Data and Methods Report
August 2022

Florence Baillet | Thorsten Euler | Andreas Franken | Henrike Schmidtchen | Anne Weber

DZHW Graduate Panel 2009

Data and Methods Report on the Graduate Panel 2009

(1st- 3rd Survey Waves)

fdz.DZHW
Forschungsdatenzentrum
für Hochschul- und Wissenschaftsforschung

Author(s)
Florence Baillet
Thorsten Euler
Andreas Franken
Henrike Schmidtchen
Dr. Anne Weber

Impressum

Editor
Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
German Centre for Higher Education
Research and Science Studies (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Postfach 2920 | 30029 Hannover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Management board
Prof. Dr. Monika Jungbauer-Gans
Dr. habil. Thorsten Kowalke

Chairman of the supervisory board
Ministerialdirigent Peter Greisler

Registration court
Amtsgericht Hannover | B 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

August 2022

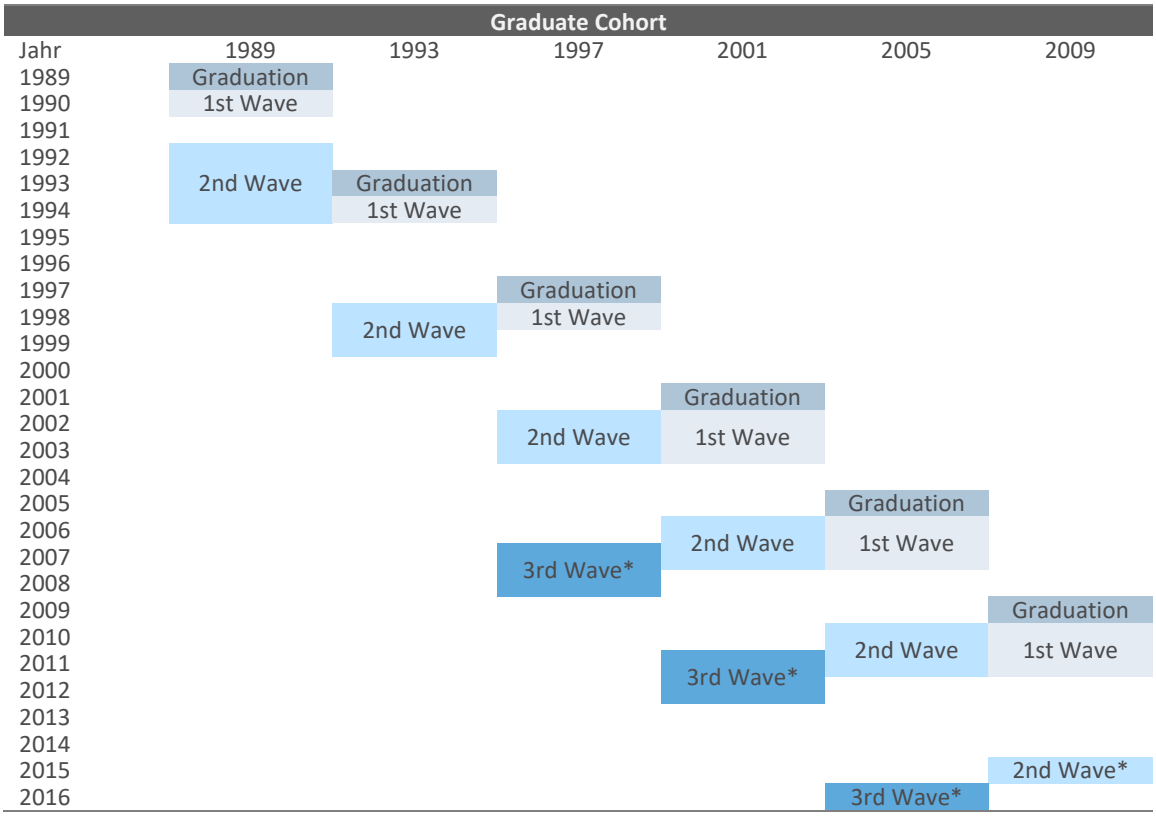
Table of Contents

List of Tables/ Table of Figures

1 Content and Design of the Studies

[Survey Series] The DZHW Graduate Panel 2009 is part of the DZHW Graduate Survey Series, which compiles information on study, career entry, career development and further qualifications of higher education graduates using standardised surveys. The first Graduate Panel was created in 1989.[1] Since then, every fourth graduate year (cohort) has been surveyed.

Figure 1: Cohort Panel Design of the DZHW Graduate Survey Series 1989-2017



*Main Survey + In-depth Surveys

The population of a cohort comprises higher education graduates who have completed a degree at a higher education institution in Germany in the winter or summer semester of the relevant examina-

[1] Since 1974, higher education graduates have been surveyed - in addition to those who discontinue their studies or change higher education institution - as part of the DZHW exmatriculation survey. This survey series have been carried out since the beginning of the 2000s under the name "Student Drop-out - Extent and Motives."

tion year.[2] For each graduate cohort, a series of survey waves are carried out, with each wave occurring at differing time intervals following the completion of degree. Thus, a combined cohort panel design is used (cf. Figure 2).

The surveys of the graduate cohorts from 1989 and 1993 comprised two waves. Since 1997, a third survey wave has been carried out. The first survey wave takes place respectively a year after graduation. The second survey wave follows approximately five years after graduation. Ten years after graduation, a third survey wave is introduced. In some cases, the second or third waves consist of a main survey and separate in-depth surveys on specific themes.

The various surveys are carried out as a written postal paper-and-pencil interview (PAPI), but also increasingly in the form of an online survey (Computer Assisted Web Interview; CAWI) (cf. Table 1).

Table 1: Outline of the DZHW Graduate Survey Series from 1989 to 2017

Wave	Survey Time Interval	Thematic Focus	Graduate Cohort			
			1989 and 1993	1997 and 2001	2005	2009
1	ca. 1 year after gradua-tion	Study progress and experience, further academic qualifica-tions, career entry	Paper & Pencil	Paper & Pencil	Paper & Pencil	Paper & Pencil
2	ca. 5 years after gradua-tion	Current occupation, employment, aca-demic and profes-sional further train-ing	Paper & Pencil	Paper & Pencil	Paper & Pencil	Online[b]
3	ca. 10 years after grad-uation	Current occupation, employment, aca-demic and profes-sional further train-ing, family circum-stances	---	Paper & Pencil[a]	Online[b]	Online (planned)

a The main survey was carried out as paper & pencil and the in-depth surveys as an online survey.
b The main survey as well as the in-depth surveys were carried out as an online survey.

The survey instruments for all cohorts contain questions on study, transition to career, further academic and professional training as well as employment, socio-demographic and educational biographical characteristics. The thematic focus of survey waves is oriented to the respective typical education, career and life phase of those surveyed at the time of the survey.

[Analytical Potential] Key information is collected in each survey wave for all cohorts. Using this information, long-term trends in higher education and labour market development can be surveyed using time series and cohort comparisons. As some of the questions in the various survey waves are repeated within a cohort, this enables the observation of intra-individual changes between the waves (e.g. causal panel analyses). It should be emphasized that continuous monthly data on individual occupational progress since graduation are generated for all cohorts across waves, which is well suited to Event History Analysis and Sequence Analyses. Moreover, some aspects can be surveyed in-depth or as a complement, depending on current developments and research interests in individual cohorts.

[2] For the 1989 cohort exclusively graduates from the federal states of the former Federal Republic of Germany were chosen.

[Research Field] The sample and survey design as well as related analysis options distinguish the DZHW graduate series from other studies of graduates carried out in Germany. For example the Bavarian Graduate Panel (BAP) of the Bavarian State Institute for Higher Education Research and Planning (IHF) is restricted to graduates of Bavarian universities.[3] The Graduate Survey Cooperation Project (KOAB) of the International Centre for Higher Education Research (INCHER) surveys graduates of its partner universities and enables individual analyses on higher education and study programme level, which can be used for evaluation and further development.[4]

[Particularities of the Graduate Panel 2009] Along with the general characteristics of the survey series, the survey under consideration here of the 2009[5] graduate cohort, for which there is currently data available from two main surveys and two in-depth surveys on the topic of 'PhD/doctorate' and 'regional mobility'[6], exhibits the following specifics. As with the 2005 graduate cohort, the study phase of the 2009 cohort is characterised by the change of higher education institution within the framework of the Bologna Process and distinguished by the career entry phase through to the economic and financial crisis of 2008. In contrast with previous cohorts of higher education leavers, the 2009 cohort indicates, along with graduates in the traditional subjects, a high number of people graduating with bachelor's degrees, which accounted for 22.9% of overall graduates in the 2009 examination year (see Dudek, Glässner & Schröder, 2010, p. 31).[7] This enables a comparison between bachelor's degree graduates and graduates from traditional degree courses. On the other hand, one can also examine whether the bachelor's degree graduates start their careers upon graduation, embark on a master's degree course or combine both of these options.

The contents of the survey tools from the main surveys of the 2009 panel of graduates are based heavily on the surveys of the 2005 panel of graduates and provide corresponding opportunities for comparison. However, more detailed information overall was gathered in the second wave through the two in-depth surveys than in the two previous survey waves. One methodological reform in the 2009 graduate cohort, when compared to previous cohorts, is that the second survey was conducted online for the first time.

The survey of the third wave of the 2009 Graduate Panel was conducted as one of three modules within the framework of the joint project "Nationwide Graduate Panel (buwap)"[8]. The other two modules comprised the second survey wave of the 2013 Gradute Panel and the first survey of the 2017 Gradute Panel.

---

[3] cf. http://www.bap.ihf.bayern.de

[4] cf. http://koab.uni-kassel.de

[5] The population consists of higher education graduates who completed their first professional degree qualification in the winter semester of 2008/2009 or in the summer semester of 2009 at a state approved higher education institution in the Federal Republic of Germany (with the exception of graduates from of German Armed Forces universities, technical universities of administration, part-time or distance learning degree programmes). As part of the first survey wave, a small special random sampling with master's degree graduates was also taken into consideration. The actual number of cases, however, was so small that any analyses based on the different subjects in particular are inappropriate. The master's graduates were therefore removed from the data set and were also given no further consideration in the following documents.

[6] The data from the second survey wave (main survey and in-depth survey) as well as the spell data from both waves cannot be published until October 2017 due to embargo periods. A third survey wave is planned for 2019.

[7] In the 2005 graduate cohort, the percentage was still 4.4 % (see Dudek, Glässner & Krause, 2007, p. 25). Therefore, graduates were only considered from subjects for which there was already a higher number of graduates. For this reason, the random sampling of bachelor's graduates from the 2005 cohort is not suitable as a basis for drawing any conclusions about this group.

[8] https://www.dzhw.eu/forschung/projekt?pr_id=606

2 Survey Instruments

In the first survey wave of the 2009 graduate panel, a standardised paper questionnaire was used as a survey instrument. In the main survey of the second survey wave, a standardised questionnaire was used in the form of an online and a paper version (see Chapter 4). Both in-depth surveys of the second survey wave were carried out exclusively via a standardised online questionnaire. All the survey instruments deployed were issued in the German language.[9]

Chapter 2.1 introduces the main contents of both survey instruments. Chapter 2.2 describes the pre-tests carried out to improve the questionnaires.

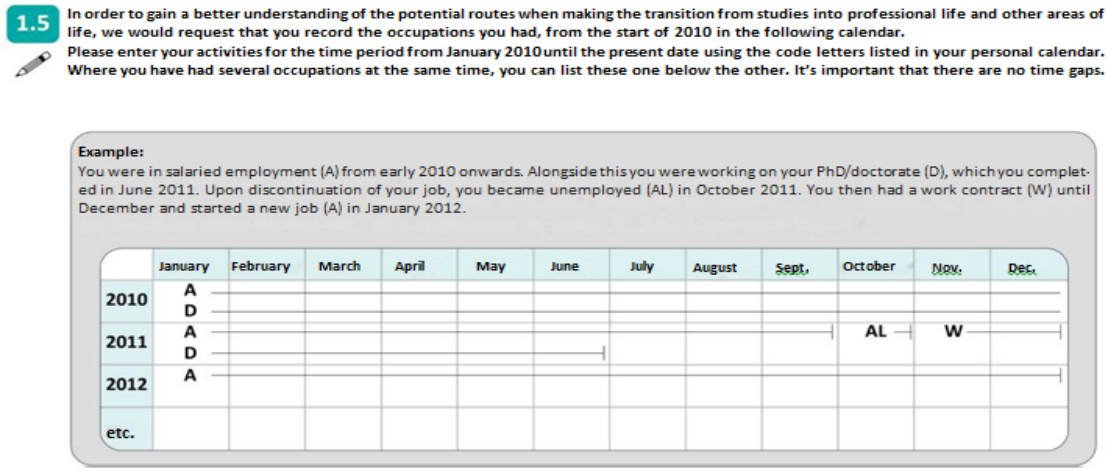
2.1 Contents of the Survey Instruments

[Characteristics of the Survey Series] The focus of the Graduate Panel 2009, as with the other co-horts in the graduate survey series, is the transition from higher education to employment and the relationship between study and career success. The starting point of the survey instrument is a re-view of the respondent's course of studies. Information is gathered on study progress and study success, on the evaluation of study circumstances as well as on qualifications gained at the higher education institution. [10]Next, information on the career of the graduates is asked.

For each of the waves, the occupation trajectory of the graduates is recorded since graduation. For each occupation (e.g. employment, PhD, parental leave), the respective spell type is recorded along with the month in which the occupation began and ended. This has been carried out since the first wave of the 2001 cohort and the second wave of the 1997 cohort in the form of a Calendar of Occu-pation (Question 4.7 in Wave 1 and Question 1.7 in Wave 2 (Paper Version) or Page 5 (Online Ver-sion) in Wave 2) for which the respondents enter their individual occupations (cf. Figure 3).[11] The Calendar of Occupation was designed by DZHW to minimise incomplete answers in the description of occupational progress

[9] The questionnaires and question flow diagrams can be downloaded in the FDZ-DZHW metadata search system (https://metadata.fdz.dzhw.eu) (from October 2017 for the second wave).
[10] cf. Section 1 "Study Progress and Study Experiences" in the questionnaire of the first wave.
[11] Before the introduction of the calendar, occupational activity was recorded using a tableau

Figure 2: Calendar of Occupation: DZHW Graduate Panel 2009, 2nd Wave, Paper Version[12]



1.5 **In order to gain a better understanding of the potential routes when making the transition from studies into professional life and other areas of life, we would request that you record the occupations you had, from the start of 2010 in the following calendar.**
Please enter your activities for the time period from January 2010 until the present date using the code letters listed in your personal calendar. Where you have had several occupations at the same time, you can list these one below the other. It's important that there are no time gaps.

Example:
You were in salaried employment (A) from early 2010 onwards. Alongside this you were working on your PhD/doctorate (D), which you completed in June 2011. Upon discontinuation of your job, you became unemployed (AL) in October 2011. You then had a work contract (W) until December and started a new job (A) in January 2012.

	January	February	March	April	May	June	July	August	Sept.	October	Nov.	Dec.
2010	A D											
2011	A D									AL	W	
2012	A											
etc.												

A	Paid work, non self-employed (as employee, civil servant)	B	Professional education, retraining, traineeship	F	Training, further education (full-time, permanent)
SE	Self-employed work (excluding contract/freelance work)	P	Internship	AL	Unemployment
T	Trainee	R	Traineeship in a school, law firm ('Referendariat'), mandatory internship ('Anerkennungspraktikum') etc.	EZ	Parental leave
J	Jobbing	ST	Course of study / studies	H	House wife/husband, family work
W	Contract for work, freelance work	D	PhD/doctorate	SO	Other (e.g. military service/alternative civilian service ['Zivildienst'], extended vacation, illness)
		JP	Junior professor/habilitation		

As a complement to the calendar, additional information on individual types of occupation is gathered. A large part of the questions refers to employment. Introduced at the same time as the calendar, the employment tableau (cf. Question 5.4 in Wave 1 and Question 4.5 (Paper Version) or Page 19 (Online Version) in Wave 2 was used to gather all declared employment information regarding the time period, the type of employment relationship, the work hours, the professional position and the place of work. For first and last employment, further characteristics are recorded, including income and sector. [13] In addition, the survey instruments also contain subjective measures to denote the respondent's employment situation, including job satisfaction, future career prospects and the adequacy of the occupation. [14]Much of the information on employment is repeatedly collected throughout the survey waves. [15] With reference to transitioning to the career, additional questions are asked on special types of employment in second training phases, e.g. traineeships or vicariates, as well as on the method of finding the career, e.g. job searches.

Besides employment other types of occupation are also handled in-depth. Questions are asked for instance on further academic qualifications, including PhDs as well as further professional training.

[12] Die Onlineversion des Kalendariums der zweiten Welle findet sich in den Screenshots auf Page 5.
[13] Before the introduction of the employment tableaus - regardless of the type of institution - additional information was generally only collected on first and last employment.
[14] See also Kerst, Fehse 2007.
[15] The wording was partly changed. Furthermore, the repeat measurement in the second wave with career change refers to a different career situation or employment than in the first wave.

Furthermore, various socio-demographic and educational biographical attributes are recorded, including parental educational background, family status and children.

[Particularities of the Graduate Panel 2009] As with the 2005 graduate cohort, the study phase of the 2009 cohort was also characterised by the structural reforms to studies within the framework of the Bologna Process and the career entry phase was influenced by the economic and financial crisis of 2008. For this reason, the content of the survey instrument in the first survey wave was based heavily on the Graduate Panel 2005 survey. Thus only particular questions within the existing topic blocks were changed or newly incorporated. For example, additional questions were asked about further academic qualifications (see questions 2.2, 2.3 and 2.5), internship (see questions 4.12, 4.13 and 4.17) and career paths (see question 5.8) as well as about career and life goals (see question 5.19).

In comparison with previous cohorts, the second survey wave included a methodological reform in that it was carried out online for the first time. In this regard, it was divided into a main survey [16] and two additional in-depth surveys on the topics of 'PhD/doctorate' and 'regional mobility'. The collection instrument for the main survey is once again based on the survey from the second wave of the Graduate Panel 2005. Additional questions were incorporated, in particular regarding professional training and further education (see question block 6) as well as on current professional activity (questions 4.7, 4.11, 4.13, 4.16 and 4.17). In the in-depth survey on the doctorate/PhD, doctoral candidates were asked about their experiences during the PhD phase, for example on their reasons for doing the PhD, the institutional frameworks or the mentoring set-up during the PhD. The data collected in the in-depth survey on regional mobility included, among other things, living history since graduation, mobility placements and purposes as well as periods of stay abroad. [17]For the in-depth surveys some questions were taken from the question instrument used in the previous two survey waves, but for the most part additional questions were asked. Overall, the second survey from the Graduate Panel 2009 is significantly more detailed than the two surveys waves of the earlier cohorts. While there was also an in-depth survey on the doctorate in the third waves of the 1997 and 2001 cohorts respectively, mobility hadn't been highlighted in any of the cohorts to date. The survey contents of the main survey from the third wave of the 2009 Graduate Panel is predominantly based on the second wave of the 2009 Graduate Panel and the third wave of the 2005 Graduate Panel. This ensures comparison both within and between cohorts. In addition to the survey of life histroies to the month since the last survey, the instruments included the current occupational situation, information on current employment, further education needs and the private life situation. In contrast to the third wave in 2005, all further academic qualifications since graduation in the examination year 2008/2009 werde addiotionally asked and for respodents with a doctoral phase, addi-

[16] For the main survey there was also a paper version of the questionnaire (see chapter 4).

[17] It should be pointed out that the Big Five Inventory (BFI-10) short scale was used in the mobility survey, see also:

- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41 (1), 203–212. doi:10.1016/j.jrp.2006.02.001
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of sociodemographic effects based on a German population representative sample. European Journal of Psychological Assessment, 23 (3), 193–201. doi:10.1027/1015-5759.23.3.193
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). A short scale for measuring the five dimensions of personality. 10 Item Big Five Inventory (BFI-10). methoden, daten, analysen, 7(2), 233–249. doi:10.12758/mda.2013.013
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). Big Five Inventory (BFI-10). Zu-sammenstellung sozialwissenschaftlicher Items und Skalen (GESIS – Leibniz-Institut für Sozialwissenschaften, Hrsg.). doi:10.6102/zis76

tional information on this was collected. This took place because no independent in-depth survey on the topic of "doctorate" was planned for the third wave.

2.2 Pre-tests

[Goal and Procedure] The survey instruments of the first survey wave and the main survey of the second and third survey waves were examined in the preliminary stages of the survey through pre-tests.[18] First, it needed to be tested whether respondents of the graduate cohort 2009 would perceive the question and response categories used in the survey instruments of previous cohorts in a similar way as previous cohorts. Secondly, it needed to be examined whether the survey instruments were also well suited to the survey of the new Bachelor graduate group (cf. Chapters 1 and 3). Thirdly, the comprehensibility and an-swerability of the new questions needed to be tested. The inclusion of new questions led to changes in the construction, layout and duration of the questionnaire that further needed to be evaluated. And fourthly, in the second survey wave, the implementation of the online version of the survey had to be tested.

A so-called field pre-test procedure was used to examine these different aspects in the first survey wave. The aim of this procedure was to study the people taking part in the pre-test "under as similar conditions as possible to those planned for the actual survey" (Häder 2015, p. 396). The collection instrument from the main survey of the second survey wave was examined within the context of expert reviews (see Häder 2015, pp 406–407).

[Test subjects]. In the 1st survey wave, employees of the DZHW having graduated from higher education in the 2009 survey year or an adjacent year were selected as test persons. By surveying these test subjects, the involvement of experts in the field of higher education research was simultaneously achieved. Through personal or project-related contacts, various external persons were recruited. Between 10 and 15 persons participated in the pre-test for the 1st survey wave.[19] In the second survey wave, the survey instrument was tested by around five members of staff [20] from the DZHW as higher education institution research experts.

[Implementation] Both pre-tests took place around two to three months before the respective survey began. The test persons were asked to complete the standardised questionnaire intended for the main survey and make a note of comprehension problems, other criticisms or remarks. In connection with the completion of the questionnaire a list of questions on various aspects of the survey was given to the test subjects. Information was gathered on completion time, content and length of the questionnaire, construction and layout, clarity of the questions and instructions for completion of the questionnaire as well as completeness of the answer options. Concrete enquiries as to new or altered questions were made.

The expert evaluations in the second survey wave were iteratively gathered in two stages. The first step involved an examination of the individual questions in terms of content. After the resulting amendment proposals had been included, the online version of the questionnaire was programmed and the paper version of the questionnaire composed. These were then examined in turn using a questionnaire for duration, structure, completion instructions, content, response options as well for the specific topic block of further education.

Based on the pre-test results, the formulations of various questions and item texts were refined, the order of particular questions and items and answer categories revised, certain questions and items

[18] There were also pre-tests for the survey instruments oft he two in-depth surveys, but only to a lesser extent (approx. 5 test persons).
[19] The exact number of participants can no longer be reconstructed.
[20] The exact number of participants is no longer reconstructable.

deleted or newly incorporated, and the layout adjusted accordingly.[21] It should be pointed out that, in this regard, the filter management within the further education topic block in the second survey wave was designed differently in the online version than in the paper version. The basic structure and scope of the questionnaires remained unchanged. For the third wave of the survey, additional findings from the pre-tests and instrument developments of the National Academics Panel Study project (Nacaps) at the DZHW, which took place in parallel, were integrated. These related primarily to the formulation of questions about doctoral studies.

[21] The specific reasons for the changes can no longer be reconstructed.

3 Population and Sample Procedure

[Population] The population of the Graduate Panel 2009 comprises all higher education graduates who completed their first professionally recognised degree at a state approved higher education institution in the Federal Republic of Germany in the winter semester of 2008/2009 or in the summer semester of 2009.[22] Higher education graduates of German Armed Forces universities, technical universities of administration, vocational academies and distance learning universities were excluded.

[Sample Procedure] Due to the lack of or non-access to the lists of higher education graduates, the individuals had to be recruited via the higher education institutions themselves. To this end, the particular degree courses were initially forecast for each federal state based on the available first-year student numbers and the duration of study for all students at universities and universities of applied sciences ('Fachhochschule'), for which graduates would be likely to register for the 2009 examination year.[23] Two separately layered cluster samplings were drawn from this, on the one hand for higher education graduates of traditional degree programmes [24] and for higher education graduates of bachelor's degree courses on the other.

[Stratified Cluster Sample] The 'primary sampling units' (clusters) from the cluster samplings were defined on the basis of the higher education institution and the area of study.[25] The 'secondary sampling units' constituted the higher education graduates from the 2009 academic year within these clusters.[26] This design was implemented by drawing basic random samplings. The clusters were thereby layered according to region (old federal states, new federal states including Berlin). The aim within the respective layers of the random samplings was to achieve a distribution proportionate to the population. As a result of the survey, however, there were slight disproportionalities among the higher education graduates from traditional degree programmes, so that a relatively high number within this group appeared in the random sampling in eastern Germany. This distortion arising from the survey design can be counterbalanced by using the appointed calibrated design weights (see chapter 7).

---

[22] The examination year 2009 began in September 2008 and ended in August 2009 at universities of applied scienc-es. At universities, it began in October 2008 and ended in September 2009.

[23] On the basis of near consistent graduate numbers, it would have been possible to take the figures from the previous examination year. However, the switch to the new degree courses resulted in too great a difference in the statistics between the academic years.

[24] This includes graduates from diploma, master's and State Examination (including teacher training) courses as well as graduates with ecclesiastical and artistic degrees.

[25] Corresponding to the classification based on the official statistics (according to the Key List of Student and Exami-nation Statistics Winter Semester 2008/2009 and Summer Semester 2009).

[26] Example of the group of people in a cluster: all bachelor's degree graduates in the area of physics at university A or all diploma graduates of economic sciences at the university of applied sciences B

In the absence of a cluster (e.g. in the event of refusal of participation at higher education institution or faculty level) – with regards to the characteristic combination of area of study, type of higher education institution, degree type and region – the most similar cluster as possible was sought as a substitute. In the event of multiple clusters with similar characteristic combinations, the biggest cluster was chosen.

It should be pointed out that the bachelor's sampling also included bachelor's graduates aiming for teaching posts. These normally always require a master's degree to enter the profession and therefore differ from the other bachelor's graduates. This should be taken into account in any analyses.[27]

[27] This concerns cases for which the following applies: inlist(astu021f_g1,4,5) & astu03a=1

4 Implementation of the Surveys

[Maintenance of Contacts and Addresses] The DZHW wrote to the sampled universities and asked them to participate. In addition the DZHW informed the universities and their examination offices of the criteria with which they could identify target persons for the Graduate Panel 2009 (e.g. examination year, first completed degree, area of study, type of degree; cf. Chapter 3).[28] In total more than 400 examination offices were contacted. Since the universities were not allowed to give out contact details of their graduates for data protection reasons, they merely informed the DZHW of the respective total number of graduates. Consequently the DZHW sent the appropriate number of survey papers for the first survey wave by post to the relevant examination offices that forwarded the survey documents to the target persons.

In order to contact persons willing to participate in the second survey wave directly through the DZHW, their contact details (postal and e-mail address) were recorded in the first wave questionnaire. Upon receiving a completed questionnaire at the DZHW, a unique identification number was stamped upon the questionnaire and also on the address section of the questionnaire. After compiling all address sections, a reference list of the identification numbers belonging to the respective addresses was generated.[29] The address lists between the survey waves were checked and updated where appropriate, so that those people were also taken into account whose e-mail addresses were still unavailable after the first survey or which had changed in the meantime or who had moved house.[30]

[Survey Documents] The survey documents for each person to be surveyed in the first survey wave consisted of a postal address (incl. data protection information), the paper questionnaire, a flyer with key information on the study and a postage paid envelope addressed to the DZHW for returning the completed questionnaire. In addition two reminder letters were sent.

The survey documents for the main survey in the second survey wave differed according to delivery method. If a valid e-mail address was available, a letter was sent out with a link to the online survey, an individual password (token) and a link to the data protection page. The letter (incl. link and password for the survey), data protection information and a flyer were otherwise sent by post. This in-

[28] At this point, it is possible that the examination offices also identified persons as belonging to the sample who did not belong to the population (i.e. overcoverage), if for example they belonged to another graduating year than 2009 (e.g. with delays in certificate production) or if it was not the first completed degree.

[29] To guarantee data protection, the address section was separated from the questionnaire and the reference list separated from the survey data and saved on a secure server.

[30] The respondents were therefore contacted in writing both following the first wave – as party of an information letter regarding the results of the survey – and before the second wave, and asked to update their addresses. The addresses of non-deliverable postage items were updated via the Deutsche Post address updating service and the register of residents' information provided by RISER ID Services GmbH. Within the filed phase of the second wave, the DZHW also carried out some further address checks whenever any survey documents were found to be unde-liverable. In preparation for the third wave, further address updates were made upon completion of the second wave in the course of the renewed sending of the results.

cluded a paper version of the questionnaire with a prepaid return envelope, so that respondents could decide which survey mode could be used for their survey participation. Three reminder e-mails and two reminders by post were also sent.

The survey documents from both in-depth surveys, which were carried out exclusively via online survey, each included – both for e-mail and postal deliveries – a letter with link and individual password for the survey. A reminder letter was also sent for each in-depth survey.

For the main survey of the third survey wave, which was also conducted exclusively via online, both the postal and the e-mail addresses of the respodents were used to contact them. First of all, the graduates, whose postal address was available, a postal invitation was sent out with a link to the online survey and an individual password (token). One week later, a further invitation was sent by e-mail (incl. link and password for the survey) to all those graduates, whose e-mail address was available. One reminder by post and three reminder e-mails were also sent. The number of contacts per person thus depended on the contact information available in each case and varied accordingly from two (postal only) to six (postal & e-mail) contacts.

[Fieldwork Phase] The time period of the first survey wave extended from February 1, 2010 to January 31, 2011.[31] Both reminder letters were sent respectively at four and eight weeks after the fieldwork phase began. Due to the contact procedure initiated through the examination offices of respective higher education institution, the DZHW could not directly influence the exact point in time that survey documents were delivered.[32] Likewise the reminder letters were sent to all persons in the sample – including those who had already completed the survey – since the examination offices had no knowledge which persons had already sent a questionnaire back to the DZHW.

The survey time period of the second survey wave extended from February 16, 2015 to October 2, 2015.[33] The fieldwork phases of the in-depth surveys, to which only participants in the main survey from the second survey wave were invited[34], took place from 4 June 2015 until 15 July 2015 (in-depth survey PhD/doctorate) and from 16 July 2015 until 31 August 2015 (in-depth survey mobility). As the DZHW now possessed the address list of respondents, exact dates for the survey documents' delivery could be specified.[35,36] In addition, the reminder letters only targeted the persons who had not yet participated in the survey.

[31] The fieldwork time was extended for as long as possible – and in parallel to project objectives – so that every questionnaire received until January 2011 was included. In the Metadata Search Portal, the term "field period" is used.

[32] After the first covering letters were delivered, the examination offices reported their respective delivery dates. Four weeks following this date, the DZHW then sent the first reminder letter to the examination office, which was then forwarded to the respective targeted persons. The examination office in turn reported the delivery date of this reminder letter. The procedure for the second reminder took place also in this manner.

[33] The fieldwork time was extended for as long as possible as well.

[34] Also for the doctorate in-depth survey, only those participants were invited who had indicated in the main survey that they had taken on a PhD/doctorate course upon graduation.

[35] Main survey: Invitation: 16.02.2015; first reminder: 26.02.2015 (e-mail), 05.03.2015 (via post); second reminder: 12.03.2015 (e-mail), 19.03.2015 (via post); third reminder: 20.03.2015 (e-mail), not via post.

[36] In-depth survey doctorate: invitation: 04.06.2015; reminder: 18.06.2015; In-depth survey mobility: invitation: 16.07.2015, reminder: 30.07.2015

The data for the main survey of the third survey wave was collected from 12 April 2019 to 3 July 2019. It was also possible here to set specific delivery times[37] and to send targeted reminders to non-participants.

[Measures to Increase Response] The measures for increasing the response rate were aimed, on the one hand, at encouraging the higher education institutions to provide organisational support for the survey during the initial contact, and at the individual motivation of respondents on the other.

The higher education institutions received an accompanying letter from the BMBF with the correspondence from the DZHW, in which the overall social importance of the survey was emphasised. Following the first wave, a summary of the results was also sent out to the higher education institutions with a view to subsequent surveys. This was sent both to higher education institutions which had taken part in the survey and higher education institutions which had refused to participate.

Along with the reminder letter method used in the main and in-depth surveys, other measures for improving response rates were also deployed at respondent level, especially in the two main surveys. Thus it was already mentioned in the letter that a summary of the key findings of the study would be sent out following the survey. In the first survey wave, the letter was also accompanied by a flyer, while in the second wave this was only sent out by post (without flyer). Furthermore, information on the project and the resulting publications was also provided on the project homepage. In addition, a prize draw was also held among all survey participants for various non-cash prizes. In the first survey wave, a notebook, five iPod nanos and five USB sticks were raffled.In the second wave a notebook, ten iPod nanos, 15 rail vouchers worth 50 EUR each and 20 USB sticks andin the third wave, a tablet PC, a smartphone and rail vouchers worth 2,000 EUR, a current tablet, a high-quality smartphone and travel vouchers worth 2,000 EUR were awarded.

[37] Main Survey: invitation (via post): 11.04.2019; invitation (e-mail): 18.04.2019; first reminder 02.05.2019 (e-mail), 16.05.2019 (via post); second reminder: 23.05.2019 (e-mail); third reminder: 06.06.2019 (e-mail)

5 Response Rate

[Response Rate] The gross sample from the first survey wave consisted of 52,550 graduates, who were registered at and contacted by the examination offices of the higher education institutions drawn from the sampling (see chapter 3) including 27,702 bachelor's degree graduates and 24,848 graduates from traditional degree programmes. These came from 237 higher education institutions from every federal state.

Once those questionnaires had been removed which were returned to the DZHW from respondents not belonging to the population (neutral sample attrition) or which could not be evaluated (relevant sample attrition) (see chapter 6.3), there were 10,494 remaining cases in the net sample of the first survey wave, including 4,883 bachelor's degree graduates and 5,611 graduates from traditional degree programmes. With regard to the gross sample, the overall response rate was around 20 per cent. The figure for graduates of traditional degree programmes (22.6%) is somewhat higher than that for bachelor's degree graduates (17.4%).

9,086 persons, over 86 per cent of the 10,494 participants in the first wave, were contacted before the second survey in the course of updating addresses and informed about the planned survey. Those comprise the gross sample of the second wave. Of these 9,086 cases, 9.004 could be invited to the survey.[38] Due to various sample relevant attrition (e.g. non-participation or questionnaires which could not be evaluated), the net sample of the second wave amounts to 4,755 cases, of which 2.110 are Bachelor graduates and 2,645 are graduates of traditional courses of study. With reference to the 9,086 cases in the gross sample, the response rate lies at 52.3 percent. As only those persons who had agreed to further contact in the first wave were invited to the second wave, the response rate of the second wave is significantly higher than that of the first wave (cf. Table 2).

For the in-depth survey "mobility", all participants of the second survey were contacted again. The gross sample amounts to 4,755 cases. A net sample of 2,465 cases could be real-ized, which results in a response rate of 51.8 per cent.

The in-depth survey "PhD/Doctorate" only comprised those 1,136 persons, which had stated having started a PhD after graduation in the second survey.[39] This amounts to a net sample of 676 cases and therefore a response rate of 59.5 per cent.

The basis for the main surve of the third survey wave were ther participants of the second survey wave who had provided their personal contact addresses in previous survey and had agreed to be contacted again with the framework of the panel. A total of 4733 people were contacted and invited to take part in the survey. 3,642 evaluable questionnaires were received by the end of the field phase, resulting in a response rate of 76.9%.

[38] Attritions are due to non-participation for following waves as well as invalid addresses.
[39] if inlist(bfec12,1,2,3,4); for one case, that took part in the PhD survey, bfec12=5 applies. How this exception arose can no longer be reconstructed. The case was not deleted.

[Panel Attrition] Consideration over time shows that the gross sample in the second wave only amounts to around 17 per cent of the gross sample in the first wave; in the main survey of thethird survey wave, it is only about 9 per cent.Of the 10,494 cases in the net sample of the first wave, around 54 per cent in the second survey wave and 34,7 per cent in the third wave were surveyed again (cf. Table 2). In comparing the net sample of the second wave with the gross sample of the first wave, only 9 percent of the initial gross sample participated in both survey waves. With the in-depth surveys, the response rates are accordingly lower. In addition, only just under 7 per cent of the first gross sample participated in the main survey of the third survey wave.

The Graduate Panel 2009 is further subject to attrition processes[40] typical for panel data. Refusing to participate in further surveys (e.g. no disclosure of address for contact in the second wave) or not participation after (attempted) contact in the second survey wave are a few examples. Furthermore, attrition due to contact difficulties (e.g. change of address) arise immediately after the completion of studies as high mobility of graduates is to be expected (cf. Fabian, Briedis 2009, pp. 71)

Table 2: Gross and Net Samples und Response Rates of the DZHW Graduate Panel 2009

	Wave 1	Wave 2 (main survey)	Wave 2 (PhD)	Wave 2 (Mobility)	Wave 3 (main survey)
Gross sample	52.550	9.086	1.136	4.755	4733
Net sample	10.494	4.755	676	2.465	3.642
Response rate	20,0 %	52,3 %	59,5 %	51,8 %	76,9%
Proportion gross sample of gross sample wave 1	--	17,3 %	2,2 %	9,0 %	9,0%
Proportion net sample of net sample wave 1	--	45,3 %	6,4 %	23,5 %	34,7%
Proportion net sample of gross sample wave 1	--	9,0 %	1,3 %	4,7 %	6,9%

[40] For attrition processes typical for panels, cf. Schnell et al. 2005, p. 241.

6 Data Preparation

In the following sections, various steps in data preparation are described. These took place during the first and second survey waves and are identical in each wave. 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 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 anonymisation) are explained separately in Chapters 7 and 8.

6.1 Data Transfer

[PAPI Surveys] In the PAPI surveys, for the purposes of further processing, the respondents' data were transferred from the paper questionnaires to a computer-readable format using a code plan. For this purpose, the questionnaires were marked to show which questions and subquestions were to be assigned a variable, what names these variables would carry and what numerical codes should be used for the responses of the interviewees. This also involved recording numerical codes for the open responses (see chapter 6.2).[41] The variables were additionally numbered to set the order of data collection.

In the first survey wave, the code plan, other instructions for data collection and the prepared paper questionnaires were handed over to an external service provider. The data collection was performed there manually by typists. In the second survey wave, data from the paper questionnaires were recorded twice by each different DZHW staff member using the data collection software EpiData[42]. The duplicate data was collated in Stata, and rechecked and cleansed in the event of any discrepancies.

[Online Surveys] It was possible to export and process the data from the online surveys directly from the survey software as a csv. file.

6.2 Coding of Open Responses

The (semi-) open responses were only entered into the SUF/CUF in coded form. The encoding decisions made by the primary research project thereby remained unchanged. 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 graduate panels. For some variables, new code lists were developed on the basis of the entries from the Graduate Panel 2009. For some semi-open questions, no new variables with numerical codes were created. Instead, entries were only assigned to the existing (closed) response categories.

[41] In the first survey wave, preliminary manual corrections were also made to facilitate the data transfer (see chapter 6.3).
[42] With the exception of calendar data, which was simply collected.

Some of the open questions were not encoded as they were mainly collected as context information for the encoding of other open data or due to insufficient time resources. [43]

Coded topics and respective code lists are presented in Table 3. The data set contains exclusively the coded numerical variables. The open entries themselves are not contained in the data set. The values of the variables are documented in the Data Set Report as well as in the Metadata Search Portal.[44]

The encoding of the (semi-) open responses for the paper questionnaires took place as part of the data transfer (see chapter 6.1). As a general rule, the encoding for the online data was performed using Excel. Although the encoding for the PAPI data in the second survey wave was technically separate from the encoding of the online data, the data from both survey modes were constantly taken into account for the preparation of the new code lists. The single exception to this separate approach was made when performing the encoding of the open professional data in the second survey wave. In this case, the open data from the paper questionnaires and the online data were collated and encoded together using coding software, in which the classification of occupations from 2010 had been stored. Where possible, the software assigns automated job codes in the process, while codes can be added manually for the remaining entries.

Table 3: Coded Topics and Code Lists in the DZHW Graduate Panel 2009

Topics	Code List Resource	Code List-ID[b]
Subject	Destatis Subject Classification 2008/2009 [accord-ing to the Key List of Student and Examination Statistics (Winter Semester 2008-2009 and Sum-mer Semester 2009), Key 3.1]	cl-destatis-studienfach-2009c
Degree	Project's Own Coding	cl-dzhw-22
Higher education institution	Destatis Key List of Student and Examination Statistics (Winter Semester 2014-2015 and Summer Semester 2015), Key 2.2	cl-destatis-hochschule-2015d
Federal State	Destatis Federal State Codebook (corresponds to both first numbers of the Official Municipality Codebook, AGS)	cl-destatis-bundesland-1990d
Foreign / Nationality	Wave 1: Project's Own Coding Wave 2: Destatis Nationality and Region Codebook 2015	cl-dzhw-23 cl-destatis-ausland-2015
Professional Title	Profession of respondents: Destatis German Classification of Occupations 2010 Training profession/parents' profession: Destatis German Classification of Occupations 1992	cl-destatis-kldb-2010d cl-destatis-kldb-1992
Professional Area of Responsibilitya	Project's Own Coding	cl-dzhw-4
Type of School Teacher training	Project's Own Coding	cl-dzhw-21
Other open enquiries	Assignment to given categories, project's own coding or deletion	---

a cf. Question 5.2 in Wave 1 (in Wave 2 the professional area of responsibility was not coded)
b 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.
c supplemented by a code from the Key List of 2012/2013
d supplemented by project's own codes if not assignable

[43] This affected in particular the open data in the in-depth survey on mobility, but also some of the open responses in the other surveys. Thus the encoding for the study focuses (question 1.8) was omitted in the first survey wave. Furthermore, no encoding of the typical work priorities took place either in the first survey wave or the main sur-vey, which were collated along with the professional title and sphere of competence (question 5.2 in wave 1 and question 4.9 in wave 2 (paper version). The professional spheres of competence in the same questions were only encoded for the first wave.
[44] https://metadata.fdz.dzhw.eu/#!/en

6.3 Data Checking and Data Cleansing

[Software-Assisted Correction] Various consistency checks were performed for both survey waves, particularly for the two main surveys. 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 combinations of characteristics, which were already tested in the preliminary manual correction, more complex feature combinations could also be tested here.

Any identified inconsistencies were, if possible, eliminated by the comparison with other responses in the questionnaire or alternatively by assigning a corresponding missing code (see chapter 6.7).

In the first survey wave, the initial consistency checks were carried out manually on the paper questionnaires before the data transfer.[46] Following the data transfer, a comprehensive review and correction of the data took place with the aid of the DZHW's own in-house software.[47] In the second survey, the consistency check was also performed with the support of software on the one hand, and via Stata-Do-Files on the other.

[Deletion of Cases] In both waves, some cases were removed from the data set. 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.[48] These were likewise removed from the data set.[49]

6.4 Generation of Variables

In addition to the variables containing the coded answers of the respondents, the Graduate Panel 2009 also generates variables. One the one hand, this includes variables that were numerically coded from the originally open entries (cf. Chapter 6.2). On the other hand, variables were changed due

---

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 The number of corrections was not recorded centrally, but simply on the paper questionnaires, and can therefore no longer be systematically reconstructed.

47 The data captured in the questionnaires were imported into a database for this purpose. Following this, valid value ranges and response combinations were defined and checked based on formal rules.

48 This occurred for example if the examination offices mistakenly wrote to persons who belonged to another graduating year or to graduates who already had further qualifications.

49 Please note that the data set contains several cases with a graduation date several months after the actual exam-ination period of the examination year 2009. These cases were kept because several higher education institutions assign specific cases to the previous examination year if the examination was postponed without the intervention of the graduate (e.g. due to illness of the examiner). The population also contains cases with an examination date several months before the actual examination period of the graduate year 2009 which were assigned to the ex-amination year 2009 by the higher education institutions due to missing formalities (e.g. internship certificate) or if a supplementary exam was necessary. Due to these procedures, cases with a divergent graduation date were only deleted if they didn't belong to the population.

to data protection reasons (cf. Chapter 8) 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 Graduate Panel 2009 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]. In addition, there is special documentation for certain auxiliary variables generated by the primary research project with regard to the information collected on gainful employment in order to simplify and standardise ecaluations. Where possible generated variables were positioned in the data set according to the respective output variable. 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, the generated variable was inserted at the end of the data set.

6.5 Generation of the Data Sets

[Merging of the Waves] The data from the first and second waves (incl. in-depth surveys) were merged.[51] Case assignment was made using the identification numbers of the respond-ents produced in the fieldwork phase (cf. Chapter 4).

[Generation of Individual and Spell Data Set] The merged data were stored in two sepa-rate data sets. The Individual Data Set contains a large part of the survey data as well as the additionally generated variables. For this format, there is a data record for each respondent (wide format). The sequence of the variables is oriented to the sequence of related questions in the questionnaire. The Spell Data Set contains only the answers from the calendars (Question 4.1 of the 1st wave, Question 1.5 or Page 5 of the 2nd wave, page 103 of the third wave). For each respondent, one or more spells are recorded. A spell is thus defined as a time period distinguished by a specific occupation (e.g. employment or training) or other status (e.g. parental leave or unemployment). Each spell of one respondent corresponds to one data row (long format). The structure corresponds to the standard structure for spell data (cf. Scherer, Brüderl 2010, p. 1042). The spells were sorted by case, i.e. all spells of the same respondent follow each other directly. Different types of occupation in the same time period were coded as independent spells. If activities of the same type immediately followed each other, or were practised simultaneously, they were summarised as one spell. Thus it cannot be discerned from the spell data whether a spell comprised one or more activities of the same type. However, detailed information is contained in the corresponding variables of the individual data set regarding employment activity and academic qualification. The data from these variables can be connected with the spell data. Individual and spell data sets can be merged using the respondent's identification number (variable: pid).

[Data Format] All data sets are available in Stata as well as SPSS format (cf. Section III)

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

[51] It's worth pointing out that overlaps had been purposely designed in the calendars and employment tableaus of both survey waves. Thus, within the context of the first wave, the respective occupations were recorded from graduation until the time of the survey. This period, depending on whenever the respondents completed the questionnaires, fell between February 2010 and January 2011). In the second wave, the calendar and the em-ployment tableau start at the beginning of 2010, so that occupations were recorded multiple times in some cases. The overlapping months were basically omitted from the data belonging to the second wave and the data from the first wave correspondingly retained.

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. Variable and value labels are available in German and English. In the SPSS format, there is a separate data set for each language. In the Stata format, bilingual labels were created in the same data set.

[Naming Variables in the Individual Data Set] A consistent naming system was created at the RDC for the naming of variables. With the exception of the identifier variable (pid) as well as the wave variable (wave),[52] variable names in the individual data set 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 Graduate Panel 2009 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 used in both waves, names of related variables were harmonised through the assignment of an identical root.

Table 4: Thematic Areas and Abbreviations for DZHW Graduate Panel 2009 Variable Labels

Themengebiets-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
stu	studies	Studium
occ	occupation	Beschäftigung
ski	skills	Fähigkeiten
fvt	further vocational training	Berufliche Fort- und Weiterbildung
fec	further education	Aus- und Weiterbildung
dem	demographic information	demographische Informationen
abr	(experiences) abroad	Auslandserfahrung
mov	move	Umzug
per	personality	Persönlichkeit
res	residence	Wohnsitz
rsa	research activity	Forschungsaktivität
wgt	weights	Gewichtungsvariablen
sys	system variables	Systemvariablen

[Variable Labels in the Spell Data Set] Variables in the spell data set include the respondent's identification number (pid), the identification number of the respective spell (eid), activity (status) as well as the beginning and end dates of the spell time period. The latter is coded using four variables (Month: begin_m and end_m; Year: begin_y; end_y). For the episodes that originate from the calen-

[52] This contains information on case participation in both waves (participation only in the first or in both waves).

dar of the third wave, the survey project also added a source variable. This indicates whether the episode was given by the respondents in the calendar or was added by the survey project in the course of dara processing. More detailed information can be found in the special documentation.

6.7 Coding of Missing Values

The coding of missing values varies between the first and second wave compared to the third wave.

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 Graduate Panel 2009 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 (cf. Chapter 8). The fourth group comprises missing values assigned for individual items in the context of data preparation of a specific data set (i.e. item-specific missing values, including „still active“ with items astu012c and astu012d, question 1.1, 1st wave).

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

Range of Values	Code	Value Label
-999 to -990: Non-response	-999	don’t know
	-998	no answer
	-997	no answer (response category)
	-996	interview break-off
	-995	not participated (panel)
	-994	refused
-989 to -970: Not applicable	-989	filtered
	-988	does not apply
	-987	missing by design (questionnaire split)
	-986	missing by design (wave)[a]
	-985	missing by design (cohort)[b]
-969 to -950: edited missing value	-969	unknown missing[c]
	-968	implausible value[d]
	-967	anonymised
	-966	not determinable[e]
	-965	invalid multiple answer
-949 to -930: Item-specific missing value	-948	still active
-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 for 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.

In the third survey wave, the single-figure negative missing codes assigned by the primary research project werde adopted (cf. Table 6).

Table 6: Systematics of the primary research project for missing values (wave 3)

Code	Value Label
-9	not participated (panel)
-8	unknown due to subsequent episode
-7	n. participated in in-depth survey
-6	no assignment possible
-4	no episodes avilable
-3	stil active
-2	filter
-1	no answer

7 Weighting

The weighting of the data serves to balance distortions in the sampling in comparison with the defined population. The text below begins with a general introduction to the procedure applied. This is followed by a detailed description of the weighting procedure and usage instructions are provided.

[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.[53] However, they mostly result from a systematic process (Missing at Random, Not Missing at Random), which requires modelling.[54]

[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.

[Cross-sectional Weighting] Three cross-sectional weights were created for the 2009 Graduate Panel: a total weight for the evaluation of all graduates and two separate weights for the evaluation of graduates from traditional degree courses and bachelor's degree graduates.

The resulting design weight for each layer is thereby as follows[55]:

$$dwgt_{sci} = \frac{n_{sc}}{N_{sc}}^{-1}$$

Due to the lack of information on non-participants in the first wave, no comprehensive adjustment of the design weights was possible on an individual basis for the attrition process through non-participation (non-response). But a calibration was performed, which was aimed at the alignment of the design weights with the distribution of characteristics in the population. The calibration took place both for the entire test sample using marginal distributions of all bachelor's graduates as well as all graduates from traditional degree courses. The region (east/west), gender and subject groups were used as characteristics; for the total weight, the type of degree was additionally used as well.[56]

---

[53] Insofar as the loss of statistical test strength through the reduction of the sample is considered irrelevant.

[54] For the various forms of attrition processes see essentially Rubin, 1976.

[55] Whereby n_{sc} the number of clusters in a layer corresponds to N_{sc} the number of individuals in the respective layer of the population. As the clusters were collected in full, the selection probability of an individual corresponds with the selection probability of the associated cluster.

[56] The information from the population was derived from the Federal Statistical Office (Statistical data 2008/2009).

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 design weights $dwgt_{sci}$ was performed using the raking algorithm[57].

[Longitudinal Sectional Weighting] A non-response weight was also calculated for the main survey in the second wave, which uses the information from the previous respective waves to model the non-participation in the second wave. This information served as covariates (σ_t) in a probit regression model. The Probit regression model that aimed to predict the probability of participation at time $P(Res_{t+1})$. In the case of variables with missing values, these were included in the model as additional categories of the variables, so that cases with item non-response could also be included in the model. This also allowed us to test the assumption that item non-response is a significant predictor of unit non-response in future waves. A number of predictors from the first wave proved to be significant for predicting the probability of participation in the second wave. From the model, the conditional probability of participation could be derived, the inverse of which represents the failure weight for the second wave:[58]

$$NR_{i_{gew_{t2}}} = P\left(Res_{t2_i} \middle| \sigma_{t1_i}\right)^{-1}$$

The longitudinal weight for the two- or three-wave panel of the dataset (main survey) is the product of the design weight ($dwgt_{sci}$) with the non-response weight ($NR_{i_{gew_{t2}}}$):

$$wgt_{i_{t1t2}} = dwgt_{sci} \times NR_{i_{gew_{t2}}}$$

The longitudinal weight for the three-wave panel of the dataset (main survey) is the product of the design weight $(dwgt_{sci})$ with the non-response weight of the second wave $(NR_{i_{gew_{t2}}})$ and the non-response weight of the third wave $(NR_{i_{gew_{t3}}})$:

$$wgt_{i_{t1t2t3}} = dwgt_{sci} \times NR_{i_{gew_{t2}}} \times NR_{i_{gew_{t3}}}$$

Subsequently, the non-response-adjusted longitudinal weight was calibrated to population characteristics using the ranking-algorithm.[59]

[Normalisation to the number of cases in the sample] The calculated weights were normalised to the number of cases in the sample, as is usual in social science research practice.

[Trimming the weights] The initially calculated weights have a small portion of weighting factors that represent outliers. To eliminate these, all weights were subjected to trimming according to Potter (1990) (cf. also Valliant, Dever & Kreuter, 2013, p. 388). The procedure is based on the assumption that the weights follow a probability distribution (beta distribution). All those weights that lie above the 99 per cent quantile are truncated to this limit. In the following, the surplus beyond the truncation is distributed among the remaining weight.

[Instructions for Use of the Weights] The weights created for the first survey wave are illustrated in Table 7.

[57] Raking is also known as 'iterative proportional fitting' (ipf) (see Kolenikov, 2014).
[58] The logic of the procedure corresponds to Propensity Score Matching, which goes back to Rosenbaum and Rubin, 1983 (cf. Blumenstiel & Gummer, 2015).
[59] For this purpose, the same characteristics were used as for the calibration of the design weight in the first survey wave.

Table 7: Weights provided for the DZHW Graduate Panel 2009

Variable name	Description
wgt01_t1d	Cross-sectional weight 1st wave: Overall graduates
wgt02_t1d	Cross-sectional weight 1st wave: Graduates from traditional degree courses
wgt03_t1d	Cross-sectional weight 1st wave: Bachelor's degree graduates
wgt04_t1t2d	Longitudinal weight 2-wave panel (main survey): Overall graduatesl
wgt05_t1t2d	Longitudinal weight 2-wave panel (main survey): graduates from traditional degree course
wgt06_t1t2d	Longitudinal weight 2-wave panel (main survey): Bachelor's degree graduates
wgt07_t1t2t3d	Longitudinal weight 3-wave panel (main survey): Overall graduatesl
wgt08_t1t2t3d	Longitudinal weight 3-wave panel (main survey): graduates from traditional degree course
wgt09_t1t2t3d	Longitudinal weight 3-wave panel (main survey): Bachelor's degree graduates

The generated weights are probability weights that can be incorporated into Stata with the aid of .ado-specific options.[60] For each main survey, there is one total weight for the evaluation of all graduates and two weight for the separate evaluation of graduates from traditional degree courses and bachelor's degree. If only cases of a sub-sample are considered, the use of the corresponding weight is recommended. Otherweise, the use of the total weight is suggested. No weight were created for the in-depth survey. The weight wgt01_t1d is intended for evaluations for all graduates. The weights wgt02_t1d and wgt03_t1d should be used for the separate evaluation of graduates from traditional degree courses and bachelor's degree graduates.

It is important to note that weights only represent useful correction variables if the analysis model applied contains or is related to the variables used for the weighting. For this reason, weights must always be used with a focus on the analysed question. The weights generated here relate to nationwide levels. It is therefore inadvisable to carry out evaluations with these weights on the basis of individual federal states or regions.

[60] See also the Stata guide (Command: help weights).

8 Anonymisation

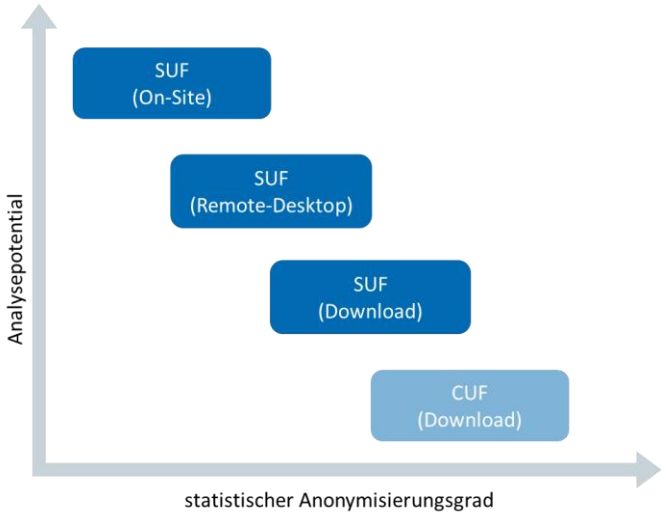
[Data Protection Legal Framework] The Federal Data Protection Act (BDSG) applies to personal data that the DZHW collected through volunteer surveys.[61] Accordingly, personal data that are collected during scientific research may be processed or used exclusively for the purposes of scientific research (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 Graduate Panel 2009, the RDC makes two types of data files available. Whereas SUF for scientific secondary use are de-facto anonymised, CUF for teaching and exercise purposes are absolutely anony-mised. 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.

While the CUF is directly transmitted by the RDC after registration, the SUF is provided using three different modes of access: download, 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/CUF) and mode of access.

[61] 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 3: Modes of Access, Statistical Level of Anonymisation and Analytical Potential of the Data of the DZHW Graduate Panel 2009



[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 email addresses, were placed in a separate data set (cf. Chapter 4) during the field phase of the Graduate Panel 2009 and are neither contained in the CUF nor in the various SUF variants. To further prevent any re-accessing of this information, the original identification number was removed and replaced with a new randomly assigned identification number.

Additionally, quasi-identifiers were determined, i.e. information which, in combination with or by the allusion to external information, allows for indirect identification.[62] For the Graduate Panel 2009, the following quasi-identifiers were used, which are present in external data sources [63] as well as in the data of the Graduate Panel: higher education institution, subject, type of degree, career information, regional information (higher education institution, location where higher education entry qualifications were obtained and place of work), nationality and country of birth. To prevent a clear association with the data of the Graduate Panel, these key attributes – according to data product and mode of access – were aggregated or deleted (cf. Table 7). For example, the attribute "higher education institution" in the SUF for on-site use becomes "NUTS-2 regions", in the remote desktop SUF it becomes "federal states" and in the download SUF and CUF it is aggregated to two categories "old vs. new federal states". Open responses are likewise quasi-identifiers (cf. Ebel 2015, p. 3) and were coded or deleted during anonymisation.

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 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 Data Protection Directive [EG-DSRL]). In the Graduate Panel 2009, information on health was collected without further consent of the respondents for secondary use. Hence, these answers were deleted in the CUF and all SUF variants.

[62] 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.
[63] E.g. student and examination statistics of the Federal Statistical Office, alumni networks of the higher education institutions or also professional networks.

To guarantee absolute anonymisation of the CUF data, more restrictive statistical anonymisation measures on the variable level in comparison to the SUF variants were performed (cf. Table 7). In addition, a randomly selected sub-sample of the data was drawn (10 percent of the surveyed graduates of traditional courses of study).

Table 8: Statistical Anonymisation Measures for the Data of the DZHW Graduate Panel 2009 by Mode of Access[64]

Characteristic	On-Site SUF	Remote Desktop SUF	Download SUF	Download CUF (Sub-sample)
Direct identifiers	Deletion and assignment of random ID	Deletion and assignment of random ID	Deletion and assignment of random ID	Deletion and assignment of random ID
Questionnaire receipt	Available	Deletion	Deletion	Deletion
Subject	Available	Aggregation to areas of study[a]	Aggregation to areas of study[a]	Aggregation to areas of study[a]
Higher education institution	Aggregation to type of higher education institution and location of higher education institution to NUTS 2: basic regions for the application of regional policies[b]	Aggregation to type of higher education institution and location of higher education institution to federal states	Aggregation to type of higher education institution and location of higher education institution to both new and old federal states	Aggregation to type of higher education institution and location of higher education institution to both new and old federal states
Place of work (federal state/abroad)	Available	Available	Aggregation to federal states and abroad	Aggregation to both old and new federal states and abroad
Place of work (postcode)	Available	Aggregation to NUTS 2: basis regions for regional political measures[b]	Aggregation to NUTS 2: basis regions for regional political measures[b]	Deletion
Place where course entry qualification was gained (federal state/abroad)	Available	Available	Aggregation to federal states and abroad	Aggregation to both old and new federal states and abroad
Place where course entry qualification was gained (postcode)	Available	Aggregation to NUTS 2: basic regions for the application of regional policies[b]	Aggregation to NUTS 2: basic regions for the application of regional policies[b]	Deletion
Place of permanent residence (postcode)	Available	Aggregation to NUTS 2: basic regions for the application of regional policies[b]	Aggregation to NUTS 2: basic regions for the application of regional policies[b]	Deletion
Place of permanent residence (federal state/abroad)	Available	Available	Aggregation to federal states and abroad	Aggregation to both old and new federal states and abroad
Occupation	Aggregation to occupational types (5 digits)[c]	Aggregation to occupational groups (3 digits)[c]	Aggregation to occupational groups (3 digits)[c]	Aggregation zu occupational main groups (2-Steller)[c]
Training profession	Aggregation to occupational orders (3 digits)[d]	Aggregation to occupational orders (3 digits)[d]	Aggregation to occupational orders (3 digits)[d]	Aggregation to occupational groups (2 digits)[d]
Parents' professi-	Aggregation to	Aggregation to	Aggregation to	Aggregation to

[64] 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)

on	occupational orders (3 digits)[d]	occupational orders (3 digits)[d]	occupational orders (3 digits)[d]	occupational groups (2 digits)[d]
Nationality (abroad)	Available	Aggregation according to NEPS classification[d]	Aggregation to world regions	Deletion
Country of birth (abroad)	Available	Aggregation according to NEPS classification[d]	Aggregation to world regions	Deletion
Age	Available	Available	Available	TOP coding[f]
Note on state of health	Deletion	Deletion	Deletion	Deletion
other open responses[g]	Coding/ Deletion	Coding/ Deletion	Deletion	Deletion

a According to the Key List of Student and Examination Statistics Winter Semester 2008-2009 and Summer Semester 2009 from the Federal Statistical Office.

b Statistical Office of the European Union (Eurostat): Nomenclature of Territorial Units for Statistics (NUTS) http://ec.europa.eu/eurostat/web/nuts/overview.

c According to German Classification of Occupations from 2010 from the Federal Statistical Office.

d According to German Classification of Occupations from 1992 from the Federal Statistical Office.

e The aggregation of states to world regions is based on the classification of the NEPS with adjustments for European countries https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf (in the first wave with adaptions for European countries).

f Age responses above a certain limit were aggregated to one category.

g Individual responses have not been coded in the on-site and remote SUF either, but were deleted completely (cf. Chapter 6.2).

9 Bibliography

Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-10459-7

Briedis, K., Brandt, G., Fabian, G. & Rehn, T. (2011). Bachelorabsolventen im Fokus. In K. Briedis, C. Heine, C. Konegen-Grenier & A.-K. Schröder (Hrsg.), *Mit dem Bachelor in den Beruf. Arbeitsmarktbefähigung und -akzeptanz von Bachelorstudierenden und -absolventen* (S. 53–82). Essen: Edition Stifterverband.

Briedis, K., Klüver, S. & Trommer, M. (2016). *Zwischen Etablierung, Stabilisierung und Aufstieg: Berufliche Entwicklung der Hochschulabsolvent*innen 2009* (Forum Hochschule 4/2016). Hannover: DZHW.

Dudek, K., Glässner, B. & Krause, I. (2007). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Sommersemester 2007* (Statistiken zur Hochschulpolitik 1/2007). Bonn: Hochschulrektorenkonferenz.

Dudek, K., Glässner, B. & Schröder, M. (2010). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Wintersemester 2010/2011* (Statistiken zur Hochschulpolitik 2/2010). Bonn: Hochschulrektorenkonferenz.

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

Fabian, G. & Briedis, K. (2009). *Aufgestiegen und erfolgreich. Ergebnisse der dritten HIS-Absolventenbefragung des Jahrgangs 1997 zehn Jahre nach dem Examen* (HIS: Forum Hochschule 2/2009). Hannover: HIS.

Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6

Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Kerst, C. & Fehse, S. (2007). Arbeiten unter Wert? Vertikal und horizontal inadäquate Beschäftigung von Hochschulabsolventen der Abschlussjahrgänge 1997 und 2001. *Beiträge zur Hochschulforschung,* 29(1), 72–98.

Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2

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

Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy,* 1(1), 2–16.

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

Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment,* 23(3), 193–201. doi:10.1027/1015-5759.23.3.193

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality,* 41(1), 203–212. doi:10.1016/j.jrp.2006.02.001

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen,* 7(2), 233–249. doi:10.12758/mda.2013.013

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). *Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis76

Rehn, T., Brandt, G., Fabian, G. & Briedis, K. (2011). *Hochschulabschlüsse im Umbruch. Studium und Übergang von Absolventinnen und Absolventen reformierter und traditioneller Studiengänge des Jahrgangs 2009* (HIS: Forum Hochschule 17/2011). Hannover: HIS.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70(1), 41–55. doi:10.2307/2335942

Rubin, D. B. (1976). Inference and missing data. *Biometrika,* 63(2), 581–592.

Saniter, N. & Siedler, T. (2014). *Door opener or waste of time? The effects of student internships on labor market outcomes* (IZA Discussion Paper No. 8141). Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).

Schaeper, H., Grotheer, M. & Brandt, G. (2014). Familiengründung von Hochschulabsolventinnen. Eine empirische Untersuchung verschiedener Examenskohorten. In D. Konietzka & M. Kreyenfeld (Hrsg.), *Ein Leben ohne Kinder* (2. Aufl., S. 47–80). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94149-3_2

Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1031–1051). Wiesbaden: VS Verlag für Sozialwissenschaften.

Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.

Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York. doi:10.1007/978-1-4614-6449-5