	demographic information	demografische Informationen
eng	engineering	Wahl ingenieurwissenschaftlicher Studieng-änge
fea	fears	(Studienbezogene) Sorgen
fin	financing	Finanzierung
fut	future prospects	Zukunftsaussichten
goa	goals (occupational, life)	(Berufs- und) Lebensziele
inc	income	Einkommen
inf	information	Information(-sverhalten)
int	interests	Interessen
job	job	Jobs
mot	motives	Tätigkeitsmotive (Studium / Education / Erwerbstätigkeit)
per	personality	Persönlichkeit
sat	(job) satisfaction	(Berufs-)Zufriedenheit
sch	school	Schulzeit
ski	skills	Fähigkeiten
stu	studies	Studium
sys	system variables	Systemvariablen
tra	transition	Übergang (Schule-Beruf)
voc	vocational training/education	(Berufs-)Ausbildung
wgt	weighting variables	Gewichtungsvariablen

6.7 Coding of Missing Values

For coding missing values, a comprehensive system was created in the RDC, in order to guarantee unified coding for missing values across various data sets of the DZHW. Missing responses were coded using three-figure negative values. Table 5 presents an overview of the system for coding missing values. The coding for missing values used in the 2008 Panel Study of School Leavers is highlighted.

Missing values can be assigned to four different groups. First, missing values may arise if the respondent does not answer the survey questions (i.e. non-response). Second, missing values may be assigned due to the filter procedure, i.e. if questions are not relevant to the respondent (not applicable). The third group contains missing values assigned through the primary research project or the RDC in the course of the data preparation (i.e. edited missing value). This includes missing variables for certain variables due to anonymisation measures (see Chapter 8). The fourth group comprises missing values assigned for individual items in the context of data preparation of a specific data set (e.g. 'Dropout in the broadest sense' next to item cstu30_g1, question 4.9, 3rd wave).

Table 5: System of the RDC-DZHW for Missing Values

Value range	Code	Value label
-999 to -990: Non-response	**-999**	**don't know**
	-998	**no answer**
	-997	no answer (answer category)
	-996	aborted interview
	-995	**non-participation (panel)**
	-994	refused
-989 to -970: Non-applicable	**-989**	**missing due to filtering**
	-988	**does not apply**
	-987	missing due to design (split questionnaire)
	-986	missing due to design (wave)[a]
	-985	missing due to design (cohort)[b]
-969 to -950: Edited missing value	**-969**	**unknown missing value**[c]
	-968	implausible value[d]
	-967	**anonymised**
	-966	Non determinable[e]
	-965	**invalid multiple answer**
-949 to -930: Item-specific missing values[f]	**-949**	**already started second course of study**
	-948	**joint course of study**
	-947	**broader definition of abandonment of studies (BA/VerFH degree)**
	-946	**degree course without more detailed description**
	-945	**chosen out of sequence**
-929 to -920: other missing values	-929	loss of data

[a] This value is only assigned for data sets in long format.
[b] This value is only assigned to pooled data sets.
[c] This value is assigned when no cause can be reconstructed.
[d] Responses which are classified as implausible due to various factors in the coding phase receive this value. An exact reconstruction may no longer be possible.
[e] This category is assigned when clear coding is not possible, e.g. open response which could not be coded because it is illegible.
[f] The characteristics of these missing categories are, by definition, specific for every data set.

7 Weighting

Weighting the data adjusts for sample bias compared to the defined population. The text below begins with a general introduction to the procedure and an illustration of the weights created. This is followed by a detailed description of the weighting procedure.

7.1 Procedure and Instructions for Use

[Causes of sample bias] Two processes are relevant for sample bias:

- **Bias due to Design:** Disproportionalities are deliberately produced to increase the number of cases in certain relevant subgroups (cf. Chapter 3).
- **Bias through non-response:** Attrition processes (e.g. non-participation, unreachable, postal error) lead to reduced response and thus to a difference between gross and net sample (cf. Chapter 5). If these processes are non-systematic (Missing Completely at Random), they can be ignored.[51] However, they mostly result from a systematic process (Missing at Random, Not Missing at Random), which requires modelling.[52]

[Conceptual Procedure] In the course of the weighting procedure, at first disproportionalities due to design should ideally be offset. In case of random sampling, the *design weights* are directly derived from the sample plan. Related to this, an adjustment of the design weights – using cross sectional and longitudinal *non-response weights* – should be produced on the basis of information on participants and non-participants. As a last step, the non-response adjusted design weights can be calibrated using distributions of characteristics from the population.

Given the description of the population and sample procedure in Chapter 3, such an ideal-type method cannot be implemented with the data in the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008. The sample design does not allow exact probabilities of inclusion to be derived. As there was also no information available in the first survey wave regarding non-participants, no individual dropout weighting is created for this. An estimated design weight is therefore calculated as a cross-sectional weight for the first wave which is calibrated using information from the population. As information on non-participants can be gained from the distribution of characteristics of the population, a form of non-response adjustment also takes place herewith. A non-response weight is also calculated for the second and third waves, which uses the information from the previous respective waves to model the non-participation in the second and third waves. The weights provided in the data set are shown in Table 6.

[51] Insofar as the loss of statistical test strength through the reduction of the sample is considered irrelevant.
[52] For the various forms of attrition processes see essentially Rubin, 1976.

Table 6: Weights provided for the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008

Variable Name	Description
wgt_t1	Cross-sectional weight 1st wave
wgt_t1t2	Longitudinal weight 2-wave panel t1t2
wgt_t1t2t3	Longitudinal weight 3-wave panel t1t2t3

[Instructions for Use of the Weights] The generated weights involve *probability weights,* which can be taken into account in Stata with the help of .ado-specific options.[53] The weight wgt_t1 is intended for evaluations of the first wave, while wgt_t1t2 and wgt_t1t2t3 are intended for evaluations of the second and third wave surveys. It is essential to note that weights only represent meaningful correction quantities if the analysis model contains the variables used for the weighting or in relation to them. For this reason, weights must always be used with a focus on the research question. In the following section, the procedure for producing the weights will be presented in more detail.

7.2 Weighting of the Data

[Cross-Sectional Weighting] Due to sample design, the design weights could not be derived exactly and therefore had to be estimated. The estimation of the design weight was performed at every stratum s as follows[54]:

$$\widehat{dwgt}_{sci} = \frac{n_{cs}}{N_{cs}}^{-1}$$

Due to missing information on non-participants in the first wave, no comprehensive adjustment of the estimated design weights was possible on an individual basis for the attrition process through non-participation (non-response). However, a calibration was performed, which had the adaptation of the estimated design weights using characteristics of the population as a model. For this, characteristics included the sex, school branch and federal state.[55] Since the characteristics are reflective of the population as a whole, information on the non-participants additionally allowed for a non-response adjustment with respect to the characteristics used for the calibration. The calibration of the estimated design weight $\widehat{dwgt}_{sci}$ was performed using the Raking algorithm[56], resulting in a cross-sectional weight wgt_{t1_i} in the first wave of the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008. For the second and third survey waves, no cross-sectional weight was produced since no newly added individuals were surveyed (i.e. refreshment sample) and because no person could participate in Waves 2 and 3 if they had not participated in the first survey wave.

[53] See Stata help (Command: *help weights*).

[54] Where n_s corresponds to the number of clusters in a stratum, N_s corresponds to the number of clusters in the respective stratum of the population. Since the clusters were fully surveyed, the selection probability of an individual corresponds to the selection probability of the corresponding cluster.

[55] The information from the population was derived from data of the Federal Statistical Office (Examination Statistics 2009a, 2009b).

[56] Raking is also termed 'iterative proportional fitting' (ipf) (cf. Kolenikov 2014).

[Longitudinal Sectional Weighting] For the weighting of the two- and three-wave-panel studies, attrition processes had to be considered in time sequence (panel mortality; see also Chapter 5). For this purpose, an attrition weight was calculated in each case, which illustrated the probability of participation in the next wave. In contrast to the non-response adjustment in the first wave, more information from the previous respective waves was available for the non-participants of waves 2 and 3 (σ_t). This information served as covariates in a probit regression model, which estimated probability of participation at a given point in time $P(R_{t+1})$. For variables with missing values, these were used as additional variable categories so that cases with item non-response could also be included in the model. Furthermore, the assumption that item non-response represents a significant predictor for unit non-response in future waves could be tested. A row of predictors from the first wave demonstrated their significance for the prediction of probability of participation in the second and third waves. Conditional probability of participation could be derived from the model, the reciprocal value of which represents the nonresponse weight for the second and third waves[57]:

$$NR_{gew_{t+1_i}} = P\left(Res_{t+1_i} \middle| \sigma_{t_i}\right)^{-1}$$

The total weight (non-response adjusted design weight) for each part of the panel results from the product of the estimated design weight with the respective non-response weights:

$$wgt_{i_{t1t2}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t2}}}$$

$$wgt_{i_{t1t2t3}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t2}}} \times NR_{i_{gew_{t3}}}$$

Then the respective total weight was calibrated to the target population using the raking algorithm. The same characteristics were used here as for the calibration of the design weight in the first survey wave.

[Standardisation for the Number of Cases in the Sample] As is customary with the practice of social science research, the calculated weights were standardised for the number of cases in the sample.

[Trimming of the Weights] The initially calculated weights exhibit a small proportion of outlying weighting factors. In order to remove them, all weights were subjected to a trimming according to Potter 1990 (see also Valliant et al. 2013, pp. 388). The procedure is based on the assumption that the weights conform to a probability distribution (beta distribution). All those weights that lie above the 99 percent quantile are truncated to this limit. Excess on the other side of the truncation is distributed among the remaining weights.

[57] The procedure corresponds to its logic according to *Propensity Score Matching*, which derives from Rosenbaum, Rubin 1983 (cf. Blumenstiel, Gummer 2015).

8 Anonymisation

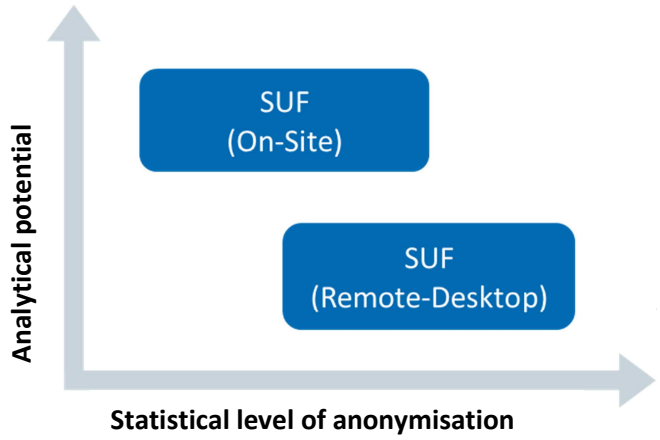
[Data Protection Legal Framework] The Federal Data Protection Act (BDSG) applies to personal data that the DZHW collected through volunteer surveys.[58] Accordingly, personal data that are collected during scientific research may be processed or used exclusively for the purposes of scientific research (cf. §40 para. 1 BDSG). Moreover, personal data must be anonymised (cf. §40 para. 2 BDSG) in order to protect respondents. According to the BDSG, the procedure of anonymisation is defined as "the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual" (§3 para. 6 BDSG). Regarding the disclosure of data from scientific research projects to third parties, the data must either be *absolutely anonymised* so that no reference to the person can any longer be produced, or at least *de facto anonymised* so that the construction of a reference to a person would mean a disproportionally high expenditure and thus the likelihood of re-identification of a person is minimal.

[Data Access, Level of Anonymisation and Analytical Potential] For the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008, the RDC makes a de facto anonymised SUF available for scientific secondary use. The anonymity of the surveyed persons is thus protected by a combination of statistical measures and technical access barriers. The more strongly data access is technically controlled, the lower is the risk of de-anonymisation of the data, the less the data must be limited in terms of information by statistical measures and the greater their analytical potential remains.

The SUF is provided using two different modes of access: remote desktop and on-site (for further information cf. Section III). For each mode of access a different SUF variant is made available, which is varyingly strongly anonymised and correspondingly contains less or more information. Figure 4 gives an overview of the respective level of statistical anonymisation and the related analytical potential. In the following the statistical anonymisation measures performed are explained according to data product (SUF) and mode of access.

[58] The BDSG is applicable since the DZHW GmbH is legally a public body of the federal government (cf. § 2 para. 3 BDSG). The federal government possesses an absolute majority of the shares in DZHW GmbH and the institute performs duties of public administration of the federal government in the broadest sense. For interpretation of individual legal aspects the European Data Protection Guidelines can be used as a complement.

Figure 4: Data Access, Statistical Level of Anonymisation and Analytical Potential of the Data of the Panel Study of School Leavers with a Higher Education Entrance Qualification 2008



[Statistical Anonymisation Measures] In the course of anonymisation, all information that directly allows individuals or institutions to be identified is deleted These so-called *direct identifiers* such as names, addresses and e-mail addresses were placed in a separate data set during the field phase of the School Leavers Panel 2008 (see Chapter 4) and are not contained in the various SUF variants. To further prevent any re-assessing of this data, both the original identification numbers of the respondents and the institutions (school, institution of higher education) were removed and replaced with new, randomly assigned identification numbers.

In order to ensure the de facto anonymization of the SUF data of respondents, *quasi-identifiers* were determined, i.e. information which, in combination with or by the allusion to external information, allows for indirect identification.[59] For the School Leavers Panel 2008, the following quasi-identifiers were identified, which are present in external data sources[60] as well as in the data of the School Leavers Panel: regional information (country of birth, place where higher education institution entrance qualification was gained, institute of higher education and workplace), nationality, language in parental home, type of school, school exam subjects, institution of higher education, subject studied, degree type, job details. To prevent a clear association with the data from the School Leavers Panel, these key attributes – according to data product and mode of access – were aggregated or deleted (see Table 7). The language spoken in the parental home, for example, is available on-site without any restrictions, while in the remote desktop SUF, 13 languages are shown separately and the rest are aggregated into 'other languages'. Open responses are likewise quasi-identifiers (cf. Ebel 2015, p. 3) and were coded or not transferred word-for-word (see Chapter 6.2).

Finally it was checked whether the data contained *sensitive information*, e.g. on health, sexual orientation or political views. This information, although not suited for re-identification of

[59] It is pointed out that the identification of a person is already made more difficult by the sample selection, since uncertainty arises whether a respondent has a unique combination of characteristics in the population.

[60] E.g. student and examination statistics of the Federal Statistical Office, alumni networks of the higher education institutions or also professional networks.

individuals or institutions, can be used in case of de-anonymisation (cf. Koberg 2016, p. 694). Therefore, its protection is particularly important (cf. §3 para. 9 BDSG, Art. 8 para. 1 and 2a EG-DSRL). In the School Leavers Panel 2008, information on health was collected without further consent of the respondents for secondary use. Hence, these answers were deleted in all SUF variants.

Table 7: Statistical Anonymisation Measures for the Data of the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification 2008 by Mode of Access[61]

Characteristic	On-Site SUF	Remote Desktop SUF
Direct Identifiers	Deletion and assignment of a random ID for schools and pupils	Deletion and assignment of a random ID for schools and pupils
School location	Aggregation to federal state	Aggregation to federal state
Type of school	Aggregation of poorly populated types of school	Aggregation to general education and vocational schools
Examination subjects/main subject (school, vocational school)	Available	Aggregation to (vocational) school subject groups
Subject	Available	Aggregation to areas of study[a]
Higher education institution	Information on type of higher education institution	Information on type of higher education institution[b]
Location of higher education institution, residence in December 2008	Aggregation to federal state and country of a foreign higher education institution	Aggregation to federal state and country of a foreign higher education institution[c]
Place of work (postcode)	Available	Reduction to two digits
Profession	Available	Aggregation to occupational orders[d]
Nationality	Available	5 countries shown separately, otherwise aggregation (country groups, world regions)
Language spoken in parental home	Available	13 languages shown separately, other languages aggregated to 'other languages'
Health characteristics	Deletion	Deletion

[a] according to the Destatis Key List of Student and Examination Statistics Winter Semester 2007/08
[b] Distinction only between university of applied sciences and university (including teacher training colleges, theological colleges, art and music colleges) and not for the Saarland (only one university)
[c] Only foreign countries with more than one higher education institution
[d] Based on the Destatis German Classification of Occupations 1992

[61] Detailed information on the anonymised variables can be found in the Data Set Report and the Metadata Search Portal (https://metadata.fdz.dzhw.eu/#!/en).

9 Bibliography

Baier, Tina; Helbig, Marcel (2014): Much ado about € 500: do tuition fees keep German students from entering university? Evidence from a natural experiment using DiD matching methods. In *Educational Research and Evaluation: An International Journal on Theory and Practice* 20 (2), pp. 98–121. DOI: 10.1080/13803611.2014.881745.

Blumenstiel, Jan Eric; Gummer, Tobias (2015): Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In Jürgen Schupp, Christof Wolf (Eds.): Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 13–44.

Ebel, Thomas (2015): Empfehlungen zur Anonymisierung quantitativer Daten. GESIS - Leibniz-Institut für Sozialwissenschaften. Mannheim.

Häder, Michael (2015): Empirische Sozialforschung. Eine Einführung. 3rd ed. Wiesbaden: Springer Fachmedien Wiesbaden, checked on 6/12/2015.

Heine, Christoph; Quast, Heiko (2009): Studierneigung und Berufsausbildungspläne. Studienberechtigte 2008 ein halbes Jahr vor Schulabgang. HIS. Hannover (HIS: Forum Hochschule, 4/2009).

Heine, Christoph; Quast, Heiko; Beuße, Mareike (2010): Studienberechtigte 2008 ein halbes Jahr nach Schulabschluss. Übergang in Studium, Beruf und Ausbildung. HIS. Hannover (HIS: Forum Hochschule, 3/2010).

Helbig, Marcel; Baier, Tina; Kroth, Anna J. (2012): Die Auswirkung von Studiengebühren auf die Studierneigung in Deutschland. Evidenz aus einem natürlichen Experiment auf Basis der HIS-Studienberechtigtenbefragung. In *Zeitschrift für Soziologie* 41 (3), pp. 227–246.

Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum. Institut für Arbeitsmarkt- und Berufsforschung (IAB). Nürnberg (06).

Koberg, Tobias (2016): Disclosing the National Educational Panel Study. In Hans-Peter Blossfeld, Jutta von Maurice, Michael Bayer, Jan Skopek (Eds.): Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study. Wiesbaden: Springer VS, pp. 691–708.

Kolenikov, Stanislav (2014): Calibrating survey data using iterative proportional fitting (raking). In *The Stata Journal* 14 (1), pp. 22–59.

Lane, Julia; Heus, Pascal; Mulcahy, Tim (2008): Data access in a cyber world: Making use of cyberinfrastructure. In *Transactions on Data Privacy* 1 (1), pp. 2–16.

Lörz, Markus (2012): Mechanismen sozialer Ungleichheit beim Übergang ins Studium: Prozesse der Status- und Kulturreproduktion. In Rolf Becker, Heike Solga (Eds.): Soziologische Bildungsforschung. Wiesbaden: VS Verlag für Sozialwissenschaften (Kölner Zeitschrift für Soziologie und Sozialpsychologie: Sonderheft, 52), pp. 302–324.

Lörz, Markus; Quast, Heiko; Roloff, Jan (2015): Konsequenzen der Bologna-Reform: Warum bestehen auch am Übergang vom Bachelor- ins Masterstudium soziale Ungleichheiten? In *Zeitschrift für Soziologie* 44 (2), pp. 137–155.

Mertens, Dieter (1976): Beziehungen zwischen Qualifikation und Arbeitsmarkt. In Winfried Schlaffke (Ed.): Jugendarbeitslosigkeit. Unlösbare Aufgabe für das Bildungs- und Beschäftigungssystem? Köln: Dt. Inst.-Verl., pp. 68–117.

Potter, Frank J. (1990): A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the Survey Research Methods Section*, pp. 225–230.

Prüfer, Peter; Rexroth, Margit (2000): Zwei-Phasen-Pretesting. In Peter Ph. Mohler, Paul Lüttinger (Eds.): Querschnitt. Festschrift für Max Kaase. Mannheim: ZUMA, pp. 203–219.

Quast, Heiko; Scheller, Percy; Lörz, Markus (2014): Bildungsentscheidungen im nachschulischen Verlauf. Dritte Befragung der Studienberechtigten 2008 viereinhalb Jahre nach Schulabschluss. HIS. Hannover (HIS: Forum Hochschule, 9/2014).

Quast, Heiko; Spangenberg, Heike (2008): Mehrmalige Befragung der Studienberechtigten 2007/08. Erste Befragung. Sachstand März 2008 (unveröffentlicht). Forschungsprojekt im Auftrag des Bundesministeriums für Bildung und Forschung. HIS Hochschul-Informations-System GmbH. Hannover (HIS: Sachstandsbericht).

Rosenbaum, Paul R.; Rubin, Donald B. (1983): The central role of the propensity score in observational studies for causal effects. In *Biometrika* 70 (1), pp. 41–55. DOI: 10.2307/2335942.

Schneider, Heidrun; Franke, Barbara (2014): Bildungsentscheidungen von Studienberechtigten. Studienberechtigte 2012 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss. Edited by Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). Hannover (Forum Hochschule, 6).

Schnell, Rainer; Hill, Paul B.; Esser, Elke (2005): Methoden der empirischen Sozialforschung. 7.th ed. München: Oldenbourg.

Statistisches Bundesamt, Wiesbaden (Ed.) (2009a): Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2008/09. Statistisches Bundesamt. Wiesbaden (DESTATIS. wissen. nutzen., 1).

Statistisches Bundesamt, Wiesbaden (Ed.) (2009b): Bildung und Kultur. Berufliche Schulen. Schuljahr 2008/09. Statistisches Bundesamt. Wiesbaden (DESTATIS. wissen. nutzen., 2).

Valliant, Richard; Dever, Jill A.; Kreuter, Frauke (2013): Practical tools for designing and weighting survey samples. New York (NY): Springer New York.

Appendix 1: Probit regression for the formulation of the panel dropout weight in wave 2

Variable *(AV: Participation Wave 2)*	**b-coefficient**	**Z-value**
Sex *(reference: female)*	-0.286	-15.7900
Mother: University degree (*reference: yes*)		
no	0.065	3.1700
no response	-0.319	-4.5700
Father: University degree (*reference: yes*)		
no	0.058	2.8900
no response	-0.115	-1.8900
Study intention (*reference: yes*)		
no	-0.174	-8.4800
no response	-0.644	-2.5000
Residence: Federal state (*reference: North Rhine-Westphalia*)		
Schleswig-Holstein	0.118	1.9300
Hamburg	-0.138	-2.0700
Lower Saxony	0.116	2.9600
Bremen	-0.109	-1.5500
Hessen	0.065	1.6600
Rhineland-Palatinate	0.134	3.1200
Baden-Württemberg	0.060	1.8300
Bavaria	0.255	7.2100
Saarland	-0.241	-2.8400
Berlin	-0.172	-3.2400
Brandenburg	-0.071	-1.4600
Mecklenburg-West Pomerania	-0.054	-1.2600
Saxony	-0.096	-2.0100
Saxony-Anhalt	-0.055	-1.2800
Thuringia	0.050	1.0600
Type of school *(reference: grammar school)*		
Comprehensive school	-0.009	-0.2200
Steiner Waldorf Schools	0.330	1.9400
Night school	0.025	0.2800
College (not vocational college)	0.041	0.6400
Vocational grammar school	-0.159	-4.8800
Upper class of a vocational school	-0.216	-2.2900
Upper class of a sixth-form college	-0.062	-0.3200
Upper vocational school	-0.281	-5.4000
Technical college	-0.428	-14.6400
(Upper) vocational school	-0.423	-9.2600
Technical school	-0.679	-12.8500
Professional academy	-0.563	-2.8600
Final grade (rounded up) *(reference: 3.0 to 3.9)*		
Grade 1.0 to 1.9	0.686	17.6000
Grade 2.0 to 2.9	0.315	16.0400
Grade 4.0 to 4.9	-0.313	-4.2200
Grade 5.0 to 6.0	-0.767	-2.5000
no response	-0.159	-3.0200
Constants	-0.752	-25.3500
Number of observations	27,997[62]	

[62] Due to lack of values for the variable 'gender', no likelihood of participation could be generated from the gross sample for 185 cases. Though this only affected cases who didn't take part in the second wave, hence all 5,933 cases from the two-wave panel were given a weight.

Appendix 2: Probit regression for the formulation of the panel dropout weight in wave 3

Variable *(AV: Participation Wave 3)*	**b-coefficient**	**Z-value**
Sex *(reference: female)*	-0.184	-5.0200
Father: University degree (*reference: yes*)		
no	0.001	0.0300
no response	-0.348	-2.9200
Country of birth: Germany *(reference: yes)*		
no	0.154	1.7300
no response	0.312	0.7800
Residence: Federal state (*reference: North Rhine-Westphalia*)		
Schleswig-Holstein	-0.022	-0.1900
Hamburg	0.034	0.3100
Lower Saxony	-0.030	-0.4000
Bremen	-0.098	-0.6900
Hessen	-0.090	-1.1400
Rhineland-Palatinate	0.027	0.3100
Baden-Württemberg	-0.033	-0.5400
Bavaria	0.022	0.3400
Saarland	-0.171	-0.9400
Berlin	-0.198	-2.1200
Brandenburg	0.237	1.9900
Mecklenburg-West Pomerania	-0.215	-2.1900
Saxony	0.042	0.4600
Saxony-Anhalt	-0.230	-2.4900
Thuringia	-0.043	-0.4300
Foreign country	0.032	0.3500
no response	-0.140	-0.4900
Type of school *(reference: grammar school)*		
Comprehensive school	-0.210	-2.6800
Steiner Waldorf Schools	-0.311	-1.1100
Night school	-0.401	-2.5400
College (not vocational college)	-0.454	-3.9600
Vocational grammar school	-0.160	-2.5400
Upper class of a vocational school	-0.073	-0.3800
Upper class of a sixth-form college	-0.522	-1.3300
Upper vocational school	-0.131	-1.3600
Technical college	-0.281	-4.6200
(Upper) vocational school	-0.431	-4.3500
Technical school	-0.255	-2.1500
Professional academy	-0.657	-1.7000
Tuition fees *(reference: yes)*		
no	0.096	2.2000
no response	-0.115	-0.6700
Parental home: number of books *(reference: 201 to 300)*		
0 to 50	-0.148	-2.2000
51 to 100	-0.003	-0.0400
101 to 200	0.001	0.0200
301 to 500	0.005	0.0900
501 to 1000	0.097	1.5500
1000 and greater	-0.039	-0.5600
no response	0.126	0.6900
Country of birth: Germany *(reference: yes)*		
no	-0.247	-3.6700
no response	-0.547	-2.5100
Final grade mathematics (rounded up) *(reference: 3.0 to 3.9)*		

Grade 1.0 to 1.9	0.261	4.8900
Grade 2.0 to 2.9	0.170	3.6300
Grade 4.0 to 4.9	-0.065	-1.1300
Grade 5.0 to 6.0	-0.147	-2.0200
no response	-0.077	-0.3600
Final grade German (rounded up) *(reference: 3.0 to 3.9)*		
Grade 1.0 to 1.9	-0.062	-1.1100
Grade 2.0 to 2.9	-0.044	-1.0400
Grade 4.0 to 4.9	-0.094	-1.4100
Grade 5.0 to 6.0	-0.323	-2.0500
no response	-0.066	-0.3100
Constants	0.428	5.4500
Number of observations	5,933	