Florence Baillet / Andreas Franken / Anne Weber

DZHW Graduate Panel 2005

Data and Methods Report on the Graduate Panel 2005 (1st, 2nd, and 3rd Survey Waves)

Data and Methods Report

January 2021

fdz.DZHW
Research Data Centre for
Higher Education Research and Science Studies

Autor(s):
Florence Baillet
Andreas Franken
Dr. Anne Weber

With assistance from:
Andreas Daniel (Chapters 3 and 7)
Dr. Karsten Stephan (proofreader)

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

List of Figures

List of Tables

I Introduction

The DZHW Graduate Panel is a series of surveys on the career paths of higher education graduates.[1] The German Centre for Higher Education Research and Science Studies (DZHW)[2] carries out the surveys. Funded by the Federal Ministry of Education and Research (BMBF), the DZHW serves as a complement to official higher education statistics and contributes to national education monitoring. Since 1989, every fourth graduating cohort has been surveyed.

In the Research Data Centre for Higher Education and Science Studies at the DZHW (RDC-DZHW), the data of some graduate cohorts are subsequently processed and documented for the purpose of secondary use.[3] Through various modes of access, data are made available as *Scientific Use Files* (SUF) for secondary scientific use and as *Campus Use Files* (CUF) for teaching and exercise purposes. In addition to preparing the survey data sets, documentation materials on the data sets and survey implementation practices are also made publically available online.

The Data and Methods Report is part of the documentation for the first, second, and third survey waves of the Graduate Panel 2005 (doi: 10.21249/DZHW:gra2005:2.0.1). Further documentation materials for the study (e.g. data set report, questionnaires, question flow diagrams etc.) can be downloaded from the search portal of the RDC (https://metadata.fdz.dzhw.eu/#!/en).

Section II of this report presents an overview of the Graduate Panel 2005. Key information on the use of the data follows in Section III. Chapter 1 introduces the content and structure of the Graduate Panel Series until 2005[4] in general and the Graduate Panel 2005 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 anonymization practices used are presented.

[1] Current updates on the DZHW Graduate Panel can be found on the project website (www.dzhw.eu/absolventen).
[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.
[4] Only cohorts preceding the 2005 cohort are documented here.

II Overview of the DZHW Graduate Panel 2005

Study Series	DZHW Graduate Survey Series
Cohort	Graduate Cohort 2005 (5th cohort in the Graduate Survey Series)
Surveying Institution	German Centre for Higher Education Research and Science Studies (DZHW)
Sponsored by	Federal Ministry of Education and Research (BMBF)
Project Contributors (Project Leader)	Kolja Briedis, Thorsten Euler, Michael Grotheer, Sören Isleib, Karl-Heinz Minks, Nicolai Netz, Hildegard Schaeper, Fabian Trennt, Maximilian Trommer
Themes	Educational Biography Transition to Career Career Development Further Qualifications
Survey Design	Cohort-Panel-Design
Survey Data Type	Quantitative Data
Population	Graduates of higher education who completed their first professionally recognised degree in the winter semester of 2004-05 or in the summer semester of 2005 at a state approved institution of higher education in the Federal Republic of Germany (with the exception of graduates of German Armed Forces universities, technical universities of administration, vocational academies and distance learning universities)
Sample	Graduates of traditional courses of study: Quoted stratified cluster sample Graduates of Bachelor degrees: Deliberative sample
Survey Method	1st wave: Standardised self-administered survey 2nd wave: Standardised self-administered survey 3rd wave (main survey): Standardised self-administered online survey
Field Period	1st wave: January 1, 2006 to May 18, 2007 2nd wave: December 6, 2010 to February 21, 2012

	3rd wave (main survey): April 14, 2016 to June 7, 2016
Net Sample Size	1st wave: n = 11,787 (Bachelor graduates=1,622) 2nd wave: n = 6,472 (Bachelor graduates=798) 3rd wave (main survey): n = 4,279 (Bachelor graduates=519)
Response Rate	1st wave: 24.7 % 2nd wave: 60.5 % 3rd wave (main survey): 66.3 %
Data Products and Access Ways	CUF: Download SUF: Download, Remote-Desktop, On-Site
Data Set Structure	Individual data in wide-format Spell data in long-format
DOI	10.21249/DZHW:gra2005:2.0.1
Annotations	Data sets for graduates of traditional courses of study and Bachelor graduates are separated because of their different sampling procedures. The CUF contains only the data of graduates of traditional courses of study.
Further Information	https://metadata.fdz.dzhw.eu/#!/en/studies/stu-gra2005$ www.dzhw.eu/absolventen2005 https://fdz.dzhw.eu

Project Publications*

Briedis, K. (2007). *Übergänge und Erfahrungen nach dem Hochschulabschluss. Ergebnisse der HIS-Absolventenbefragung 2005* (HIS: Forum Hochschule 13/2007). Hannover: HIS.

Briedis, K. & Minks, K.-H. (2007). *Generation Praktikum. Mythos oder Massenphänomen*. Hannover: HIS.

Grotheer, M., Isleib, S., Netz, N. & Briedis, K. (2012). *Hochqualifiziert und gefragt. Ergebnisse der zweiten HIS-HF Absolventenbefragung des Jahrgangs 2005* (HIS: Forum Hochschule 14/2012). Hannover: HIS.

Euler, T., Trennt, F., Trommer, M. & Schaeper, H. (2018). *Werdegänge der Hochschulabsolventinnen und Hochschulabsolventen 2005. Dritte Befragung des Prüfungsjahrgangs 2005 zehn Jahre nach dem Abschluss.* (Forum Hochschule 01/2018). Hannover: DZHW.

*All project publications are available for download on the project website (www.dzhw.eu/absolventen2005).

Publications using the Data Set (selected)

Schaeper, H. (2009). Development of competencies and teaching–learning arrangements in higher education: findings from Germany. *Studies in Higher Education, 34* (6), 677–697. doi:10.1080/03075070802669207

Jaksztat, S. (2014). Bildungsherkunft und Promotionen: Wie beeinflusst das elterliche Bildungsniveau den Übergang in die Promotionsphase? *Zeitschrift für Soziologie, 43* (4), 286–301.

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

Kratz, F. & Netz, N. (2016). Which mechanisms explain monetary returns to international student mobility? *Studies in Higher Education.* doi:10.1080/03075079.2016.1172307

III Data Use Instructions

[Data Use Requirements] The data of the Graduate Panel 2005 are anonymised and made available by the RDC of the DZHW in accordance with the European General Data Protection Regulation (GDPR) and released exclusively for scientific use.[5] The RDC provides *Scientific Use Files* (SUF) for scientific secondary use and a *Campus Use Files* (CUF) for teaching and exercise purposes.

Requirements for the use of a SUF are an employment at a scientific institution and the conclusion of a data use agreement. Students or doctoral candidates without a position at an scientific institution must conclude a data use agreement together with a supervising staff member. In the course of concluding the contract, the RDC also checks whether there is any scientific interest in using the data. The data usage application form can be downloaded from the RDC website. Registration with the RDC is required in order to use the CUF. The CUF will then be transmitted by the RDC. A data use agreement does not have to be concluded.

[Data Access] The CUF of the Graduate Panel 2005 can be used at the local computer. The SUF is provided using three modes of access, which differ in their restrictions with respect to storage location, the opportunity for autonomous linking with external data and RDC control options.

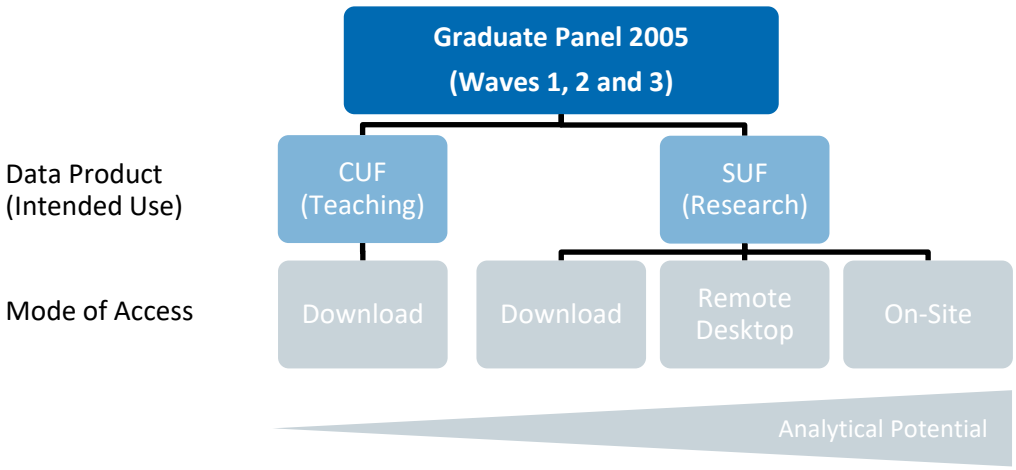
- **Download:** Data are available for download via a secure connection from the RDC website. Users can save the data on their local computer to link with data from external sources as well as perform analysis using their own software.
- **Remote Desktop:** Data are available on a RDC terminal server. 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, pp. 6, on upon which the Leibniz Institute for Educational Trajectories (LIfBi) (cf. Koberg 2016, pp. 699) and the RDC of the Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner et al. 2012, p. 9) have oriented themselves. The RDC 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 anonymization and analytical potential of the CUF and the differing SUF variants.

Figure 1: Modes of Access and Analytical Potential



[Charges for Data Access] Currently SUF and CUF are available free of charge (effective September 2019). 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 agreement. CUF may only be disclosed in the context of specified teaching activities.
- **Duty to delete:** SUF downloads must be deleted after expiry of the agreed period of use (as a rule 1.5 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.
- **Notification/Provision of Publications:** The RDC has to be immediately notified of all types of publications that are produced using data of the RDC. An electronic version of the publication shall be provided immediately.
- **Citation rules**: The data used must be cited in publications, other work (e.g. theses) and lectures according to the RDC guidelines.

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

1 Content and Design of the Study

[Survey Series] The DZHW Graduate Panel 2005 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.[8] Since then, every fourth graduate year (cohort) has been surveyed. 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 examination year.[9]

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).

Figure 2: Cohort Panel Design of the DZHW Graduate Survey Series

	Graduate Cohort				
Year	**1989**	**1993**	**1997**	**2001**	**2005**
1989	**Graduation**				
1990	1st Wave				
1991					
1992	2nd Wave				
1993		**Graduation**			
1994		1st Wave			
1995					
1996					
1997			**Graduation**		
1998		2nd Wave	1st Wave		
1999					
2000					
2001				**Graduation**	
2002			2nd Wave	1st Wave	
2003					
2004					
2005					**Graduation**
2006				2nd Wave	1st Wave
2007			3rd Wave		
2008					
2009					
2010					2nd Wave
2011				3rd Wave	
2012					
2013					
2014					
2015					
2016					3rd Wave

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

[8] 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 survey of exmatriculated students. This survey series have been carried out since the beginning of the 2000s under the name "Student Drop-out - Extent and Motives."

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

after graduation. The second survey wave follows approximately five years after graduation. Approximately ten years after graduation, the third survey wave follows. To some extent, the second and third waves consist of a main survey and separate in-depth surveys on specific topics.

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 2005

Wave	Survey Time Interval	Thematic Focus	Graduate Cohort 1989 and 1993	Graduate Cohort 1997 and 2001	Graduate Cohort 2005
1	ca. 1 year after graduation	Study progress and experience, further academic qualifications, career entry	Paper & Pencil	Paper & Pencil	Paper & Pencil
2	ca. 5 years after graduation	Current occupation, employment, academic and professional further training	Paper & Pencil	Paper & Pencil	Paper & Pencil
3	ca. 10 years after graduation	Current occupation, employment, academic and professional further training, family circumstances	---	Paper & Pencil[a]	Online[b]

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

[b] The main surveys 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.

[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.[10] The

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

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.[11]

[Particularities of the Graduate Panel 2005] In addition to the general characteristics of the study series the Graduate Panel 2005[12] has the following specific characteristics. In contrast to preceding graduate cohorts, the study phase of the 2005 cohort is defined by the transformation of higher education through the Bologna Process. In order to investigate the goals of the reform regarding the internationalisation of study and employability of higher education graduates, the survey instrument was extended to include new questions of international mobility, competence development and study organisation. With respect to the academic structural reform of the Bologna Process, a significant number of Bachelor graduates were surveyed for the first time in addition to graduates of traditional courses of study. Of the 2005 graduate cohort, Bachelor graduates made up 4.4 percent of the total number of graduates (cf. Dudek et al. 2010, p. 25).[13] To survey this group, a separate sample was drawn (cf. Chapter 3). Only graduates of subjects with a large number of graduates were considered.[14]

In addition to the changing academic environment, the graduate cohort 2005 also faced different labour market conditions than previous cohorts, whereby the beginning of the economic and financial crisis in 2008 marks their career entry. Due to the socio-political discourse on "Generation Praktikum" (“generation internship”) at the time of the survey, additional information on internships after graduation was collected for the first wave of the Graduate Panel 2005.[15]

[11] cf. http://koab.uni-kassel.de
[12] The population is made up of university graduates who obtained their first vocational qualification in the winter semester 2004/2005 or in the summer semester 2005 at a state-recognised university in the Federal Republic of Germany (with the exception of graduates of universities Federal Armed Forces, administrative colleges, vocational academies and distance learning colleges).
[13] A further 4 percent had a Master’s degree. For the 2001 graduate cohort, the proportion of Bachelor and Master’s graduates was 0.6 percent (Bachelor: 0.1 percent).
[14] Due to differing sampling procedures for graduates of traditional courses of study and Bachelor graduates, separate data sets for the surveyed groups are available.
[15] see also Briedis, Minks 2007.

2 Survey Instruments

In the first two survey waves of the Graduate Panel 2005 a standardised paper questionnaire in German was used as a survey instrument. The third survey wave was conducted as an online survey.[16] Chapter 2.1 introduces the main contents of 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 2005, as with the other cohorts 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 review 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.[17] 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 Occupation (Question 4.7 in Wave 1; Question 1.7 in Wave 2; Page 9 in Wave 3) for which the respondents enter their individual occupations (cf. Figure 3).[18] The Calendar of Occupation was designed by DZHW to minimise incomplete answers in the description of occupational progress.

[16] Questionnaires as well as question flow diagrams for all waves can be downloaded from the Metadata Search Portal of the RDC (https://metadata.fdz.dzhw.eu/#!/en).
[17] cf. Section 1 "Study Progress and Study Experiences" in the questionnaire of the first wave.
[18] Before the introduction of the calendar, occupational activity was recorded using a tableau.

Figure 3: Calendar of Occupation: DZHW Graduate Panel 2005, 1st Wave

4.7 In order better to understand the transition from studying to career and other areas of life, we ask you please to enter your occupations since graduating in the following calendar.
Please mark the month in which you achieved the final examination results of your completed studies with an **X** and enter in the calendar your occupations from the time of graduating until now using the listed code letters. If you had more than one occupation simultaneously, you can list them. It's important not to leave any time gaps.

Example:
In October 2004 you achieved your final examination results (**X**). Between November 2004 and February 2005 you first did an internship (**P**), then you were unemployed (**AL**). From March 2005 until the time of the survey you were (possibly in various jobs) in non self-employed employment (**A**). Parallel to your employment (e.g. in a doctoral post) you worked on your dissertation (**D**) from June 2005.

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
2004										X	P	AL
2005	AL		A			D						
2006	A D											

Employment		*Other Occupations*			
R	Traineeship in a school, certification internship	D	PhD	EZ	Parental leave
W	Work contract/fee work	ST	Studies	H	Housewife/husband, family work
J	Side job	P	Internship	AL	unemployed
SE	Self-employment (non-work/fee contract)	B	Vocational training, re-training, volunteering	SO	Other (e. g. military/civil service, holiday)
A	Non self-employed employment (e. g. as employee or public official)	F	Further education (full-time, longer-term)		

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 (Question 5.2 in Wave 1; Question 4.3 in Wave 2; Page 24-24h in Wave 3), 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.[19] 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.[20] Much of the information on employment is repeatedly collected throughout the survey waves.[21] 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. Furthermore, various socio-demographic and educational biographical attributes are recorded, including parental educational background, family status and children.

[Particularities of the Graduate Panel 2005] These general themes are contained in the survey instruments of all cohorts in the graduate survey series. In addition, specific questions

[19] Before the introduction of the employment tableaus - regardless of the type of institution - additional information was generally only collected on first and last employment.
[20] see also Kerst, Fehse 2007.
[21] The wording was partly changed. Furthermore, the repeat measurements in the second and third wave with career change refer to a different career situation or employment than in the first and second wave.

were included in the questionnaires of the Graduate Panel 2005 that were not asked of the preceding cohorts. More specifically, two central goals of the Bologna Process (i.e. internationalisation of study and employability of higher education graduates) were taken into consideration in the first wave. For example, a question on study experiences abroad was added to the questionnaire (Question 1.6). Additionally, information on study organisation, on forms of teaching and learning as well as on qualification and competence development was collected in greater detail than before (Questions 1.14 to 1.18).[22] To test the theory of "Generation Praktikum", information on internships after study was gathered in the first wave (Questions 4.10 to 4.16).[23] The survey instrument of the second wave includes only a few specific questions comparable to the preceding cohort (e.g. Question 8.11 and Question 8.12 on spatial mobility). In particular, many opportunities exist for comparison with the corresponding survey instrument of the 2001 cohort. Similarly, the survey instrument of the main third wave survey is completely based on the content of the 2001 cohort instrument, but in the form of an online questionnaire.

2.2 Pre-tests

[Goal and Procedure] The survey instruments were examined in the preliminary stages of the survey through pre-tests. First, it needed to be tested whether respondents of the graduate cohort 2005 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 answerability of the new questions needed to be tested. Fourthly, the inclusion of new questions and the change in the selection of questions also entailed changes in the structure and layout of the questionnaire and the duration of the survey, which were to be evaluated. In the third wave of the survey, the implementation of the online version of the questionnaire also had to be checked.

In order to test these various aspects, a so-called *field pre-test* was used in the first two survey waves. The aim of this procedure is to ensure that persons participating in the pre-test are monitored under conditions as similar as possible to those participating in the actual survey (Häder 2015, p. 396).[24] The survey instrument of the main survey of the third survey wave was examined within the framework of *Expert evaluations* (cf. Häder, 2015, pp. 406-407).

[Test subjects] In the first two survey waves, employees of the DZHW having graduated from higher education in the 2005 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

[22] Question 1.17 in this form was already a part of the survey instrument in graduating year 2001. In contrast to previous surveys, information is gathered not only on the extent to which the listed competencies are required in the career, but also on the extent to which the respondents in their own estimation possess such knowledge and competencies. In addition the three-level answer scale was replaced by a five-level scale (cf. Schaeper, Briedis 2004, p. 8).

[23] see also Briedis, Minks 2007.

[24] It should be noted that the pre-tests respondents did not all originate from the test population – as is usual with these pre-test procedures. Rather other examination years were also considered.

the pre-test for both waves.[25] In the third survey wave, the survey instrument was tested by about five DZHW employees[26] as experts in higher education research.

[Implementation] The pre-tests of the first two survey waves 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 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. In the first wave, concrete enquiries as to new or altered questions were made.[27] On the basis of the pre-test results, the formulations of various 'question' and 'item' texts were clarified, the sequences of individual questions and items and answer categories were revised, individual questions and items were deleted or newly included, and the layout was adapted accordingly.[28] The basic structure and scope of the questionnaires were retained unchanged.

In the expert evaluations of the third survey wave, it was, in particular, examined whether the questions from the established questionnaire of the wave 3 of the 2001 cohort were also appropriate in content for the 2005 cohort and whether they could be adopted.

[25] The exact number of participants can no longer be reconstructed.
[26] The exact number of participants can no longer be reconstructed.
[27] In the first survey wave, the cover letter to respondents was evaluated with respect to content, clarity and length.
[28] The specific reasons for changes can no longer be reconstructed.

3 Population and Sample Procedure

[Population] The population of the Graduate Panel 2005 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 2004-05 or in the summer semester of 2005.[29] Higher education graduates of German Armed Forces universities, technical universities of administration, vocational academies and distance learning universities were excluded.

[Sample Procedure] Due to lacking or inaccessible lists of higher education graduates, the individuals had to be recruited from the universities. This resulted in separate sampling processes for higher education graduates of traditional courses of study[30] and higher education graduates of Bachelor degrees. For graduates of traditional courses of study, a quoted stratified cluster sample was performed. The sample of the Bachelor graduates was based on deliberate sampling.[31]

[Quoted stratified cluster sample] The "primary sampling units" – or clusters – of the quoted stratified cluster sample were defined according to the higher education institution, area of study[32] as well as type of degree[33]. The "secondary sampling units" comprised higher education graduates of the examination year 2005 within these clusters.[34] The implementation of this design was achieved with a number of steps. In the first step, a multi-level random sample was taken.[35] At first, the clusters were disproportionately stratified according to old and new federal states in order to increase the size of the sample of the new federal states. The sample size of the clusters amounted to 30 percent in the new federal states (i.e. 150 clusters) and 18 percent in the old federal states (i.e. 400 clusters). Within both of these strata, a proportional stratified cluster sample was performed. This procedure enables a more exact estimation of the stratification characteristics of the population than a simple random sample. Stratification characteristics included location in federal states,[36] size and type of higher education institution[37]. The random allocation/rounding according to Cox (1987) was used for this stratification.

In the second step, there was a quoted sampling[38] of 87 further clusters[39] that partly replaced the clusters drawn in the first step and were partly considered in addition. The aim

[29] The examination year 2005 began in September 2004 and ended in August 2005 at universities of applied sciences. At universities, it began in October 2004 and ended in September 2005.
[30] This applies to graduates in courses of study with the final qualifications graduate diploma, master's degree ("Magister"), state examination (incl. teacher training) as well as graduates with church or art qualifications.
[31] For the procedure of deliberate sampling cf. Schnell et al. 2005, pp. 298.
[32] According to the official statistics (i.e. Key List of Student and Examination Statistics Winter Semester 2004/2005 and Summer Semester 2005).
[33] Here there is differentiation between a graduate diploma and comparable qualifications, e.g. a master's degree ("Magister") or various teacher training qualifications.
[34] Example of persons in a cluster: all higher education graduates of the Leibniz-Universität Hannover who achieved their graduate diploma in the area of study "Mathematics".
[35] This selection was made by the "Centre for Survey Research and Methodology (ZUMA)" in Mannheim.
[36] Some federal states were grouped together in one stratum. The following units were grouped together: Bavaria, Baden-Württemberg, Lower Saxony, North Rhine-Westphalia, the remaining western federal states (Berlin, Bremen, Hamburg, Hessen, Rheinland-Pfalz, Saarland, Schleswig-Holstein), Saxony and the remaining eastern federal states (Brandenburg, Mecklenburg-Vorpommern, Saxony-Anhalt, Thüringen).
[37] There is further differentiation between university and university of applied sciences.
[38] Quotation characteristics: area of study, type of degree and university.
[39] This draw was performed by the DZHW.

was to achieve the best possible alignment of distributions by area of study, type of degree and higher education institution in the sample to population distribution.

In the case of cluster attrition (e.g. due to refusal to participate at institutional or faculty level) with respect to stratification characteristics, clusters as similar as possible were sought as replacement in a third step.

[Deliberate Sample of Bachelor Graduates] Selection of the Bachelor graduates was conducted through deliberate sampling. This selection procedure was chosen because in the examination year 2005 relatively few higher education graduates took a Bachelor's degree. However it was desired to survey as many people in this new graduate group as possible, in order to gain initial investigative knowledge. On the basis of previous first-year students numbers, average student drop-out rates as well as average duration of studies was estimated in subject areas where larger numbers of Bachelor graduates were to be expected. Based on these estimates, all Bachelor graduates of the identified subject areas were selected.

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 2005 (e.g. examination year, first completed degree, area of study, type of degree; cf. Chapter 3).[40] 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 (mail address, 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.[41] The address list was checked and updated between the waves accordingly in order to take into account those participants whose e-mail addresses were not available after the first wave yet or whose e-mail addresses changed in the meantime.[42]

[Survey Documents] The survey documents for the first two survey waves included a cover letter (with data protection information), the questionnaire, a flyer with key information on the study and a postage paid envelope addressed to the DZHW for returning the completed questionnaire. [43] In addition two reminder letters were sent.

In the main survey of the third survey wave, a cover letter with a link to the online survey, an individual password (token) and a link to the data protection sheet was sent by e-mail. If no valid e-mail address was available, the survey documents were also sent by post. In addition, three reminders were sent by e-mail.

[Fieldwork Phase] The time period of the first survey wave extended from January 1, 2006 to May 18, 2007.[44] 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

[40] 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 2005 (e.g. with delays in certificate production) or if it was not the first completed degree.

[41] 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.

[42] Respondents were contacted both after the first wave and before the second wave and asked to update their contact information. The addresses of undeliverable mail items were updated via the Deutsche Post address update service and the RISER ID Services GmbH registration information service. In the field phase of the second wave, the DZHW also searched for addresses if survey documents were undeliverable. For the third wave, further address updates were carried out after completion of the second wave, before field start and within the survey period.

[43] In the second wave, there was also the possibility, upon request, to receive the questionnaire in electronic form by e-mail. This service was used by 140 people (see Grotheer, Isleib, Netz & Briedis, 2012, p. 424).

[44] The fieldwork time was extended for as long as possible – and in parallel to project objectives – so that every questionnaire received until May 2007 was included.

exact point in time that survey documents were delivered.[45] 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 lasted from December 6, 2010 to February 21, 2012.[46] As the DZHW now possessed the address list of respondents, exact dates for the survey documents' delivery could be specified.[47] In addition, the reminder letters only targeted the persons who had not yet participated in the survey.

The survey period for the main survey of the third survey wave ran from April 14, 2016 to June 7, 2016. As in the second survey wave, concrete dispatch times[48] could be determined in the third wave and specific reminders sent to non-participants.

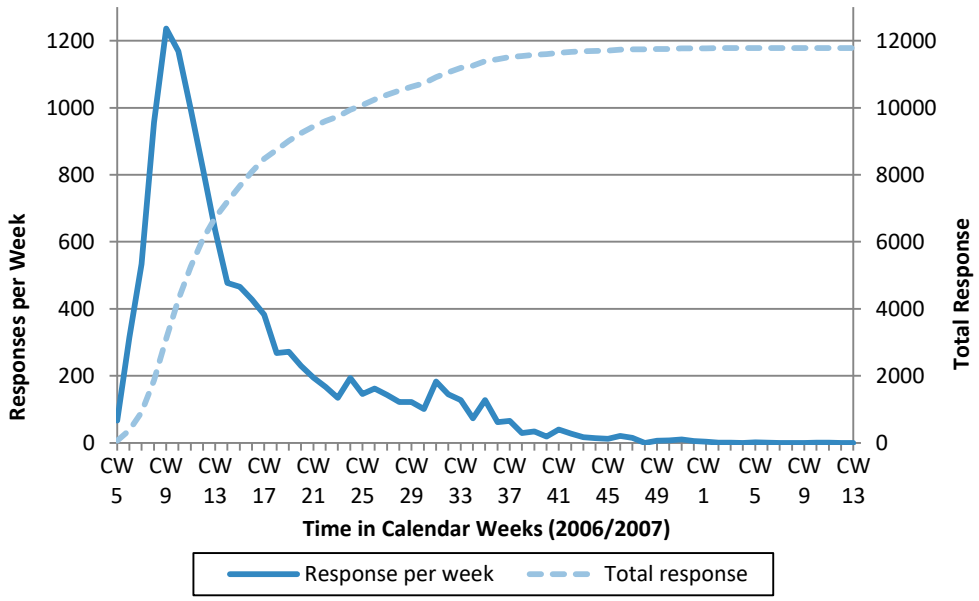
[Measures to Increase Response] In addition to the especially effective instrument of the reminder letter for increasing responses and the flyer sent together with the covering letter of the first and second waves, it was announced in the covering letter that a summary of the key results of the study would be sent together with the survey. Furthermore on a project homepage information on the project and arising publications was made available. Finally a prize draw was made for all survey participants. In the first survey wave 40 book tokens worth €50 each were drawn as prizes, in the second wave a notebook worth around €1,200, five flight gift tokens worth €200 each and ten book tokens worth €30 each. In the third wave, the draw included a Tablet PC at around 1,000 euros, two smartphones at around 550 euros and five train travel vouchers at 100 euros each.

[45] 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.
[46] The field time differs from the SUF version 1.0.0, because latecomers were added to the data set later.
[47] Invitation: 6.-10.12.2010; first reminder: 14.01.2011; second reminder: 11.02.2011.
[48] Main questionnaire: Invitation: 14.04.2016 (e-mail), 21./22.04.2016 (postal); First reminder: 03.05.2016 (e-mail); Second reminder: 19.05.2016 (e-mail); Third reminder: 31.05.2016 (e-mail)

5 Response Rate

[Response Rate] The gross sample from the first survey wave contained ca. 47,800 graduates registered with the examination offices of their respective institutes of higher education. Using the sample procedure described in Chapter 3, these individuals were identified and contacted.[49] In total, 12,114 questionnaires were returned to the DZHW. Of these, 327[50] were excluded, as they did not belong to the target population (neutral sample attrition) or could not be evaluated (relevant sample attrition) (cf. Chapter 6.2). In the final sample of the first survey wave, 11,787[50] cases remained; of which 1,622 were Bachelor graduates. With regard to the gross sample the response rate lies at around 25 percent (cf. Table 2). Figure 4 represents the response rate to the questionnaires during the fieldwork phase of the first survey wave. It can be observed that a large proportion of the completed questionnaires reached the DZHW in the first half of the field phase, during which the reminders were also sent. At the same time, questionnaires were also sent back at later points in time, e.g. after the second reminder had been sent some time ago.

Figure 4: Response Rate of the DZHW Graduate Panel 2005 over Time, 1st Wave



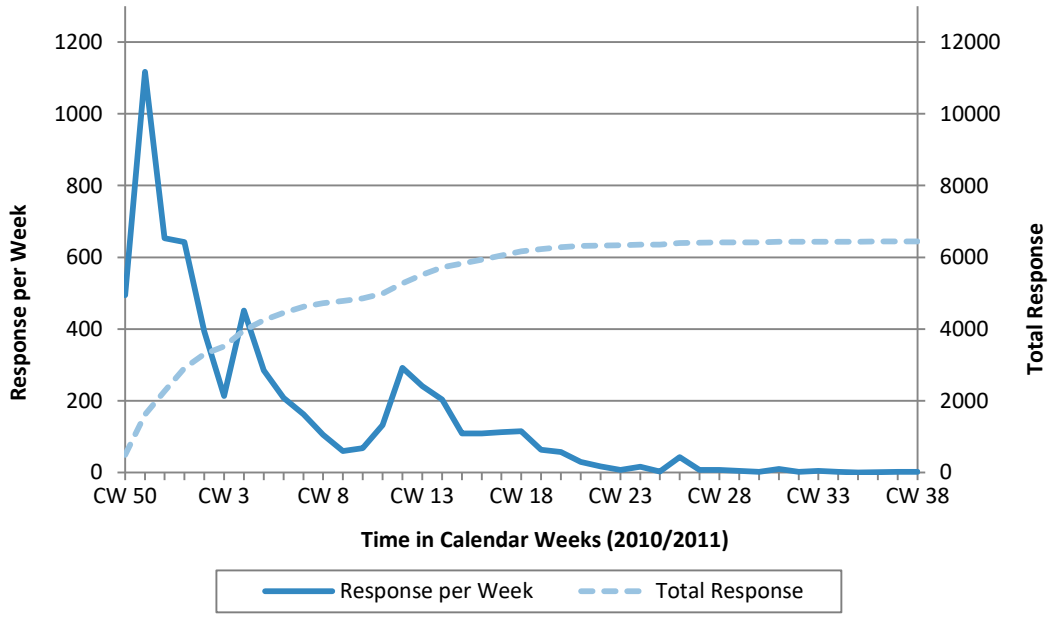
NOTE: Only cases that could be evaluated are included.

10,706 persons, over 90 percent of the 11,787 participants in the first wave, confirmed that they would be willing to be contacted for further surveys. This constitutes the gross sample of the second wave. Due to various sample relevant attrition (e.g. non-participation, invalid addresses[51] or questionnaires which could not be evaluated[52]), the net sample of the second

[49] Since differentiation between an (unadjusted) gross initial sample and an adjusted gross sample is not possible, in the following only the description "gross sample" is used. Likewise it is not possible to reconstruct how many of these graduates were Bachelor graduates.
[50] From the SUF version 1.0.0 a case had to be deleted later because of a duplication in the (original) ID.
[51] This affected around 300 cases.
[52] This affected 4 cases.

wave amounts to 6,472 cases, of which 798 are Bachelor graduates. With reference to the 10,706 cases in the gross sample, the response rate lies at 60.5 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).

Figure 5: Response Rate of the DZHW Graduate Panel 2005 over Time, 2nd Wave



NOTE: Only cases that could be evaluated are included.

Over 70 percent of the questionnaires sent between December 2010 (CW 50) and February 2011 (CW 8) were received (cf. Figure 5). Within this time period, the reminder letters were also sent. In March and April 2011 (CW 9 to CW 17), an additional 20 percent of the questionnaires were received.

Out of the 6,472 cases from the second wave, 6,453 cases were contacted for the third wave.[53] After deducting the sample-relevant attrition (non-participation, non-evaluable questionnaires), 4,279 cases remained in the net sample of the third wave, including 519 Bachelor graduates and 3,760 graduates of traditional courses of study. The response rate is thus 66.3% and higher than in the second wave, but significantly lower than the response rate in the third survey waves of the previous graduate cohorts in 1997 (89%) and 2001 (88%) (Euler et al., 2018, p. 10).

[53] The attrition results on the one hand from refusals to participate in subsequent waves and on the other from invalid addresses.

Table 2: Gross and Net Samples and Response Rates of the DZHW Graduate Panel 2005

	Wave 1	Wave 2	Wave 3
Gross sample	47,800	10,706	6,453
Net sample	11,787[a]	6,472[b]	4,279[c]
Response Rate	24.7 %	60.5 %	66.3 %
Proportion Gross Sample W2/3 of Gross Sample W1	--	22.4 %	13.5 %
Proportion Net Sample W2/3 of Net Sample W1	--	54.9 %	36.3 %
Proportion Net Sample W2/3 of Gross Sample W1	--	13.5 %	9.0 %

[a] of which 1,622 were Bachelor graduates.
[b] of which 798 were Bachelor graduates.
[c] of which 519 were Bachelor graduates.

[Panel Attrition] The Graduate Panel 2005 is further subject to attrition processes[54] 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 or third 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).

Consideration over time shows that the gross sample in the second wave only amounts to around 22 per cent of the gross sample in the first wave. Of the 11,787 cases in the net sample of the first wave, around 55 per cent were surveyed in the second survey wave (cf. Table 2). In comparing the net sample of the second wave with the gross sample of the first wave, only 14 per cent of the initial gross sample participated in both survey waves. This development continues in the third wave.

[54] 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. The procedures described in Chapters 6.1 and 6.2 had already been conducted by the primary research project. The generation of variables (Chapter 6.3) was carried out by the primary project as well as the RDC during data preparation. Procedures described in Chapters 6.4 to 6.6 were carried out by the RDC building on the work of the primary research project. Additional procedures (weighting and anonymization) are explained separately in Chapters 7 and 8.

6.1 Data Transfer

[PAPI surveys] In the PAPI surveys, the information provided by the respondents in the paper questionnaires had first to be transferred into a computer-readable format for further processing on the basis of a code plan. To this end, the questionnaire noted to which question or sub-question a variable is assigned, which name this variable bears and which numerical codes should be used for the standardised answers of the respondents. Numerical codes for the open responses were already noted (cf. Chapter 6.3) and manual preliminary corrections were made to facilitate data transmission (cf. Chapter 6.2). The variables were additionally numbered in order to determine the order in which they were recorded. 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.

[Online Surveys] The data from the online surveys could be exported directly from the survey software as a .csv file and processed further.

6.2 Data Checking and Data Cleansing

[Consistency Checks] Various consistency checks were carried out for all survey waves. 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).
- *Test of Combination of Characteristics:* The consistency of responses within a questionnaire as well as between survey waves was checked. In particular, content and time-related data on the same topics were compared. For example, it was checked within both waves whether the information on occupational activities in the employment tableau corresponded with the corresponding information in the calendar.

Any inconsistencies found were - if possible - resolved by comparison with other entries in the questionnaire (using the paper questionnaires if necessary) or otherwise assigned a corresponding missing code (see Chapter 6.6).

In the first two survey waves, initial consistency checks were carried out manually on the paper questionnaires before the data was transferred.[55] Following the data transfer, a comprehensive check and correction of the data was carried out using DZHW's own software.[56] In the third wave of the survey, the consistency check was also software-supported on the one hand and - for the calendar - via Stata-Do-Files on the other.

[Deletion of Cases] In all 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. Moreover, some cases were identified after the first wave as not belonging to the target population.[57] These were likewise removed from the data set. [58]

6.3 Generation of Variables

In addition to the variables containing the direct answers of the respondents, the data set of the Graduate Panel 2005 also contains generated variables. On the one hand, this includes variables that were numerically coded from the originally (semi-)open responses.[59] For the paper questionnaires, the (semi-)open responses were already coded during data transmission (cf. Chapter 6.1). For the online data, the coding was usually carried out with Excel. The coding decisions made by the primary research project were retained unchanged. On the other hand, variables were changed due to data protection reasons (cf. Chapter 8) (e.g. aggregation of course subjects into areas of study and subject groups or deriving the location and type of the higher education institution from the higher education institution variables). With a few exceptions, variables generated in the course of anonymization measures were created by the RDC. In addition, the primary research project generated auxiliary variables to facilitate work with the employment tableau.[60]

The variable name of a generated variable is identified in the data set by the suffix "_g#" (see Chapter 6.5). Where possible, generated variables were positioned in the data set according to the respective output variable. If a variable was generated from various source

[55] The number of corrections made was not documented centrally but only on the paper questionnaires and can therefore no longer be systematically reconstructed.

[56] For this purpose, the collected questionnaire data were read into a database. Valid value ranges and answer combinations were then defined and checked using formal rules.

[57] 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.

[58] Please note that the data set contains several cases with a graduation date several months after the actual examination period of the examination year 2005. 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 2005 which were assigned to the examination year 2005 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 obviously didn't belong to the population.

[59] Individual open questions were not coded because they were mainly collected as contextual information for coding other open responses or due to insufficient time resources. This concerns, for example, the typical focal points of work that were surveyed in addition to the job title and field of activity (question 5.1 in the first wave; question 4.12 in the second wave; page 14 in the third wave). The information on the typical focal points of work was only used to obtain additional information for coding the occupational title, which was also openly queried, as well as the occupational field of activity.

[60] see https://metadata.fdz.dzhw.eu/#!/en/instruments/ins-gra2005-ins3$ the comparison figure documentation as well as the parental leave flag variable documentation

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.

An overview of all variables generated for the 2005 graduate panel and information on how they were generated can be found in Appendix 1. Detailed documentation of the individual variables, including their respective characteristics and calculation rules, can be found in the metadata search system on the respective variable detail page.[61]

6.4 Generation of the Data Sets

[Merging of the Waves] Data of the individual survey waves were merged. Case assignment was made using the identification numbers of the respondents produced in the fieldwork phase (cf. Chapter 4).

[Generation of Individual and Spell Data Set] The merged data were stored in two separate 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 4.7 of the first wave, Question 1.7 of the second wave, Page 9 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*).

[Separation of the Bachelor Data] Due to different sampling procedures used (cf. Chapter 3), the sample for the Bachelor graduates is not suited to making inferences about the target population of this group. For this reason, individual and spell data sets for Bachelor graduates were compiled and saved separated from those of graduates of traditional courses of study.

[Data Format] All data sets are available in Stata as well as SPSS format.

6.5 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

[61] E.g. https://metadata.fdz.dzhw.eu/#!/en/variables/var-gra2005-ds1-astu011e_g1o$

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.[62]

[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)[63], 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 identifier based on a letter.[64] The root consists of three components. First, the variable is assigned to a thematic area using a three-character English letter abbreviation. Table 3 presents an overview of the various thematic areas of the Graduate Panel 2005 as well as the related abbreviations for the root of the variable name. Secondly, the variables within the defined thematic areas are numbered consecutively to two to four digits. Thirdly, a letter at the end of the root can be used to identify different variables belonging to the same question and thus having the same thematic differentiation and numbering (e.g. item batteries, multiple answers, or questions combining closed and open questions) (e.g. stu01a, stu01b, stu01c, ...). For indicators collected in several survey waves, the names of the associated variables ("panel variables") are harmonised by assigning an identical root.

Table 3: Thematic Areas and Abbreviations for DZHW Graduate Panel 2005 Variable Labels

Thematic Area Abbreviation	Thematic Area (English)	Thematic area (German)
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
wgt	weights	Gewichtungsvariablen
sys	system variables	Systemvariablen

The suffix separated from the root by an underscore contains distinct additional information:

- Generated variables (see Chapter 6.3) were marked with the abbreviation g# (g1 or g2, g3, ... for other derivatives). The type of generated variable includes all variables that were generated from one or more variables of the original data set (e.g. coded variables, indices, aggregations).
- For information that has been repeated in several waves but queried with modified question, item or answer category formulations[65], the new variable versions have the

[62] The command "label language en" switches to the English labels. With "label language de" you can switch back to the German labels.
[63] This contains information on which cases participated in which waves.
[64] Wave 1: a; Wave 2: b; Wave 3 (Main Questionnaire): c
[65] The presentation and arrangement of the respective question in the survey instrument are irrelevant. Likewise, it is not necessary that the corresponding question was repeatedly asked to the same groups of persons.

same variable root as the original variable, but with a version abbreviation v# in the suffix (_v1 for the first change, _v2 for the second change, etc.). The original version of the variable is not flagged.

- For anonymization reasons, certain variables cannot be viewed via all potential access ways (Download-CUF, Download-SUF, Remote-Desktop-SUF, On-Site-SUF; see Chapter III). In these cases, the suffix of the variable name specifies the access way from which the variable can be used[66]:
  - **d:** Variable is not usable in CUF, but it can be used in Download-SUF, Remote-Desktop-SUF and On-Site-SUF.
  - **r:** Variable is not usable in CUF and Download-SUF, but it can be used in Remote-Desktop-SUF and On-Site-SUF.
  - **o:** Variable is not usable in CUF, Download-SUF and Remote-Desktop-SUF, but it can be used in On-Site-SUF.
  - **a:** Variable cannot be used via any access way. However, it is documented, since there are related questions in the questionnaire.
- Weighting variables contain (in addition to the fixed thematic area abbreviation "wgt") the suffix t# as an indicator of the waves to which they refer (see Chapter 7.1).[67]

[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).

6.6 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 4 presents an overview of the system for coding missing values. The coding for missing values used in the Graduate Panel 2005 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 anonymization 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 (i.e. item-specific missing values, including "not given" with items aocc17a, aocc17b and aocc17c, Question 4.16, 1st wave).

[66] "Usable" means: the variable does not contain the missing "anonymized".
[67] They therefore do not contain a wave prefix.

Table 4: 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 values	-969	unknown missing[c]
	-968	**implausible value**[d]
	-967	**anonymised**
	-966	**not determinable**[e]
	-965	invalid multiple answer
-949 to -930: Item-specific missing values[f]	**-949**	**not given**
	-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.

7 Weighting

Weighting the data adjusts for sample bias compared to the defined population. However, weighting was only performed on data for the graduates of traditional courses of study. For Bachelor graduates, there was no weighting due to its explorative character and the correspondingly selected sample design (cf. Chapter 3). A general introduction to the procedure and the presentation of the final weights follows in Section 7.1. The weighting procedure is also described in detail in Section 7.2.

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, unavailability, 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.[68] However, they mostly result from a systematic process (Missing at Random, Not Missing at Random), which requires modelling.[69]

[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 sampling frame. 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-typical method cannot be implemented with the data of the Graduate Panel 2005. The sample design does not allow exact probabilities of inclusion to be derived. Because there is no information on non-participants in the first wave available, individual non-response weights can also not be derived. The cross-sectional weight for the first wave is therefore an estimated design weight, which is calibrated using information from the population. As information on non-participants can be gained from the population the calibration procedure encompasses also a non-response adjustment. For the second and third waves, an additional non-response weight is calculated which models the non-participation in the second and third wave respectively with the help of information from the respective preceding waves. From this, the longitudinal section weights were calculated. The weights produced are represented in Table 5.

[68] Insofar as the loss of statistical test strength through the reduction of the sample is considered irrelevant.
[69] On the different forms of attrition processes see fundamentally Rubin, 1976.

Table 5: Weights provided for the DZHW Graduate Panel 2005

Variable Name	Description
wgt_t1d	Cross-Sectional Weight for Wave 1
wgt_t1t2d	Longitudinal Section Weight for Two-Wave-Panel
wgt_t1t2t3d	Longitudinal Section Weight for Three-Wave-Panel

[Instructions for Use of the Weights] All weights were calculated only for the data set of graduates of traditional study courses. The weights calculated involve probability weights, which can be taken into account in Stata with the help of .ado-specific options.[70] The weight wgt_t1d is intended for evaluations of the first wave, the weights wgt_t1t2d and wgt_t1t2t3d for evaluations of the two-/three-wave-panel.[71] 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[72]:

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

Due to missing information on the non-participants in the first wave, no comprehensive adjustment of the estimated design weights was possible for the attrition process via non-participation (non-response). However, the estimated design weights were calibrated using characteristics of the population. The characteristics included the federal state, gender, type of educational qualification and area of study.[73] 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,[74] resulting in a cross-sectional weight $wgt_{i_{t1}}$ for the first wave of the Graduate Panel 2005. For the second and third survey wave, no cross-sectional weight was produced since no newly added individuals were surveyed (i.e. refreshment sample) and because no person could participate in Wave 2 or 3, respectively, if they had not participated in the first survey wave.

[70] See Stata help (Command: *help weights*).
[71] For evaluations that only refer to variables of the second wave, the longitudinal sectional weight is also to be used. The same applies to evaluations in the third wave.
[72] Where n_{sc} corresponds to the number of clusters in a stratum, N_{sc} 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.
[73] The information on the population was derived from data of the Federal Statistical Office (Examination Statistics 2005).
[74] Raking is also termed 'iterative proportional fitting' (ipf) (cf. Kolenikov 2014).

[Longitudinal Sectional Weighting] For weighting the two- and three-wave-panel, attrition processes had to be considered in time sequence (cf. Chapter 5). For this purpose, an attrition weight was calculated, 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 waves was available for the non-participants of the second and third wave. This information served as covariates (σ_t) in a probit regression model, which estimated probability of participation at a given point in time $P(Res_{t+1})$. 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 series of predictors demonstrated their significance for the prediction of probability of participation in the second or third wave. Conditional probability of participation could be derived from the model, the reciprocal value of which represents the nonresponse weight for the second or third wave:[75]

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

The total weight (nonresponse-adjusted design weight) for the respective subpanel arises from the product of the estimated design weight and the respective non-response weights:

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

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

Hence, the non-response adjusted longitudinal sectional weight was calibrated to characteristics of the population using the Raking algorithm.[76]

[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.

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

[76] It involved the same attributes that were used for the calibration of the estimated design weight in the first survey wave.

8 Anonymisation

[Data Protection Legal Framework] For personal data[77], which is collected in voluntary surveys carried out by the DZHW, the EU General Data Protection Regulation (EU-GDPR) and the German Federal Data Protection Act (BDSG) in its most recent version dated 30 June 2017 apply.[78] According to this provision, personal data must generally be prepared for disclosure for scientific secondary use (without a declaration of consent for secondary use of the personal data) in such a way that "the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (Art. 4 (5) GDPR; see also Art. 89 GDPR and recital 26 GDPR). This means that for the transfer of data from scientific research projects to third parties, the data must be made anonymous in such a way that no reference to the individual persons can be made.

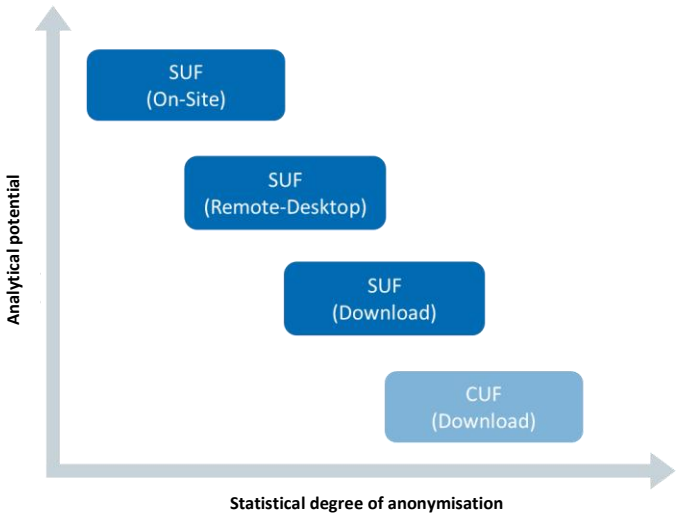
[Data Access, Level of Anonymization and Analytical Potential] For the Graduate Panel 2005, the RDC for Higher Education Research and Science Studies provides a SUF for scientific secondary use and a CUF for teaching and exercise purposes. The anonymity of the respondents is thus protected by a combination of statistical measures and technical access barriers. The more strongly data access is technically controlled, the lower is the risk of de-anonymization 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 for Higher Education Research and Science Studies 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 6 gives an overview of the respective level of statistical anonymization and the related analytical potential. In the following the statistical anonymization measures performed are explained according to data product (SUF/CUF) and mode of access.

[77] "'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Art. 4 GDPR, p. 1).

[78] The GDPR applies in principle within the EU and thus also to the DZHW. The BDSG in its new version of 30 June 2017 (Act for the Adaptation of the Data Protection Act to the Regulation (EU) 2016/679 and for the implementation of the Directive (EU) 2016/680 (Data Protection Adaptation and Implementation Act EU DSAnpUG-EU)) also partially applies in addition, as DZHW GmbH is legally regarded as a public body of the Federation (§ 2 Para. 3 BDSG). The Federal Government holds the absolute majority of the shares of DZHW GmbH and the institute fulfils tasks of the Federal Government's public administration in the broadest sense.

Figure 6: Modes of Access, Statistical Level of Anonymization and Analytical Potential of the Data of the DZHW Graduate Panel 2005



[Statistical Anonymization Measures] In the course of anonymization, 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 2005 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.[79] For the Graduate Panel 2005, the following quasi-identifiers were used: 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 6). 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".

[79] 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.

In addition, Ebel and Meyermann recommend deleting open responses even if the respective questions are unproblematic in themselves. For there is a danger that study participants have disclosed critical information that could lead to identification when answering questions that are actually unobjectionable and openly answered (cf. Ebel & Meyermann, 2015, p. 5). Most of the open responses have already been coded by the primary research project as part of the data preparation process and will be made available in this form (partly aggregated). Non-coded open responses were deleted in the CUF and in all SUF variants.

Finally, it was checked whether the data contained sensitive information, e.g. on health, sexual orientation and political attitudes. Although these are not necessarily suitable for the re-identification of individuals or institutions, the information can be useful in the case of de-anonymization (cf. Koberg, 2016, p. 694) and are therefore particularly worthy of protection (Art. 9 GDPR, recital 51 GDPR). In the Graduate Panel 2005, health information was collected for which no additional consent for secondary use was obtained from the respondents. These answers were therefore deleted from the CUF and all SUF variants.

For the realization of the anonymity of the respondents in CUF data, more restrictive statistical anonymization measures on the variable level in comparison to the SUF variants were performed. In addition, a randomly selected sub-sample of the data was drawn (2/3 of the surveyed graduates of traditional courses of study).

Table 6 below summarizes the statistical anonymization measures carried out depending on the data form or access path.

Table 6: Statistical Anonymization Measures for the Data of the DZHW Graduate Panel 2005 by Mode of Access[80]

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 date	Available	Release of month/year	Release of month/year	Deletion
subject[a]	Available	Aggregation to areas of study[a]	Aggregation to areas of study[a]	Aggregation to subject areas[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
Further academic qualification (country)	Available	Available	Aggregation to Germany and abroad	Aggregation to Germany and abroad
Place of work (federal state/country)	Available	Available	Aggregation to federal states and abroad	Aggregation to both old and new federal states and abroad
Place of work (three-digit 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 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 (three-digit 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

[80] 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).

Characteristic	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Sub-sample)
Place of permanent residence (abroad)	Available	Available	Aggregation to missing/abroad	Aggregation to missing/abroad
Place of permanent residence (three-digit 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]	Aggregation to missing/Germany
Occupation (KldB occupational classes (4-digit) or occupational categories (5-digit))[c]	Available	Aggregation to occupational orders/ groups (3-digit)[c]	Aggregation to occupational orders/ groups (3-digit)[c]	Aggregation to occupational groups/ Main occupational groups (2-digit)[c]
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[e]
Note on state of health	Deletion	Deletion	Deletion	Deletion
Other open responses	Coding/Deletion	Coding/Deletion	Deletion	Deletion

[a] According to the Key List of Student and Examination Statistics Winter Semester 2004/2005 and Summer Semester 2005 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 (KldB) from 1992 (wave 1) and 2010 (waves 2 and 3)
[d] 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.
[e] Age responses above a certain limit were aggregated to one category.

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. (2007). Übergänge und Erfahrungen nach dem Hochschulabschluss. Ergebnisse der HIS-Absolventenbefragung 2005 (HIS: Forum Hochschule 13/2007). Hannover: HIS.

Briedis, K. & Minks, K.-H. (2007). Generation Praktikum. Mythos oder Massenphänomen. Hannover: HIS.

Cox, L. H. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association,* 82(398), 520–524. Zugriff am 17.03.2016 UTC.

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.

Euler, T., Trennt, F., Trommer, M. & Schaeper, H. (2018). Werdegänge der Hochschulabsolventinnen und Hochschulabsolventen 2005. Dritte Befragung des Prüfungsjahrgangs 2005 zehn Jahre nach dem Abschluss. (Forum Hochschule 01/2018). Hannover: DZHW.

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.

Grotheer, M., Isleib, S., Netz, N. & Briedis, K. (2012). *Hochqualifiziert und gefragt. Ergebnisse der zweiten HIS-HF Absolventenbefragung des Jahrgangs 2005* (HIS: Forum Hochschule 14/2012). 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). *Datenschutz am Forschungsdatenzentrum* (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Jaksztat, S. (2014). Bildungsherkunft und Promotionen: Wie beeinflusst das elterliche Bildungsniveau den Übergang in die Promotionsphase? *Zeitschrift für Soziologie,* 43(4), 286–301.

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.

Kratz, F. & Netz, N. (2016). Which mechanisms explain monetary returns to international student mobility? *Studies in Higher Education*. doi:10.1080/03075079.2016.1172307

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.

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.

Schaeper, H. & Briedis, K. (2004). Kompetenzen von Hochschulabsolventinnen und Hochschulabsolventen, berufliche Anforderungen und Folgerungen für die Hochschulreform. Hannover: HIS.

Schaeper, H. (2009). Development of competencies and teaching–learning arrangements in higher education: findings from Germany. *Studies in Higher Education,* 34(6), 677–697. doi:10.1080/03075070802669207

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

Appendix 1: Variables generated in the DZHW Graduate Panel 2005

Variables	Information on generation[a]
astu011e_g1o astu011g_g1o astu011i_g1o astu012e_g1o astu012g_g1o astu012i_g1o astu013e_g1o astu013g_g1o astu013i_g1o astu014e_g1o astu014g_g1o astu014i_g1o astu015e_g1o astu015g_g1o astu015i_g1o astu021c_g1o astu021d_g1o astu021e_g1o astu022c_g1o astu022d_g1o astu022e_g1o astu023c_g1o astu023d_g1o astu023e_g1o afec021h_g1o afec022h_g1o bfec161g_g1o bfec162g_g1o bfec163g_g1o	Coding of open responses on the subject of study Coding list: cl-destatis-studienfach-2005[b] (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer semester 2005), key 3.1)
astu011e_g2d astu011g_g2d astu011i_g2d astu012e_g2d astu012g_g2d astu012i_g2d astu013e_g2d astu013g_g2d astu013i_g2d astu014e_g2d astu014g_g2d astu014i_g2d astu015e_g2d astu015g_g2d astu015i_g2d astu021c_g2d astu021d_g2d astu021e_g2d astu022c_g2d astu022d_g2d astu022e_g2d astu023c_g2d astu023d_g2d astu023e_g2d	Aggregation of subjects into fields of study Coding list: cl-destatis-studienfach-2005[b] (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer semester 2005), key 3.1)

Variables	Information on generation[a]
afec021h_g2d afec022h_g2d bfec161g_g2d bfec162g_g2d bfec163g_g2d	
astu011e_g3 astu011g_g3 astu011i_g3 astu012e_g3 astu012g_g3 astu012i_g3 astu013e_g3 astu013g_g3 astu013i_g3 astu014e_g3 astu014g_g3 astu014i_g3 astu015e_g3 astu015g_g3 astu015i_g3 astu021c_g3 astu021d_g3 astu021e_g3 astu022c_g3 astu022d_g3 astu022e_g3 astu023c_g3 astu023d_g3 astu023e_g3 afec021h_g3 afec022h_g3 bfec161g_g3 bfec162g_g3 bfec163g_g3	Aggregation of subjects into subject groups Coding list: cl-destatis-studienfach-2005 (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer semester 2005), key 3.1)
astu011f_g1 astu011h_g1 astu011j_g1 astu012f_g1 astu012h_g1 astu012j_g1 astu013f_g1 astu013h_g1 astu013j_g1 astu014f_g1 astu014h_g1 astu014j_g1 astu015f_g1 astu015h_g1 astu015j_g1 astu021f_g1 astu022f_g1 astu023f_g1	Coding of open responses on type of degree Coding list: cl-dzhw-2 (project-specific coding list)
astu09a_g1r astu09b_g1r	Coding of open responses on the major field of study (for selected economics subjects) Coding list: cl-dzhw-3 (project-specific coding list)
astu011k_g1a astu012k_g1a astu013k_g1a	Coding of open responses on the higher education institution Coding list: cl-destatis-hochschule-2005[b] (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer

Variables	Information on generation[a]
astu014k_g1a astu015k_g1a astu021g_g1a astu022g_g1a astu023g_g1a bfec161h_g1a bfec162h_g1a bfec163h_g1a	semester 2004/2005) 2005), key 2.2)
astu011k_g2 astu012k_g2 astu013k_g2 astu014k_g2 astu015k_g2 astu021g_g2 astu022g_g2 astu023g_g2 bfec161h_g2 bfec162h_g2 bfec163h_g2	Aggregation of the higher education institution to the type of higher education institution (Uni/FH) Coding list: cl-destatis-hochschule-2005 (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer semester 2005), key 2.2) In the key index, the higher education institutions in Germany are differentiated into 6 types (universities, colleges of education, theological/church colleges, art colleges, universities of applied sciences (without administrative colleges), administrative colleges). These types were again aggregated to the two categories "university" and "university of applied sciences". The generated variable has the value 1 "University", if it is a university, college of education, theological/church college or college of art. It has the value 2 "university of applied sciences" if it is a university of applied sciences or "administrative college".
astu011k_g3o astu012k_g3o astu013k_g3o astu014k_g3o astu015k_g3o astu021g_g3o astu022g_g3o astu023g_g3o bfec161h_g3o bfec162h_g3o bfec163h_g3o	Aggregation of higher education institutions to location of higher education (NUTS 2 regions) Coding list: cl-eurostat-nuts-2010 (NUTS Classification of the Statistical Office of the European Union (Eurostat) of 2010)
astu011k_g4r astu012k_g4r astu013k_g4r astu014k_g4r astu015k_g4r astu021g_g4r astu022g_g4r astu023g_g4r bfec161h_g4r bfec162h_g4r bfec163h_g4r	Aggregation of higher education institutions to location of higher education (federal states) Coding list: cl-destatis-hochschule-2005[b] (Destatis key index for student and examination statistics (winter semester 2004/2005 and summer semester 2005), key 2.2)
astu011k_g5 astu012k_g5 astu013k_g5 astu014k_g5 astu015k_g5 astu021g_g5 astu022g_g5 astu023g_g5 bfec161h_g5 bfec162h_g5 bfec163h_g5	Aggregation of higher education institutions (federal states) to location of higher education (new/old federal states)
aocc221j_g1r aocc222j_g1r aocc223j_g1r	Coding of open responses on the location of work (federal state/country) Coding list: cl-destatis-bundesland-1990[b] (Destatis federal state key (corresponding to the first two digits of the official municipality key

Variables	Information on generation[a]
aocc224j_g1r aocc225j_g1r aocc226j_g1r bocc221j_g1v1r bocc222j_g1v1r bocc223j_g1v1r bocc224j_g1v1r bocc225j_g1v1r bocc226j_g1v1r bocc227j_g1r bocc228j_g1r bocc229j_g1r cocc221j_g1v1r cocc222j_g1v1r cocc223j_g1v1r cocc224j_g1v1r cocc225j_g1v1r cocc226j_g1v1r cocc227j_g1r cocc228j_g1r cocc229j_g1r cocc2210j_g1r cocc2211j_g1r cocc2212j_g1r	(AGS)); cl-dzhw-1 (project-specific coding list, waves 1 and 2); cl-destatis-ausland-2016[b] (Destatis State and Territorial Nomenclature 2016, wave 3)
aocc221j_g2d aocc222j_g2d aocc223j_g2d aocc224j_g2d aocc225j_g2d aocc226j_g2d bocc221j_g2v1d bocc222j_g2v1d bocc223j_g2v1d bocc224j_g2v1d bocc225j_g2v1d bocc226j_g2v1d bocc227j_g2d bocc228j_g2d bocc229j_g2d cocc221j_g2v1d cocc222j_g2v1d cocc223j_g2v1d cocc224j_g2v1d cocc225j_g2v1d cocc226j_g2v1d cocc227j_g2d cocc228j_g2d cocc229j_g2d cocc2210j_g2d cocc2211j_g2d cocc2212j_g2d	Aggregation of the countries of the location of work (for abroad) to the category "abroad" (federal states were retained)
aocc221j_g3 aocc222j_g3 aocc223j_g3 aocc224j_g3 aocc225j_g3 aocc226j_g3 bocc221j_g3v1	Aggregation of the location of work (federal states) to new/old federal states or countries of the location of work (for abroad) to the category "abroad".

Variables	Information on generation[a]
bocc222j_g3v1 bocc223j_g3v1 bocc224j_g3v1 bocc225j_g3v1 bocc226j_g3v1 bocc227j_g3 bocc228j_g3 bocc229j_g3 cocc221j_g3v1 cocc222j_g3v1 cocc223j_g3v1 cocc224j_g3v1 cocc225j_g3v1 cocc226j_g3v1 cocc227j_g3 cocc228j_g3 cocc229j_g3 cocc2210j_g3 cocc2211j_g3 cocc2212j_g3	
cocc221k_g2v2o cocc222k_g2v2o cocc223k_g2v2o cocc224k_g2v2o cocc225k_g2v2o cocc226k_g2v2o cocc227k_g2v1o cocc228k_g2v1o cocc229k_g2v1o cocc2210k_g2o cocc2211k_g2o cocc2212k_g2o	Aggregation of the location of work (5-digit postal code) to 3-digit postal code
aocc221k_g1d aocc222k_g1d aocc223k_g1d aocc224k_g1d aocc225k_g1d aocc226k_g1d bocc221k_g1v1d bocc222k_g1v1d bocc223k_g1v1d bocc224k_g1v1d bocc225k_g1v1d bocc226k_g1v1d bocc227k_g1d bocc228k_g1d bocc229k_g1d cocc221k_g1v2d cocc222k_g1v2d cocc223k_g1v2d cocc224k_g1v2d cocc225k_g1v2d cocc226k_g1v2d cocc227k_g1v1d cocc228k_g1v1d cocc229k_g1v1d cocc2210k_g1d cocc2211k_g1d	Aggregation of the location of work (postal code) into NUTS 2 regions Coding list: cl-eurostat-nuts-2010 (NUTS Classification of the Statistical Office of the European Union (Eurostat) of 2010, waves 1 and 2); cl-dzhw-5 (project –specific coding list, waves 1 and 2); cl-eurostat-deplznuts-2010 (Tercet NUTS-postal codes matching tables of the Statistical Office of the European Union (Eurostat) of 2010, wave 3)

Variables	Information on generation[a]
cocc2212k_g1d	
adem05a_g1r	Coding of open responses on the location of higher education entrance qualification (federal state/state) Coding list: cl-destatis-bundesland-1990[b] (Destatis federal state key (corresponding to the first two digits of the official municipality key (AGS)); cl-dzhw-1 (project-specific coding list)
adem05a_g2d	Aggregation of the countries of the location of the higher education entrance qualification (for abroad) to the category "abroad" (federal states were retained)
adem05a_g3	Aggregation of the federal states of location of the higher education entrance qualification to new/old federal states or countries (for abroad) to the category "abroad"
adem05b_g1d	Aggregation of the location of the higher education entrance qualification (postcode) to NUTS 2 regions Coding list: cl-eurostat-nuts-2010 (NUTS Classification of the Statistical Office of the European Union (Eurostat) of 2010); cl-dzhw-5 (project-specific coding list)
astu061c_g1 astu062c_g1 astu063c_g1 astu064c_g1	Coding of open responses about the stay abroad Coding list: cl-dzhw-1 (project-specific coding list)
afec021k_g1r afec022k_g1r	Coding of open responses on the country of further academic qualification Coding list: cl-dzhw-1 (project-specific coding list)
afec021k_g2 afec022k_g2	Aggregation of the country of further academic qualification into the categories "Germany" and "Abroad"
bocc46b_g1	Coding of open responses on the country of corporate headquarters Coding list: cl-dzhw-1 (project-specific coding list)
bdem18b_g1o bdem19b_g1o	Coding of open responses on the country of birth or nationality Coding list: cl-dzhw-1 (project-specific coding list)
bdem18b_g2r bdem19b_g2r	Aggregation of country of birth or nationality into NEPS regions (adapted) Coding list: cl-dzhw-1 (NEPS regions)
bdem18b_g3d bdem19b_g3d	Aggregation of country of birth or nationality into world regions Coding list: cl-dzhw-1 (project-specific coding list)
aocc19_g1o	Coding of open responses on occupation (KldB-1992-4 digit) Coding list: cl-destatis-kldb-1992 (Destatis Classification of Occupations 1992)
aocc19_g2d	Aggregation of the occupation (KldB-1992-4-digit) to KldB-1992-3-digit Coding list: cl-destatis-kldb-1992 (Destatis Classification of Occupations 1992)
aocc19_g3	Aggregation of the occupation (KldB-1992-4-digit) to KldB-1992-2-digit Coding list: cl-destatis-kldb-1992 (Destatis Classification of Occupations 1992)
aocc35b_g1d	Coding of open responses on occupation (KldB-1992-3 digit) Coding list: cl-destatis-kldb-1992 (Destatis Classification of Occupations 1992)
aocc35b_g2	Aggregation of the occupation (KldB-1992-3-digit) to KldB-1992-2-digit Coding list: cl-destatis-kldb-1992 (Destatis Classification of Occupations 1992)
bocc19_g1v1o cocc19_g1v1o	Coding of open responses on the occupation (KldB-2010-5 digit) Coding list: cl-destatis-kldb-2010Vorversion (previous version of Destatis Classification of Occupations 2010, wave 2); cl-destatis-kldb-2010[b] (Destatis Classification of Occupations 2010, wave 3)
bocc19_g2v1d cocc19_g2v1d	Aggregation of the occupation (KldB-2010-5-digit) to KldB-2010-3-digit Coding list: cl-destatis-kldb-2010[b] (Destatis Classification of Occupations 2010)
bocc19_g3v1	Aggregation of the occupation (KldB-2010-5-digit) to KldB-2010-2-digit

Variables	Information on generation[a]
cocc19_g3v1	Coding list: cl-destatis-kldb-2010[b] (Destatis Classification of Occupations 2010)
aocc20a_g1r aocc20b_g1r bocc20a_g1v1r bocc20b_g1v1r	Coding of open responses on the professional area of responsibility Coding list: cl-dzhw-4 (project-specific coding list)
bdem231c_g1d bdem232c_g1d bdem233c_g1d bdem234c_g1d bdem235c_g1d	Aggregation of residence (postcode) to NUTS 2 regions Coding list: cl-eurostat-nuts-2010 (NUTS Classification of the Statistical Office of the European Union (Eurostat) of 2010); cl-dzhw-5 (project-specific coding list)
bdem231c_g2 bdem232c_g2 bdem233c_g2 bdem234c_g2 bdem235c_g2	Aggregation of residence (postcode) to category 'Germany'.
bdem231d_g1r bdem232d_g1r bdem233d_g1r bdem234d_g1r bdem235d_g1r	Coding of open responses on residence abroad Coding list: cl-dzhw-1 (project-specific coding list)
bdem231d_g2 bdem232d_g2 bdem233d_g2 bdem234d_g2 bdem235d_g2	Aggregation of the country of residence abroad into the category 'abroad'
bocc221m_g1 bocc222m_g1 bocc223m_g1 bocc224m_g1 bocc225m_g1 bocc226m_g1 bocc227m_g1 bocc228m_g1 bocc229m_g1 cocc221m_g1 cocc222m_g1 cocc223m_g1 cocc224m_g1 cocc225m_g1 cocc226m_g1 cocc227m_g1 cocc228m_g1 cocc229m_g1 cocc2210m_g1 cocc2211m_g1 cocc2212m_g1	cf. documentation of comparison figure https://metadata.fdz.dzhw.eu/#!/en/instruments/ins-gra2005-ins3$
cocc221n_g1 cocc222n_g1 cocc223n_g1 cocc224n_g1 cocc225n_g1 cocc226n_g1 cocc227n_g1 cocc228n_g1 cocc229n_g1 cocc2210n_g1 cocc2211n_g1	cf. documentation of the flag variable for parental leave https://metadata.fdz.dzhw.eu/#!/en/instruments/ins-gra2005-ins3$

Variables	Information on generation[a]
cocc2212n_g1	
adem11_g1 bdem11_g1	This variable was usually generated on the basis of the given birth dates of the children. In special cases, however, the number of children does not always correspond to the sum of the specified children (birth dates).
aocc14a_g1	Coding of open responses on sectors of the internship
aocc14b_g1	Coding list: analogous to the answer categories for question 5.7 in wave 1
adem07_g1	Aggregation of the respondent's age (top coding for birth years up to 1959)
astu064d_g1r	Coding of other open responses
afec03o_g1r	Assignment to predefined categories or project-specific coding list
aocc05l_g1r	
aocc06s_g1r	
aocc15h_g1r	
aocc231v_g1r	
aocc232v_g1r	
aocc261b_g1r	
aocc262b_g1r	
aocc271b_g1r	
aocc272b_g1r	
adem01b_g1r	
adem02b_g1r	
afvt01l_g1r	
bocc37p_g1r	
bocc06s_g1v1r	
bfec03o_g1v1r	
bfec11h_g1r	
bocc232v_g1v1r	
bocc272b_g1v1r	
bocc262b_g1v1r	
bocc292j_g1r	
bdem16o_g1r	
bfec161j_g1r	
bfec162j_g1r	
bfec163j_g1r	
cocc272b_g1v1r	

[a] A Coding List ID was only assigned if the categories were not derived from the actual responses in the data set, but from a system (classification). All coding lists can be found in the metadata search system under the survey instruments.

[b] if necessary, supplemented by special codes, if not assignable