

September 2026

Gesche Brandt | Kolja Briedis | Susanne de Vogel | Steffen Jaksztat | Anne-Marie Lapstich | Tina Oestreich | Henrike Steins | Carola Teichmann | Sandra Vietgen | Marten Wallis

DZHW PhD Panel 2014

Data and Methods Report on the Surveys of the 2014
PhD Cohort (Survey Waves 1–10);
Version 6.0.0

This work is licensed under the Creative Commons Attribution – Non-commercial – ShareAlike 3.0 Germany Licence (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Authors

Gesche Brandt
Kolja Briedis
Susanne de Vogel
Steffen Jaksztat
Anne-Marie Lapstich
Tina Oestreich
Henrike Steins
Carola Teichmann
Sandra Vietgen
Marten Wallis

Legal notice

Published by

German Centre for Higher Education Institutions
Science Studies GmbH (DZHW)
Lange Laube 12 | 30159 Hanover | www.dzhw.eu
PO Box 2920 | 30029 Hanover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Management

Dr Marcus Beiner
Regina Oelfke

Chairman of the Supervisory Board

Senior Ministerial Director Peter Greisler

Register court

Hanover Local Court | B 210251
VAT registration number:
DE291239300

September 2026



Contents Table of contents heading

List of Tables and Figures	II
1 Content and structure of the study	3
2 Survey instruments	4
2.1 Pretest.....	4
2.2 Key instruments	4
2.2.1 Employment	4
2.2.2 Professional and academic further qualificatio.....	5
2.2.3 Academic activities.....	5
2.2.4 International mobility.....	5
2.2.5 Personal circumstances	5
2.3 Learning environment and contextual conditions	5
2.4 Wave-specific (focus) topics	6
2.4.1 International mobility.....	6
2.4.2 Individual innovation behaviour.....	6
2.4.3 Personality traits	6
2.4.4 Health and disability.....	6
2.4.5 Political participation.....	7
2.4.6 Questions on COVID-19.....	7
3 Population and sampling method	8
3.1 Population.....	8
3.2 Comprehensive survey.....	8
4 Survey administration	9
4.1 Initial survey.....	9
4.2 Follow-up surveys	9
5 Response rate	11
6 Data preparation	14
6.1 Data transfer	14
6.2 Coding of open responses.....	14
6.3 Data checking and data cleaning	16
6.4 Generation of variables.....	17

6.5	Creation of the datasets	17
6.6	Assignment of variable names, variable and value labels	18
6.7	Coding of missing values.....	20
7	Weighting	21
7.1	Procedure and application notes.....	21
7.2	Modelling of non-response weights	22
7.3	Calibration.....	23
7.4	Trimming.....	23
7.5	Distribution of the weights	23
8	Anonymisation	25
9	Appendix	27
9.1	Logit regression to calculate non-response wave.....	27
10	Bibliography	38

List of Tables and Figures

Table 1:	Gross and net response rates of the DZHW PhD Panel 2014.....	11
Table 2 :	Participation of higher education institutions	12
Figure 1:	Response rates for the DZHW PhD Panel 2014 over time: Wave 2 to Wave 7	12
Table 3 :	Coded characteristics and code lists used	15
Table 4:	Classification of subject areas and variable sets	18
Table 5:	Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) classification system for missing values.....	20
Table 6:	Weights provided for the 2014 DZHW PhD Panel	21
Table 7:	Distribution of weighting for waves 1 to 7.....	24
Table 8:	Modes of access at the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW).....	25
Table 9:	Logit regression to calculate the non-response weight for Wave 3.....	27
Table 10:	Logit regression to calculate the non-response weight for Wave 4.....	29
Table 11:	Logit regression to calculate the non-response weight for Wave 5.....	31
Table 12:	Logit regression to calculate the non-response weight for Wave 6.....	33

1 Content and structure of the study

The DZHW PhD Panel was launched in 2013 as part of the ‘Research on Early-Career Academics’ (FoWiN) funding programme of the Federal Ministry of Education and Research¹. The study examines the influence that formal doctoral contexts and the specific learning and development conditions encountered by doctoral candidates during their doctoral phase have on their career transition after completing their doctorate and on their subsequent career paths both within and outside academia. The survey population comprises all 28,147 individuals (Federal Statistical Office 2015) who completed a doctorate in Germany in the 2014 examination year at a higher education institution with the right to award doctorates. The survey is designed as a comprehensive survey in order to obtain a sufficiently large sample size (for details on the first two waves, see: Brandt et al. 2017).

Parts of the data preparation, as well as the translation and evaluation of the translation process, were carried out as part of the project ‘Optimisation of the translation process in the context of the internationalisation of the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) using the example of additional data publications (OptiDat)’, funded by the Federal Ministry of Research, Technology and Space.

The initial survey took place in 2015; for this purpose, a paper-and-pencil questionnaire was sent by post approximately half a year to one and a half years after completion of the doctorate. The second survey of the graduating cohort took place approximately one year later as an online survey. The subsequent survey waves were conducted at annual intervals, each in the spring (waves 2–10).

In addition to basic data for educational reporting, the dataset contains detailed information on the learning and developmental conditions during the doctoral phase, as well as on the career paths of doctoral graduates following the completion of their doctorates. Furthermore, the dataset includes a range of personality traits (Big Five, self-efficacy, internal/external locus of control) as well as socio-biographical and educational background information. This offers significant, previously unavailable analytical potential for higher education institutions and science studies. The panel design and the survey of longitudinal data at monthly intervals enable causal analyses at the individual level over time (e.g. in the form of event data and sequence analyses).

¹ Funding reference numbers 16FWN014 and 16FWN017

2 Survey instruments

In the first survey wave of the 2014 PhD Graduates Panel, a standardised paper questionnaire in German was used as the survey instrument.² For survey waves two to ten, standardised online questionnaires in German were used for data collection.³ The core questions, which are repeated in every wave, cover topics such as employment, professional and academic further qualifications, academic activities, mobility experiences and information on personal circumstances. Each survey wave also has its own specific focus. The questionnaire was programmed and the survey conducted using the DZHW online survey software 'Zofar'.

For the development of the survey instruments, individual questions were taken from or adapted from existing survey instruments. An overview of the sources used can be found in Appendix **Fehler! Verweisquelle konnte nicht gefunden werden.**

2.1 Pretest

Prior to the initial survey, a cognitive pre-test was first carried out, followed by a quantitative pre-test. The primary aim of these phases was to test the newly developed items designed to survey the learning environment during the doctoral phase and to assess their suitability for use in surveying doctoral candidates and PhD graduates across various disciplines and types of doctoral programmes. The aim of the pre-tests was to ascertain doctoral candidates' and PhD graduates' understanding of the questions and their response behaviour, to determine the time taken to complete the questionnaire, and to identify any order effects. This is documented in detail in a workshop report (Brandt, Vogel & Jaksztat, 2016).

The questionnaire for the second survey was also tested in advance as part of a cognitive pre-test with PhD graduates from different formal PhD programmes and faculties.

2.2 Key instruments

The core instruments are described in more detail below regarding their various themes.

2.2.1 Employment

From the second survey wave onwards, the questionnaires included two tools for recording career trajectories following the completion of the doctorate or since the last survey: Firstly, a career timeline was used to record key career episodes on a month-by-month basis and to detail the characteristics of each episode (start and end dates, occupational position, type of contract, working hours, place of work, academic activity and employer). In addition, an event history calendar was used to record professional and non-professional activities following the doctorate on a month-by-month basis (e.g. further vocational education and training, scholarships, parental leave, family responsibilities, unemployment). To facilitate the updating of data across the waves, all activities since the last survey in which the doctorate graduates took part were recorded. Respondents who, for example, did not take part in the second survey were asked in the third wave to provide details since achieving their doctorate qualification, in order to obtain the most complete longitudinal data possible. Technically, this was

² Respondents could request a German or English PDF questionnaire by email if they wished.

³ The questionnaires, question flow diagrams and variable plans can be downloaded from the metadata system.

achieved using preloads and the display of alternative question texts. Finally, respondents were asked to identify the period corresponding to their current or, respectively, the most recent employment.

For the current or, where applicable, most recent job, a range of additional employment characteristics were also surveyed in each wave. These included job title, field of activity and job responsibilities, on the basis of which coding was carried out during data preparation in accordance with the 2010 Classification of Occupations (KldB2010) for waves 1 to 5⁴. Furthermore, respondents were asked whether their job was in the public service, and data were collected on the company size, the sector, job satisfaction, the qualification (mis)match for the role, and income. From the third wave onwards, the staff category was also surveyed for the first time for roles in academia. Finally, questions on self-employment were asked in each wave.

2.2.2 Professional and academic further qualification

In the second and third waves, information was collected on continuing education organised by professional bodies. As such continuing education programmes usually span several years, details on professional continuing education were subsequently collected in the fifth survey. At the same time, from the third survey onwards, the section was expanded to include two questions on the completion of the continuing education, which are asked annually.

The core set of questions for a survey of any academic further qualification undertaken includes questions on the nature and progress of the further qualification, on academic career intentions, on the assessment of career prospects, and more detailed information on the intention or nature of a professorship.

2.2.3 Academic activities

From the second survey onwards, information on the academic nature of the work was requested. In order to keep the survey burden on respondents as low as possible, detailed questions on academic activities were asked from the third survey onwards, following an introductory filter question. Questions asked annually covered publications, conference attendance, peer review activities, securing external funding, patents, teaching activities and the respondent's subjective sense of integration into the scientific community. In the fifth wave, two additional instruments were used to assess the mentoring situation during the postdoc phase. Participation in academic self-governance and membership of associations and professional societies were no longer surveyed from the second wave onwards.

2.2.4 International mobility

Since the third wave, questions on international mobility experiences and plans have been part of the annual questionnaire. Respondents are asked about stays abroad completed since the last survey and their plans to go abroad in the future.

2.2.5 Personal circumstances

In this section, a survey was conducted on relationship status and parenthood in every wave. This included, for example, marital status, details (where applicable) of the partner's educational background, employment status and field of work, as well as the household situation. The questions regarding existing children in the third and fourth waves differed from those in previous waves in that they only asked about children who had been born since the respondent's last participation. From the fifth wave onwards, the total number of children and all dates of birth were surveyed again. Similarly, in the tenth wave, questions were asked about care responsibilities and detailed information on any interruptions to employment due to children.

⁴ From Wave 6 onwards, the time-consuming coding of occupations was discontinued for cost reasons.

2.3 Learning environment and contextual conditions

To assess the learning environment during the doctoral phase, a theoretically grounded and empirically tested model was developed for the first wave. The theoretical framework for learning environments is based on the SSCO model (*Structure – Support – Challenge – Orientation*) by Bäumer, Preis, Roßbach, Stecher & Klieme (2011). Details of the instrument can be found in the relevant publications (Brandt et al. 2016; de Vogel et al. 2017).

2.4 Wave-specific (focus) topics

In the third wave, key questions from the second wave were re-surveyed among participants who had not taken part in the second wave. These included, for example, information on their doctorate (publications, date of degree conferral, awards received), on international mobility before and during their studies, and on their citizenship. Furthermore, in each wave, specific (focus) topics were set. In the seventh wave, additional information was also collected on doctoral supervision and the principal examiners (in each case: gender, position, sympathie).

2.4.1 International mobility

In the third wave, the focus was on experiences of international mobility. In addition to the follow-up survey on stays abroad prior to completing the doctoral degree and the core instruments relating to international mobility, future mobility plans and the use of information and communication technologies (ICT) for international exchange were also surveyed. The use of ICT was also surveyed in the seventh and tenth waves.

2.4.2 Individual innovation behaviour

The questionnaire for the fourth wave contained several established instruments on the topic of individual innovation behaviour: data were collected on the dynamics of the professional environment (German adaptation by Schilke (2014)), problem-solving and decision-making autonomy at work (Morgeson and Humphrey (2006), German adaptation by Stegmann et al. (2010)) and individual ambidexterity (German adaptation by Kobarg et al. (2017)).

2.4.3 Personality traits

The instruments used to assess key personality traits, which had already been incorporated into the initial survey, were employed once again in the fourth wave. These include the short psychometric scales for measuring general self-efficacy (ASKU) (Beierlein et al. 2014b), for measuring the 'Big Five' personality factors (BFI-10) (Rammstedt et al. 2014) and for measuring internal and external locus of control (IE-4) (Kovaleva et al. 2014). The ASKU and BFI-10 were also administered in the seventh and tenth waves.

2.4.4 Health and disability

The thematic focus of the fifth wave was on health and disability. A short instrument was used to conduct a subjective general health survey. Three questions were also asked to survey physical and mental health, as well as health-related limitations in everyday life. These questions are a German-language adaptation of the Healthy Days Core Module (CDC HRQOL-4) (Centres for Disease Control and Prevention (CDC) 2000), which was also used, for example, in the fifth initial cohort of the NEPS (National Educational Panel Study (NEPS) 2018). Finally, information was collected on disabilities and, where applicable, the type, year of diagnosis and degree of disability. These questions are instruments that were already used in the sixth cohort of the NEPS (National Educational Panel Study (NEPS) 2013). All health-related instruments are also part of the Nacaps questionnaire (National Academic Panel Study). As data on health status are classified as special categories of personal data under the General

Data Protection Regulation (GDPR), the respondents' explicit consent was sought once again for these questions.

2.4.5 Political participation

In the sixth wave, the political interest and political engagement of doctoral graduates were surveyed. The six questions were closely modelled on those used in surveys conducted as part of the Eurograduate project and were derived from the European Social Survey (ESS) 2014 (questions 5.2–5.5) and ESS 2016 (question 5.6). Question 5.1, in which political interest was assessed on a five-point scale, was combined with a methodological experiment. The reason for this was that the direction of the scale in the original instrument did not correspond to the usual scale direction used in the PhD graduate survey. Consequently, all respondents were randomly shown one of the two scale directions (Version 1: from 1 'very interested' to 5 'not at all interested'; Version 2: from 1 'not at all interested' to 5 'very interested') to test whether the responses differed. The dataset contains a generated variable in which all respondents have the same scale direction, achieved by reversing the responses of one of the two groups (k6polinter).

2.4.6 Questions on COVID-19

In the seventh and eighth survey waves, questions were asked about the impact of COVID-19. The questions are taken from a supplementary survey conducted by the National Educational Panel Study (NEPS) (cohorts 2–6) in May 2020. Respondents were asked whether their work situation had changed as a result of the COVID-19 pandemic, whether the order or income situation of self-employed individuals had changed, and to what extent various areas of life had been affected.

3 Population and sampling method

3.1 Population

In line with the definition used in the Federal Government's reporting, 'early-career researchers' are understood to mean higher education graduates who, after completing their studies, are engaged in academic work or are pursuing a doctoral degree within or outside higher education institutions or non-university institutions. Additionally, doctoral graduates who are engaged in or wish to engage in academic work within or outside higher education institutions and non-university institutions are also counted. The population comprises all doctoral graduates who completed their doctorate in the 2014 examination year (winter semester 2013/14 and summer semester 2014) at a German higher education institution with the right to award doctorates. Official statistics show that 28,147 doctoral graduates were recorded nationwide for the 2014 examination year (Federal Statistical Office, 2015).

3.2 Comprehensive survey

As a first step, the management teams of all (at that time) 146 higher education institutions with the right to award doctorates were informed about the planned survey and asked to support the research project. For data protection reasons, the initial contact with the target group was made via the higher education institutions. For the follow-up surveys, the participants' contact details were recorded and continuously updated between the survey waves (7 survey waves). The initial sample for the panel survey comprised all respondents who had agreed to continue participating and for whom valid contact details were available. The number of people contacted decreased in subsequent waves due to withdrawals and contact details that had become invalid.

4 Survey administration

4.1 Initial survey

[Initial contact and address maintenance] For data protection reasons, initial contact with doctoral graduates was made via the dean's offices or the examination offices responsible for doctorates. In most cases, responsibility lay with the individual dean's offices. At a few higher education institutions, central or decentralised examination offices or graduate academies were identified as the responsible bodies.

[Survey materials] In the first survey wave, the survey materials for each respondent consisted of a cover letter (including data protection information), the questionnaire (paper-and-pencil), an accompanying letter from the Federal Ministry of Education and Research (BMBF) and a postage-paid envelope addressed to the DZHW for returning the completed questionnaire. To encourage participation, two reminder letters were also sent by post.

[Field period] The survey period for the first survey wave ran from 15 December 2014 to 17 February 2016. The two reminder letters were sent out approximately four and eight weeks, respectively, after the start of the fieldwork. This lengthy interval was due to the fact that the administrative departments of the higher education institutions responded at varying speeds to the project's request. Not all administrative offices could be persuaded to send out the reminders. However, for the majority of administrative offices, the survey materials and the two reminder letters were dispatched as planned. Due to the contact procedure used via the examination offices, the DZHW had no direct influence over the exact timing of the dispatch of the survey materials.

[Measures to increase response rate] The measures to increase response rate were aimed, on the one hand, at providing incentives for higher education institutions to provide organisational support for the survey and, on the other hand, at motivating individual respondents. As a token of thanks for their support, the higher education institutions received exclusive preliminary results from the survey. Faculties could, upon request, receive special analyses for their own higher education institution, provided there were sufficient case numbers available. The measures to increase the response rate aimed at motivating individual respondents involved, firstly, highlighting the subjective and societal relevance of the topic in the cover letter to respondents. In addition, the accompanying letter from the Federal Ministry of Education and Research (BMBF) in the first wave outlined the study's relevance to higher education policy and sought support. Furthermore, reminder letters were sent out. By providing feedback on selected study results following the first wave, the aim was, on the one hand, to strengthen the participants' commitment to the panel in order to motivate them to take part in subsequent waves. On the other hand, the distribution of results served to update the address database once again. In addition, a laptop, three iPads and several book vouchers were raffled off amongst all participants in the first wave.

4.2 Follow-up surveys

[Establishing contact and address maintenance] An address database was created for the follow-up surveys based on the contact details collected in the initial survey. To comply with data protection regulations, survey and address data were linked via an identification number and stored separately. The address databases were continuously updated between survey waves. In addition, project results, announcements of forthcoming surveys and, from certain waves onwards, postal communications

(e.g. advance notification letters or Christmas cards) were used to strengthen panel retention and verify that contact details were up to date. From the sixth wave onwards, contact was made exclusively by email.

[Survey documentation] The follow-up surveys were conducted as online surveys using the DZHW online survey software 'Zofar'. Invitations were sent predominantly by email and contained a personalised access link to the survey, information on data protection, and details of the announced incentives. In some waves, supplementary advance notices or interim invitations were sent by post. Non-participants received up to three reminder letters in each instance. With the introduction of the General Data Protection Regulation (GDPR), consent to the data protection provisions was obtained via an opt-in procedure.

[Field period] The follow-up surveys were conducted annually between 2016 and 2024. Each field period lasted several weeks. After the invitation letters had been sent out, up to three reminder letters were sent out in stages to individuals who had not yet taken part in the respective survey. The exact dates of the respective field periods can be found in the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW)'s metadata system.

[Measures to increase response rates] To encourage participation, various measures were implemented to maintain the panel and increase response rates. These included regular updates on project findings, the provision of a project website, advance notices of the surveys, and postal contact campaigns. In addition, prizes and vouchers were raffled off in all follow-up surveys. The reminder letters and ongoing communication with participants also served to strengthen panel loyalty and ensure long-term willingness to participate.

5 Response rate

Table 1 provides an overview of the number of survey documents and invitations sent out, as well as the response rates for each survey wave. In the initial survey, around 19,900 questionnaires were sent out via the participating higher education institutions. A total of 5,423 questionnaires were received⁵, of which 5,408 cases were included in the survey dataset after adjustment for blank, duplicate and illegible questionnaires, as well as subsequent panel adjustments.

For the second survey wave, 4,816 people (89 per cent of those who took part in the initial survey) agreed to be contacted again. The response rates for the follow-up surveys ranged between 55.6 per cent and 66.1 per cent and remained largely stable up to the fifth wave. In waves six and seven, they fell to around 56 per cent and remained relatively stable after that. Across all survey waves, there was a marked increase in the response rate following the dispatch of reminder letters, which underlines their effectiveness in boosting willingness to participate.

[Panel attrition] Over the course of the panel, there were the typical panel attritions seen in longitudinal studies. These resulted in particular from active withdrawals from the study, postal or email addresses that had become invalid, and, in isolated cases, from duplicate participation⁶. Furthermore, so-called ‘soft’ refusals occurred in all follow-up surveys, where the online questionnaire was accessed but then exited without any substantive responses being provided. Such cases were not counted as definitive panel attrition, as the individuals concerned were contacted again in subsequent waves. The number of people invited to participate decreased steadily over the course of the panel study due to withdrawals and participants who could no longer be reached. From the sixth survey wave onwards, contact was made exclusively by email, meaning that individuals without a valid email address could no longer be included in the survey. Despite these dropouts, which are typical of panel studies, a stable response rate was achieved over the entire survey period.⁷

Table 1: Gross and net response rates of the DZHW PhD Panel 2014

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
Survey materials sent out	19,916	4,816	4,809	4,795	4,778	4,558	4,526
of which dispatch confirmed	19,900 ⁸	4,815	4,806	4,789	4,771	4,524	4,500
Survey documents – response rate	5,423	3,334	2,976	3,037	3,085	2,585	2,625
Survey data usable	5,408	3,183	2,924	2,981	3,037	2,535	2,574
response rate (usable/sent)	27.2%	66.1%	60.8%	62.2%	63.6%	55.6%	56.9%

⁵ The usable survey data comprises all PhD graduates who consented to participate and provided at least one response in the online questionnaire.

⁶ In isolated cases, respondents submitted two questionnaires across at least two waves, resulting in duplicate cases with different IDs. Following identification, one case was retrospectively deleted by the primary research project.

⁷ Due to incomplete data this table can not be expanded to cover waves 8 to 10 currently.

Proportion of Wave 1 (usable cases / usable cases in Wave 1)	-	58.9%	54.1%	55.1%	56.2%	46.9%	47.6%
--	---	-------	-------	-------	-------	-------	-------

Table 2 : Participation of higher education institutions

Participating higher education institutions	Number
Higher education institutions with the right to award doctorates	146
No doctorates awarded in the 2014 examination year	19
Participated	112
➤ Full participation	80
➤ Partially participated	32
Did not participate	15

Figure 1: Response rates for the DZHW PhD Panel 2014 over time: Wave 2 to Wave 7

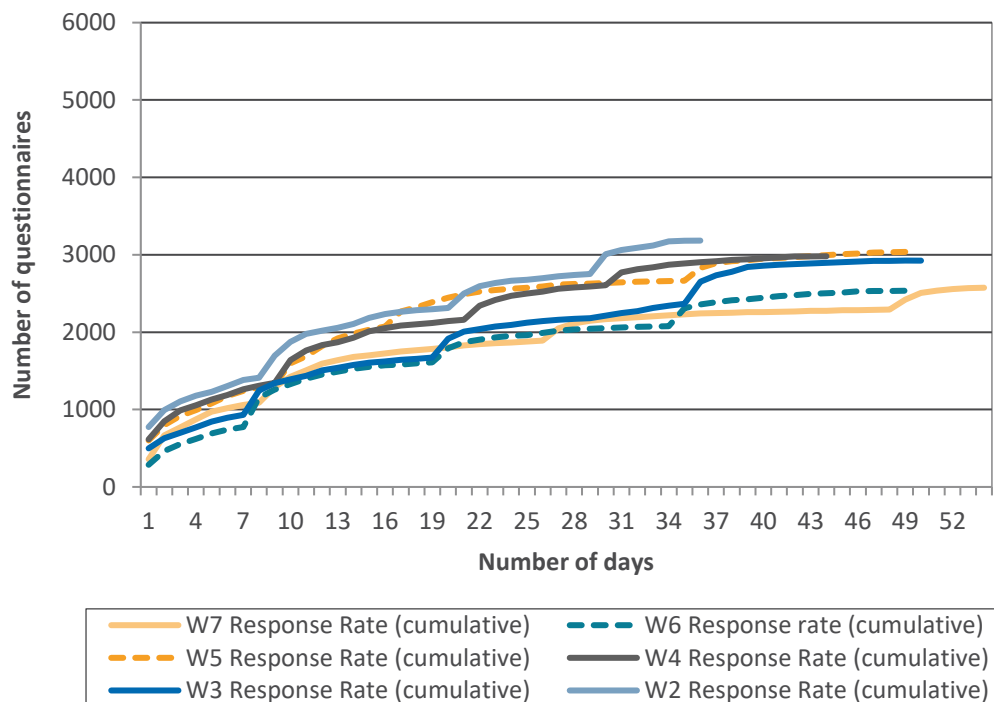


Figure 1 shows the response rates for the second to seventh rounds of the PhD Panel over time. Compared with the third wave, the curves for the second, fourth and fifth waves show a steeper rise both shortly after the start of the survey and following the individual reminder campaigns. In the fourth and fifth waves, a preliminary notification letter was sent by post before the start of each survey. For

the third wave, this was sent during the survey period. The response rates for these survey rounds converge towards the end of the survey period; however, the response rates for the fourth and fifth waves are slightly higher than those for the third wave.

The trends from the second to the fifth wave are very similar, but from the sixth wave onwards, the response rate is noticeably lower.

6 Data preparation

The data preparation steps described below were carried out following the data collection. The steps described in Chapters 6.1 to 6.3 – data transfer, coding of open responses, data checking and data cleaning – were carried out by the primary research project. For waves 8 to 10, parts of this work were carried out as part of the project ‘Optimisation of the translation process in the context of the internationalisation of the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) using the example of additional data publications (OptiDat)’, funded by the Federal Ministry of Research, Technology and Space.

The generation of variables (Chapter 6.4) was carried out both by the primary research project and as part of the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW)’s data editing process. The creation of the datasets, the assignment of variable names, variable and value labels, and the coding of missing values (Chapters 6.5 to 6.7) were carried out by the primary research team in consultation with the FDZ. The steps involved in weighting and anonymisation are described separately in Chapters 7 and 8.

6.1 Data transfer

In the first wave, the respondents’ answers from the paper questionnaires were transcribed into a computer-readable format for further processing, based on a code plan. Prior to this, numerical codes had been entered on the paper questionnaires for most open responses (see Chapter 6.2), and preliminary manual corrections had been made to facilitate the data transfer (see Chapter 6.3).

[Creation of a code plan] A code plan was drawn up based on the initial survey questionnaire. This specified which question or sub-question a variable was assigned to and what name that variable bore.

[Data entry] For data transfer, the code plan, further instructions on data entry and the prepared paper questionnaires were handed over to an external service provider. The data was entered manually there by data entry clerks.

From the second survey wave onwards, the data was collected using the online survey tool ‘Zofar’, and consequently no code plan or manual data entry was used. The raw data set was stored on a secure server.

6.2 Coding of open responses

[Initial survey] Prior to data transfer, most (semi-)open responses were coded. Numerical codes were assigned to these responses using a code list. Different code lists were used depending on the variable. These are classification codes from official statistics (e.g. classification of occupations, code list for student and examination statistics, etc.). For some variables, code lists were developed based on the responses found in the 2014 PhD Panel data. For some semi-open-ended questions, no new variables with numerical codes were created; instead, the responses were simply assigned – where possible – to the existing (closed) response categories. Individual open-ended questions were not coded because they were primarily recorded as contextual information to aid the coding of other open responses. The coding decisions made by the primary research project were retained unchanged.

[Follow-up surveys] Open responses were coded – where possible – using existing coding schemes from previous waves and classification keys from official statistics (e.g. the Classification of Occupations). The coding was carried out in the same way as in Wave 1.⁹

The company STR Coding was commissioned to code occupational activities, which are available for Waves 1 to 5 in the form of codes corresponding to the Classification of Occupations 2010 (KldB2010, the coding documentation is available in the metadata system). In addition to the open responses on job title, main area of activity and professional tasks, the stated occupational position and, in individual cases, further contextual information – such as the economic sector (industry) and the doctoral field – were also utilised. The assigned code reflects the field of activity and the occupational position. In this way, changes in activities (research, administrative or advisory, etc.) whilst remaining in the same post, as well as changes such as promotion to management positions, are captured. A change in the occupational code is therefore not always synonymous with a change of job.

Table 3 shows the coded characteristics and the code lists used for each wave. The dataset from the primary research team contains the coded numerical variables as well as the open-ended responses. In the SUF, on the other hand, only the numerical variables are reported for data protection reasons (see Chapter 8). The values of the individual variables for the SUF are documented in the metadata search portal¹⁰ (see also Chapter 6.6).

Table 3 : Coded characteristics and code lists used

Variable names ¹¹	Characteristic	code list
d1prohs	Doctors' College, higher education institution for graduation	German higher education institution: Destatis key directory for student and examination statistics (Winter semester 2013/2014 and Summer semester 2014) Foreign higher education institution: DZHW coding based on country codes
d1profach1-3, b1studber1-4, b1abs1-4	Doctoral field, subject, graduation	Destatis Key Directory for Student and Examination Statistics (Winter semester 2015/2016 and Summer semester 2016)
c1eberuf	Occupation	Destatis Classification of Occupations 2010
c1-5jberuf [c1-5jberuf]	Occupation	Destatis Classification of Occupations 2010
c*taet1plz – c*taet10plz	Place of work	Postcode directory
c*taet1land – c*taet6land	Country of employment	Destatis System of Countries and Regions 2014
m*studkomplla, m*studsempla, m*studprakla, m*studkursla, m*studsonla	Country of study	Destatis System of Countries and Regions 2014
m*grad1land, m*grad2land, m*grad3land, m*grad4land	Country of stay abroad	Destatis System of Countries and Regions 2014

⁹ This relates to areas of professional activity and duties. The open responses served as additional information for coding the job titles, which were also asked for in an open-ended format.

¹⁰ [https://metadata.fdz.dzhw.eu/#!/de/studies/stu-phd2014\\$](https://metadata.fdz.dzhw.eu/#!/de/studies/stu-phd2014$)

¹¹ The variable names given were assigned by the primary research group.

m*plandaofla, m*plandamfla, m*planzeitla, m*planforla, m*planwbila, m*plansonla	Country of the planned stay abroad	Destatis System of Countries and Regions 2014
k*natausl	Nationality	Destatis System of Countries and Regions 2014
c*weibfarzt, c*weibzarzt, c*weibpsycho, c*weibtarzt, c*weibfanw	Other professional continuing education	Classification into specified categories of continuing education, where possible
c*jbranche	Other economic sectors	Classification into predefined categories of economic sectors, where possible

The symbol '*' in the variable names stands for wave number 1–10.

For the following variables, no coding was applied to open responses:

- details of the doctoral award (surveyed only in Wave 3, Question 4.3),
- areas of activity and professional duties (in Waves 2 to 5, these were coded as occupational codes in c1eberuf or c1-5jberuf; from Wave 6 onwards, these codes are no longer used),
- the category for 'other' details when asked about the institution at which a professorship is based (surveyed only in Wave 3, Question 2.9),
- the type of professorship specified under 'Other' (surveyed in Waves 3 to 10),
- staff categories specified openly under 'Other' (surveyed in Waves 2 to 10) and
- the partner's vocational qualification, as specified openly under 'Other' (surveyed in Waves 2 and 3).

6.3 Data checking and data cleaning

[Preliminary manual correction of the initial survey] In the first wave, a manual check and, where necessary, an adjustment of respondents' answers on the paper questionnaires were carried out even before the data was transferred.¹² This was intended, on the one hand, to facilitate data capture. To this end, the format of the existing entries was modified. For example, entries that were difficult to read or crossings-out made by respondents were clarified; numerical values were entered right-aligned in the boxes provided; and verbal descriptions of grades were converted into numerical values (e.g. 'good' = 2.0).

Secondly, the manual check aimed to rectify any initial errors or inconsistencies in the respondents' answers even before the software-assisted correction (see below). For example, the entry 'none' had to be deleted if one or more articles were listed in a row under 'Total number'. Furthermore, checks were carried out to ensure that the details of employment activities in the activity matrix matched the corresponding details on employment (Wave 1, Questions 2.1 and 8.1–8.13). Any inconsistencies identified were resolved – where possible – by cross-referencing with other entries in the questionnaire; otherwise, a corresponding missing code (see Chapter 6.7) was assigned.

[Software-assisted correction] Following the data transfer in the first wave, a comprehensive check and correction of the data was carried out using software developed in-house by the DZHW and, from the second wave onwards, using the statistical programme Stata. The aim was, on the one hand, to identify errors arising from the preliminary manual correction and data transfer, and, on the other

¹² The number of corrections made was not documented centrally, but only on the paper questionnaires, and therefore cannot be systematically reconstructed.

hand, to identify further inconsistent responses from respondents that could not be checked during the preliminary manual correction stage. Valid value ranges and response combinations were defined and checked based on formal rules. The following types of checks were carried out:

- *Test of value ranges:* A test was carried out to ensure that the recorded value for a variable fell within the value range defined for that variable.
- *Test of adherence to filter procedures:* Based on the questionnaire's defined filter procedures, checks were carried out to determine, on the one hand, whether information should have been provided for the respondent in question but was missing (completeness check) and, on the other hand, whether information was provided for the respondent that should not have been (filter violations).
- *Test of variable consistency:* The consistency of the information within a questionnaire and across survey waves was checked. Detailed information on the primary research team's data editing can be found in the data editing concepts and the data versioning concept.

In the case of missing, incorrect or implausible values, the first step was to check, using the paper questionnaire or the survey data record, whether the relevant value had been transcribed incorrectly (or not at all). Otherwise, an attempt was made to deduce the correct value from other information in the questionnaire. In cases of doubt, a specific missing code was assigned (see Chapter 6.7). Error corrections were documented and checked by at least one other person.

[Deletion of cases] Over the course of the panel, several cases were removed from the dataset. Two cases were deleted because duplicate participation was identified. Deletions were also made where fewer than one question had been answered (five cases) or where there were too many inconsistencies (three cases). On three occasions, the entire case had to be removed from the dataset due to a withdrawal.

6.4 Generation of variables

In addition to the variables containing the respondents' coded answers, the 2014 PhD Graduates Panel dataset also includes generated variables. These consist, on the one hand, of variables with numerical codes derived from originally open responses (see Chapter 6.2). On the other hand, variables for the SUF were modified for data protection reasons (see Chapter 8). Furthermore, variables frequently required in the research field were generated from the values of one or more source variables (e.g. deriving the number of children from the information provided in the 'Children' table).

6.5 Creation of the datasets

[Merging of waves] Although the data from all survey waves were collected separately, they were merged into a single dataset. Cases were matched using the identification numbers assigned to respondents during the first field period (see Chapter 4).

[Creation of individual and spell data sets] The data merged in this way was stored by the FDZ in two separate datasets for the SUF. The individual data set contains the majority of the survey data as well as the additionally generated variables. There is one data row (wide format) for each respondent. The order of the variables follows the order of the corresponding questions in the questionnaire.

The spell data set, in long format, is largely based on information from the calendars, which, from the second wave onwards, provide month-by-month details of the respondents' activities (see, for example, Question 4.8, Wave 2). In Wave 2, respondents were asked about their activities since completing their doctorate up to the time of the survey; from Wave 3 onwards, the most recent survey date for the participant was used. In addition, an occupational profile was used in all waves (see, for example,

Question 2.3, Wave 1), in which respondents could provide detailed information on their professional activities. As both instruments contain information on professional activities, this was checked for plausibility in the primary research project.

One exception to this is the category ‘self-employment/teaching assignment/contract work/freelance work’ in the ‘status’ variable in the spell data set. This is based exclusively on information from the occupational profile. If information from Wave 1 was available but the respondent had not taken part in the second or third wave, the information was grouped here under the category ‘Self-employment/teaching assignment/contract for services/freelance work’, provided that no contrary information was specified in the event history calendar. As the table in the first waves cannot clearly distinguish between self-employment and a contract for services, these have been combined here.

[File format] All Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) datasets are provided in both Stata and SPSS formats.

6.6 Assignment of variable names, variable and value labels

[Assignment of variable and value labels] For variable and value labels, wording from the questionnaire was adopted or concise abbreviations of this wording were chosen. The variable labels are generally based on the corresponding question text. Depending on the question type, the value labels are based on the texts of the response categories or a combination of the texts from the question and the response categories. For generated variables based on specific classifications, the names of the classification keys were adopted verbatim for the value labels. Furthermore, the suffix ‘_g#’ is used to identify generated variables (see Chapter 6.4).

[Variable naming] The primary research team assigned variable names according to a fixed scheme. The first letter of the variable name indicates the subject area to which a variable can be assigned (e.g. ‘c’ stands for the subject area ‘Employment history, job characteristics’). The second character indicates the wave in which the variable was surveyed.¹³ Where a variable can be assigned to a specific variable block, a variable stem was assigned.¹⁴ Finally, a descriptive suffix reflects the specific content of the respective variable.¹⁵ Text variables relating to open-ended questions were marked with the suffix ‘txt’. A list of the subject areas and variable stems used to create the variable names can be found in Table 4.

Table 4: Classification of subject areas and variable sets

Thematic areas		Variable sets
A	Existing skills	
B	study trajectory	mop = month of graduation jp = year of graduation fachsem = degree semester examnote = exam mark punkte = marks abs = type of qualification studber = area of study hs = higher education institution
C	Employment history; Job characteristics	job = details in the job profile zu = Recommendations for future career ak = Academic activities ko = Contacts

¹³ Example: c3 means that the variable comes from the third wave

¹⁴ Example: c3j means that the variable belongs to the ‘Current/last job’ section

¹⁵ Example: c3jbranche identifies the variable ‘To which economic sector does/did the company or organisation in which you work/worked primarily belong?’

		neustell = Position following qualification for the PhD e = First post j = Current/most recent position zuf = Field-specific satisfaction stud = Activities during studies weib = continuing education for professionals ww = academic further qualifications beaus = anticipated employment prospects
D	Course of the doctorate; framework conditions for the doctorate	pv = doctoral agreements pub = academic publications kon = conferences/symposia mob = mobility experiences pat = patents nw = networks bpraxis = professional practice outside academia fin = funding
K	Social background (relationship status, children, demographics, school leaving qualification and social background)	geb = birth (date/place) part = relationship kind = children studb = higher education entrance qualification nat = citizenship
P	Personality traits, life goals and motivations	m = motives for doctoral study asku = General self-efficacy (short scale) kont = beliefs about control lz = importance of life goals bfi = Big Five Inventory mob = willingness to move
S	SSCO dimensions	strs = structure_security and stability strr = structure_regulations and control strg = structure_basic principles/equipment/conditions/opportunity structures chak = challenge_collaborative Research chal = challenge_general pressure to perform chav = challenge_pressure to publish orid = orientation_interdisciplinarity orii = orientation_internationality supf = support_academic support supe = support_emotional support supk = support_career planning supw = support_academic Activities sups = support_key Skills supa = support_expansion of academic networks orif = orientation_research orientation orip = orientation_practical orientation
W	Academic activities following a doctorate	pub = academic publications kon = conferences pat = patents wisskar = academic career prof = professorship gut = peer review activities foan = applications for research funding mitgl = memberships of professional associations/societies asv = academic self-governance
G	Non-employment	
M	Mobility	plan = mobility plans
X	Information about the questionnaire	

6.7 Coding of missing values

The Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) uses a comprehensive coding system for missing values to ensure consistent coding of missing data across various DZHW datasets. Missing entries have been coded as three-digit negative values. Table 5 illustrates the missing value coding system used.

The missing data can be categorised into three different groups: In the first two groups, a distinction is made between missing values due to respondents not answering questions (non-response) and missing values due to the filter procedure or the question being irrelevant to the respondents (not applicable). The third group comprises missing value codes assigned by the primary research project or the FDZ during data preparation (edited missing value). This group also includes the coding assigned to certain variables as a result of anonymisation measures (see Chapter 8).¹⁶

Table 5: Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) classification system for missing values

value range	Code	value label
-999 to -990: Non-response	-999	don't know
	-998	no answer
	-997	no answer (response category)
	-996	interview break-off
	-995	not participated (panel)
	-994	refused
-989 to -970: Not applicable	-989	filtered
	-988	does not apply
	-987	missing by design (questionnaire split)
	-986	missing by design (wave) ^a
	-985	missing by design (cohort) ^b
-969 to -950: Edited missing value	-969	unknown missing ^c
	-968	implausible value
	-967	anonymized
	-966	not determinable ^e
	-965	invalid multiple answer
-949 to -930: Item-specific missing values	-949	not applicable
-929 to -920: Other missing values	-929	data loss

^a This value is only assigned to datasets in long format.

^b This value is only assigned in pooled datasets.

^c This value is assigned when no cause can be determined.

^d Entries that are classified as implausible during the coding phase due to various factors are assigned this value. It may no longer be possible to reconstruct the exact cause.

^e This category is assigned when unambiguous coding is not possible, e.g. in the case of an open response that could not be coded because it is illegible.

¹⁶ A fourth group comprises special missing codes that were assigned only to individual items during the data preparation of a specific dataset.

7 Weighting

The weighting of the data serves to compensate for distortions in comparison with the defined population. This section begins with a general introduction to the procedure and a presentation of the weights generated. The weighting procedure is then described in detail.

7.1 Procedure and application notes

[Causes of sample bias] Two processes are generally responsible for sample bias:

- **Design-related bias:** Disproportionalities are deliberately created to increase the number of cases in certain relevant subgroups.
- **Non-response bias:** Dropout processes (e.g. non-participation, inability to contact respondents, loss in the post) lead to a reduced response rate and thus to a difference between the gross and net sample (see Chapter 5). If these attrition processes are unsystematic (*Missing Completely at Random*), they can be ignored. However, they are usually subject to a systematic attrition process (*Missing at Random*, *Missing Not at Random*), which requires modelling.¹⁷

[Conceptual approach] As the DZHW PhD Panel is a complete enumeration, no design-related disproportionalities were generated that would need to be adjusted for. Consequently, no design weights needed to be created, as all individuals have the same selection probability. For this dataset, however, an adjustment should be made using cross-sectional non-response weights, which are generated based on information about participants and non-participants.

In collaboration with weighting experts from Gesis and the DZHW, the weighting procedure was optimised in comparison with the SUF for the first two survey waves (DOI: 10.21249/DZHW:phd2014:2.0.0). As the study design of the PhD Panel allows for individual non-response in individual survey waves, only cross-sectional weights are provided in this SUF. To this end, in the first step, wave-specific non-response weights are generated using a logistic regression model with predictors from the initial survey (propensity scores). In a second step, these non-response weights are calibrated to the distribution of characteristics in the population using a raking algorithm. This results in smaller weights overall with lower variances. Outliers, defined as very small or very large weights, are finally trimmed.

In addition, an error in the allocation of doctoral graduates from Berlin higher education institutions was corrected during the calibration to the distribution of characteristics in the population. The weights provided are shown in Table 6.

Table 6: Weights provided for the 2014 DZHW PhD Panel

Variable name	Description
x1gewi	Calibrated weight for cross-sectional analyses of the first survey wave (trimmed)
x2geaus	Non-response weight for cross-sectional analyses of the second survey wave (trimmed)
x2gewi	Calibrated non-response weight (x2geaus) for cross-sectional analyses of the second survey wave (trimmed)
x3geaus	Non-response weight for cross-sectional analyses of the third survey wave (trimmed)

¹⁷ For a basic overview of the different forms of non-response processes, see Rubin 1976.

x3gewi	Calibrated non-response weight (x3geaus) for cross-sectional analyses of the third survey wave (trimmed)
x4geaus	Non-response weight for cross-sectional analyses of the fourth survey wave (trimmed)
x4gewi	calibrated non-response weight (x4geaus) for cross-sectional analyses of the fourth survey wave (trimmed)
x5geaus	Non-response weight for cross-sectional analyses of the fifth survey wave (trimmed)
x5gewi	calibrated non-response weight (x5geaus) for cross-sectional analyses of the fifth survey wave (trimmed)
x6geaus	Non-response weight for cross-sectional analyses of the sixth survey wave (trimmed)
x6gewi	calibrated non-response weight (x6geaus) for cross-sectional analyses of the sixth survey wave (trimmed)
x7geaus	Non-response weight for cross-sectional analyses of the seventh survey wave (trimmed)
x7gewi	Calibrated non-response weight (x7geaus) for cross-sectional analyses of the seventh survey wave (trimmed)

[Notes on the use of the weights] The weights created are probability weights, which can be applied in Stata using .ado-specific options.¹⁸ The weight x1gewi is used for weighting the first survey wave; the weights x2geaus and x2gewi are used for weighting in analyses of the second-wave data, and so on. The weights provided are only suitable for cross-sectional analyses. As panel dropouts and re-entries are possible between waves, longitudinal weights have not been provided as part of the SUF. However, users can create these themselves. It is essential to note that weights only constitute meaningful correction factors if the analytical model used contains the variables employed for weighting or is related to them. For this reason, weights must always be used with a focus on the specific research question. The procedure for creating the weights for waves 2 to 7 is described in more detail below.

7.2 Modelling of non-response weights

[Propensity Score Matching] For waves 2 to 7, the non-response weights were calculated using propensity scores.¹⁹ To this end, a logit estimation model was used to predict an individual's probability of participation (RP) for the relevant wave. An exploratory approach was used to identify individual characteristics that help explain the probability of participation.²⁰ In order to estimate the probability of participation for each individual, missing values for all variables were incorporated into the model as an additional category. The key criterion for modelling the non-response processes was to find models with the greatest possible explanatory power whilst remaining parsimonious, i.e. keeping the increase in variance as low as possible. Only predictors from the initial survey were used for this purpose.²¹ In a second step, the conditional probability of participation was then derived from the model, the reciprocal of this represents the non-response weight (x2geaus-x7geaus) for the respective wave: $w = 1/RP$.

[Estimation models] The estimation models used to calculate the non-response weights include the following variables: *gender, age, place of birth abroad, type of doctoral degree, doctoral field, doctoral degree mark, laziness (Big Five) and gross income*. The logit models for waves two to seven can be found in the Appendix in Table 9, Table 10, Table 11 and Table 12.

¹⁸ See the Stata help (command: help weights).

¹⁹ The method is logically consistent with propensity score matching, which dates back to Rosenbaum and Rubin 1983 (see also Blumensiel and Gummer 2015).

²⁰ See the estimation model in the appendix.

²¹ Attempts to take account of the panel nature of the present dataset—for example, by including participation (yes/no), the time of participation, or predictors from previous waves in the estimation models—did indeed increase the explanatory power of the non-response models; however, these had to be discarded due to excessively large weightings and variances.

7.3 Calibration

[IPF raking method] The non-response weights for waves 2 to 7 (x2geaus–x7geaus) were adjusted to the marginal distribution of known characteristics in the population using the IPF raking method (Iterative Proportional Fitting).²² As the raking method requires only information on the marginal distribution of the characteristics and does not necessitate cross-tables, there is little need to merge cells and fewer individual case decisions need to be made. The adjustment of the weightings follows an iterative process until no further improvement is possible. As no non-response weights were calculated for the initial survey, only calibration is carried out for this survey.²³

[Characteristic distribution of the population] The samples were matched to the population on the basis of the characteristics ‘area of study’, ‘gender’ and ‘region of the doctoral institution where the PhD was completed’ (East/West). The information on the population was taken from data provided by the Federal Statistical Office (ICE-Land data portal; dataset: 50001). Groups of individuals whose proportion in the respective wave does not correspond to their proportion in the population are weighted up or down accordingly using the calibrated non-response weight (x1gewi-x7gewi).

There was a particular issue in the procedure for creating the weighting variables for the sixth wave, as the dataset contains two individuals whose doctoral field cannot be assigned to any of the areas of study. According to official statistics, this category also exists in the population. However, in Wave 6 (and only in Wave 6), neither of these individuals took part, meaning that, in theory, this category would have to be combined with a similar or higher-level category in the population table. However, this assignment was not possible, as this specific category cannot be assigned to any other category. It was also not possible to exclude these cases from the population. The primary research team therefore decided on the following procedure: one of the two cases in the dataset was randomly recoded as a participant and included in the weighting procedure. It was then recoded as non-participating and the weighting variable set to ‘Missing’. As a result of this procedure, the mean value of the weighting variable x6gewi is not exactly 1, but deviates slightly downwards. It can be assumed that the calculated weights for the other cases also deviate slightly from the actual value.

[Missing values] When using the calibrated non-response weights, it should be noted that no calibration could be carried out for individuals who did not provide information on the variables relevant to the raking process (gender, area of study, federal state of the doctoral institution where the doctorate was obtained). Accordingly, these individuals do not receive calibrated non-response weights. The sample sizes (n) may therefore be smaller in analyses using calibrated weighting.

7.4 Trimming

To eliminate outliers, all weights were subjected to weight trimming according to Potter (1990) (see also Valliant et al. 2013, pp. 388ff.). The procedure is based on the assumption that the weights follow a probability distribution (beta distribution). All weights lying above the 99th percentile are truncated at this limit. The excess beyond the truncation is subsequently distributed amongst the remaining weights.

7.5 Distribution of the weights

The distribution statistics for the weights provided for waves one to seven can be found in Table 7:

²² The IPF raking method dates back to Deming and Stephan (1940) (see also Lomax and Norman (2016)) and was implemented by Kolenikov (2014) in the ‘stata.ado’ file ‘ipfraking’.

²³ Technically, this was achieved by first assigning a non-response weight of 1 to all participants in the initial survey, which was then calibrated in a second step using the raking procedure.

Table 7: Distribution of weighting for waves 1 to 7

Weight	n	Min	Max	SD	Variance
x1gewi	5,318	0.226	1.832	0.264	0.069
x2geaus	3,183	1.163	2.326	0.230	0.053
x2gewi	3,146	0.178	2.162	0.349	0.122
x3geaus	2,924	1,264	2.603	0.272	0.074
x3gewi	2,890	0.166	2.162	0.349	0.122
x4geaus	2,981	1.241	2.559	0.267	0.072
x4gewi	2,949	0.209	2.147	0.342	0.117
x5geaus	3,037	1.246	2.432	0.239	0.057
x5gewi	3,003	0.214	2.055	0.323	0.104
x6geaus	2,535	1.463	3.025	0.322	0.103
x6gewi	2,509	0.253	2.112	0.337	0.113
x7geaus	2,574	1.411	3.045	0.335	0.113
x7gewi	2,545	0.196	2.146	0.341	0.117

8 Anonymisation

[Legal Framework] The EU General Data Protection Regulation (GDPR) and the Federal Data Protection Act (BDSG) in its revised version of 30 June 2017 apply to personal data²⁴ collected by the DZHW in voluntary surveys.²⁵ Accordingly, personal data for secondary scientific use (without a declaration of consent for secondary use of the personal data) must be processed “in such a manner 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 para. 5 GDPR; see also Art. 89 GDPR and Recital 26 GDPR).

[Ensuring data protection at the FDZ-DZHW] At the FDZ-DZHW, the data protection of the respondents is ensured by a combination of legal-organisational, technical and statistical measures.²⁶ The combinations result in four potential access ways, which are summarized in Table 8.²⁷

Table 8: Modes of access at the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW)

	Campus Use File (CUF): Download	Scientific Use File (SUF): Download	Scientific Use File (SUF): Remote Desktop	Scientific Use File (SUF): On-Site
legal-organisational measures	very low	low	moderate	high
technical measures	very low	low	moderate	high
statistical measures	very high	high	moderate	low

The more data access is legally-organisationally and technically controlled and restricted, the lower the risk of de-anonymization of the data. As a result, the data must be anonymized less by means of statistical measures, i.e. less information has to be removed from it – which means the greater its analysis potential remains.

All modes of access are available for the data from the 2014 PhD Panel. The specific statistical anonymization measures carried out are explained in more detail below.

[Statistical anonymisation measures] As part of the statistical anonymisation measures, all information was first checked to determine whether it could be used to directly identify individuals. These

²⁴ “‘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 para. 1 GDPR).

²⁵ The GDPR generally applies within the EU and therefore also to the DZHW. The BDSG in its new version of 30 June 2017 (Act to Adapt Data Protection Law to the Regulation (EU) 2016/679 and to Implement Directive (EU) 2016/680, (Data Protection Adaptation and Implementation Act-EU (DSAnpUG-EU)) also applies in part, as the DZHW GmbH is legally considered a public body of the federal government (sect. 2 para. 3 BDSG). The federal government holds the absolute majority of shares in DZHW GmbH and the institute fulfils public administration tasks for the federal government in the broadest sense.

²⁶ The FDZ’s data protection concept is based on the portfolio approach of Lane, Heus and Mulcahy (2008, p. 6ff.), which is already used by the Leibniz Institute for Educational Trajectories (LifBi) (cf. Koberg, 2016, p. 699ff.) and the RDC of the German Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner, Müller, Schmucker, & Roß, 2012, p. 9f.).

²⁷ For further details, see <https://www.fdz.dzhw.eu/de/datennutzung>

direct identifiers, such as names, addresses and email addresses, were recorded in a separate dataset during the field period and are therefore not contained in the data per se. The original identification number was removed and replaced with a new, randomly assigned identification number.

In addition, Ebel and Meyermann's recommendation to delete open responses was followed even if the questions themselves are unproblematic. This is because there is a risk that study participants may have disclosed critical information in actually harmless questions with open response option that could lead to identification (cf. Ebel & Meyermann, 2015, p. 5).

After that, quasi-identifiers were determined, i.e. information that, in combination or by adding external information, is suitable for indirectly identifying a person. To prevent unambiguous assignment of the data to a specific person, it was determined for these key characteristics – depending on the access way – whether they may only be available in an aggregated form or not at all.

Finally, the data was checked for sensitive information, e.g. on health, sexual orientation and political attitudes. Although this information is not necessarily suitable for re-identification of individuals or institutions, it does pose a potentially increased risk of harm to respondents in the event of de-anonymization (cf. Koberg, 2016, p. 694) and is therefore particularly worthy of protection (Art. 9 GDPR, Recital 51 GDPR). For the sensitive information, it was also determined – depending on the access way – whether it may only be available in an aggregated form or not at all.

The dataset report, which can be found in the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW)'s [metadata system](#), provides further information on which variables are contained in the data provided and via which modes of access they can be utilised in each case. For example, the federal state of the place of work can be used in aggregated form, grouped into federal state groups (c*taet1bula_g1), via all modes of access. In its unaggregated form, the federal state can only be accessed via the 'Remote Desktop' and 'On-Site' modes of access (c*taet1bula).

9 Appendix

9.1 Logit regression to calculate non-response wave

Table 9: Logit regression to calculate the non-response weight for Wave 3

	Odds ratio	Std. Err	z	P > z
Gender				
Male	Ref.			
Female	1.10	0.07	1.49	0.14
Not specified	1.13	0.51	0.28	0.78
Age				
23–30	1.04	0.08	0.57	0.57
31–34	Ref.			
35–39	1.01	0.08	0.16	0.87
40–44	1.25	0.17	1.64	0.10
45–49	1.90	0.36	3.38	0.00
50–60	1.75	0.37	2.66	0.01
61–79	2.17	0.86	1.97	0.05
Not specified	0.12	0.07	-3.71	0.00
PhD funding framework				
Research assistant, funded post	0.95	0.08	-0.56	0.58
Research assistant, third-party funding post	Ref.			
Structured doctoral programme	1.03	0.12	0.28	0.78
Scholarship programme	0.79	0.09	-2.14	0.03
self-funded doctoral candidate	0.92	0.09	-0.91	0.36
Not specified	0.62	0.25	-1.17	0.24
PhD mark				
summa cum laude	1.32	0.13	2.89	0.00
magna cum laude	1.17	0.09	2.08	0.04
cum laude	Ref.			
satis bene	1.24	0.47	0.56	0.57
rite	0.96	0.20	-0.22	0.83
Not specified	0.52	0.17	-2.01	0.04

Place of birth				
Germany	Ref.			
In another country	0.68	0.07	-3.98	0.00
Not specified	0.90	0.21	-0.44	0.66
"I'm lazy and tend to be idle"				
doesn't apply at all	0.78	0.06	-3.21	0.00
2	0.90	0.07	-1.28	0.20
3	Ref.			
4	1.12	0.13	0.93	0.35
completely true	0.87	0.20	-0.61	0.54
Not specified	0.33	0.18	-2.01	0.05
subject group				
Humanities	1.19	0.14	1.43	0.15
Sport	0.68	0.23	-1.11	0.27
Law, Economics and Social Sciences	0.82	0.07	-2.23	0.03
Mathematics, Natural Sciences	Ref.			
Human Medicine/Health Sciences	0.73	0.07	-3.21	0.00
Agricultural, Forestry, Nutritional Sciences, Veterinary Medicine	0.91	0.13	-0.66	0.51
Engineering	0.82	0.08	-2.04	0.04
Art, Art History	1.02	0.25	0.08	0.94
Not specified	1.06	0.43	0.15	0.88
Current/most recent gross income				
up to €3,200	1.01	0.08	0.11	0.91
€3,201 to €5,000	Ref.			
Over €5,000	0.88	0.07	-1.63	0.10
no income	0.83	0.10	-1.51	0.13
Not specified	0.56	0.06	-5.08	0.00
Constant	1.45	0.17	3.16	0.00
N=5408; pseudo R ² =0.025				

Table 10: Logit regression to calculate the non-response weight for Wave 4

	Odds ratio	Std. Err	z	P > z
Gender				
Male	Ref.			
Female	1.07	0.07	1.17	0.24
Not specified	0.83	0.37	-0.43	0.67
Age				
23–30	1.05	0.08	0.70	0.48
31–34	Ref.			
35–39	0.89	0.07	-1.57	0.12
40–44	0.95	0.13	-0.36	0.72
45–49	1.81	0.34	3.13	0.00
50–60	1.52	0.32	2.03	0.04
61–79	1.25	0.47	0.60	0.55
Not specified	0.11	0.06	-3.87	0.00
PhD funding framework				
Research assistant, funded post	0.93	0.08	-0.94	0.35
Research Assistant, External Funding Unit	Ref.			
Structured doctoral programme	0.95	0.11	-0.45	0.66
Scholarship programme	0.87	0.09	-1.26	0.21
self-funded doctoral candidate	0.85	0.08	-1.78	0.08
Not specified	0.60	0.25	-1.25	0.21
PhD mark				
summa cum laude	1.27	0.12	2.44	0.02
magna cum laude	1.12	0.09	1.51	0.13
cum laude	Ref.			
satis bene	1.07	0.41	0.18	0.86
rite	0.75	0.16	-1.37	0.17
Not specified	0.59	0.19	-1.66	0.10
Place of birth				
Germany	Ref.			
In another country	0.66	0.06	-4.43	0.00
Not specified	1.06	0.25	0.23	0.82

"I'm lazy and tend to be idle"				
doesn't apply at all	0.84	0.07	-2.26	0.02
2	0.95	0.07	-0.72	0.47
3	Ref.			
4	1.28	0.15	2.10	0.04
completely true	1.12	0.26	0.50	0.62
Not specified	0.68	0.35	-0.75	0.45
subject group				
Humanities	0.99	0.12	-0.09	0.93
Sport	0.71	0.24	-1.00	0.32
Law, Economics and Social Sciences	0.87	0.08	-1.59	0.11
Mathematics, Natural Sciences	Ref.			
Human Medicine/Health Sciences	0.69	0.07	-3.66	0.00
Agricultural, Forestry, Nutritional Sciences, Veterinary Medicine	0.90	0.13	-0.69	0.49
Engineering	0.84	0.08	-1.76	0.08
Art, Art History	1.10	0.27	0.37	0.71
Not specified	1.07	0.44	0.17	0.87
current/most recent gross income				
up to €3,200	1.02	0.08	0.31	0.76
€3,201 to €5,000	Ref.			
Over €5,000	0.91	0.07	-1.18	0.24
no income	0.95	0.12	-0.39	0.69
Not specified	0.55	0.06	-5.22	0.00
Constant	1.61	0.19	4.04	0.00
N=5408; pseudo R ² =0.026				

Table 11: Logit regression to calculate the non-response weight for Wave 5

	Odds ratio	Std. Err	z	P > z
Gender				
Male	Ref.			
Female	1.05	0.07	0.85	0.40
not specified	1.11	0.51	0.23	0.82
Age				
23–30	1.05	0.08	0.67	0.50
31–34	Ref.			
35–39	0.99	0.08	-0.11	0.92
40–44	1.34	0.19	2.10	0.04
45–49	2.10	0.41	3.78	0.00
50–60	1.59	0.33	2.20	0.03
61–79	1.00	0.37	0.00	1.00
not specified	0.08	0.05	-3.93	0.00
PhD programme framework				
Research assistant, funded post	0.83	0.07	-2.32	0.02
Research assistant, third-party funding post	Ref.			
Structured doctoral programme	0.98	0.12	-0.20	0.84
Scholarship programme	0.85	0.09	-1.53	0.13
self-funded doctoral candidate	0.89	0.08	-1.28	0.20
not specified	0.55	0.23	-1.44	0.15
PhD mark				
summa cum laude	1.27	0.12	2.44	0.02
magna cum laude	1.18	0.09	2.15	0.03
cum laude	Ref.			
satis bene	0.89	0.34	-0.30	0.76
rite	1.17	0.24	0.75	0.45
not specified	0.83	0.26	-0.61	0.54
Place of birth				
Germany	Ref.			
In another country	0.72	0.07	-3.47	0.00
not specified	0.93	0.22	-0.29	0.78

"I'm lazy and tend to be idle"				
doesn't apply at all	0.76	0.06	-3.47	0.00
2	0.83	0.06	-2.45	0.01
3	Ref.			
4	1.07	0.13	0.53	0.59
completely true	0.83	0.19	-0.79	0.43
not specified	0.57	0.29	-1.12	0.26
subject group				
Humanities	1.07	0.13	0.60	0.55
Sport	1.11	0.39	0.29	0.77
Law, Economics and Social Sciences	0.84	0.07	-2.03	0.04
Mathematics, Natural Sciences	Ref.			
Human Medicine/Health Sciences	0.73	0.07	-3.10	0.00
Agricultural, Forestry, Nutritional Sciences, Veterinary Medicine	0.76	0.11	-1.94	0.05
Engineering	0.82	0.08	-2.05	0.04
Art, Art History	0.89	0.22	-0.50	0.62
not specified	1.22	0.51	0.48	0.63
Current/most recent gross income				
up to €3,200	0.97	0.07	-0.37	0.71
€3,201 to €5,000	Ref.			
Over €5,000	0.91	0.07	-1.26	0.21
no income	0.89	0.11	-0.94	0.35
not specified	0.50	0.06	-6.16	0.00
Constant	1.77	0.21	4.86	0.00
N=5408; pseudo R ² =0.025				

Table 12: Logit regression to calculate the non-response weight for Wave 6

	Odds ratio	Std. Err	z	P > z
Gender				
Male	Ref.			
Female	1.07	0.07	1.14	0.26
Not specified	0.78	0.35	-0.57	0.57
Age				
23–30	1.07	0.08	0.96	0.34
31–34	Ref.			
35–39	0.91	0.07	-1.25	0.21
40–44	1.27	0.17	1.75	0.08
45–49	1.40	0.25	1.83	0.07
50–60	1.19	0.24	0.87	0.39
61–79 and no information provided	0.25	0.09	-3.88	0.00
PhD programme framework				
Research Assistant (funded post)	0.89	0.07	-1.43	0.15
Research assistant, third-party funding post	Ref.			
Structured doctoral programme	1.09	0.13	0.76	0.45
Scholarship programme	0.91	0.10	-0.84	0.40
self-funded doctoral candidate	0.95	0.09	-0.52	0.61
Not specified	0.86	0.35	-0.38	0.71
PhD mark				
summa cum laude	1.21	0.12	1.98	0.05
magna cum laude	1.15	0.09	1.79	0.07
cum laude	Ref.			
satis bene	0.85	0.33	-0.41	0.69
rite	1.24	0.26	1.04	0.30
Not specified	0.98	0.31	-0.07	0.94
Place of birth				
Germany	Ref.			
In another country	0.69	0.07	-3.82	0.00
Not specified	0.75	0.18	-1.22	0.22
"I'm lazy and tend to be idle"				

doesn't apply at all	0.86	0.07	-2.00	0.05
2	1.00	0.08	0.02	0.99
3	Ref.			
4	1.32	0.15	2.42	0.02
completely true	1.01	0.23	0.04	0.97
Not specified	0.77	0.40	-0.50	0.62
subject group				
Humanities	1.07	0.13	0.60	0.55
Sport	0.64	0.22	-1.29	0.20
Law, Economics and Social Sciences	0.81	0.07	-2.36	0.02
Mathematics, Natural Sciences	Ref.			
Human Medicine/Health Sciences	0.76	0.08	-2.74	0.01
Agricultural, Forestry, Nutritional Sciences, Veterinary Medicine	0.79	0.11	-1.69	0.10
Engineering	0.88	0.09	-1.32	0.19
Art, Art History	1.13	0.28	0.50	0.62
Not specified	0.89	0.35	-0.30	0.77
current/most recent gross income				
up to €3,200	1.01	0.08	0.17	0.87
€3,201 to €5,000	Ref.			
Over €5,000	0.87	0.07	-1.76	0.08
no income	0.84	0.10	-1.41	0.16
Not specified	0.60	0.07	-4.42	0.00
Constant	1.04	0.12	0.32	0.75
N=5408; pseudo R ² =0.019				

	Odds ratio	Standard error	z	P > z
Gender				
Male	Ref.			
Female	1.06	0.06	0.94	0.35
not specified	0.80	0.37	-0.48	0.63
Age				
23–30	1.18	0.09	2.21	0.03
31–34	Ref.			
35–39	0.96	0.07	-0.49	0.63
40–44	1.30	0.18	1.89	0.06
45–49	1.91	0.35	3.49	0.00

50–60	1.67	0.34	2.50	0.01
61–79	0.63	0.25	-1.17	0.24
not specified	0.12	0.08	-3.33	0.00
PhD funding framework				
Research assistant, funded post	0.95	0.08	-0.66	0.51
Research assistant, third-party funding post	Ref.			
Structured doctoral programme	1.08	0.12	0.63	0.53
Scholarship programme	0.89	0.10	-1.10	0.27
self-funded doctoral candidate	0.94	0.09	-0.63	0.53
not specified	0.72	0.30	-0.79	0.43
PhD mark				
summa cum laude	1.26	0.12	2.36	0.02
magna cum laude	1.13	0.09	1.64	0.10
cum laude	Ref.			
satis bene	0.68	0.27	-0.96	0.34
rite	1.19	0.25	0.84	0.40
not specified	1.17	0.36	0.50	0.62
Place of birth				
Germany	Ref.			
In another country	0.66	0.06	-4.25	0.00
not specified	0.96	0.23	-0.17	0.86
"I'm lazy and tend to be idle"				
doesn't apply at all	0.83	0.06	-2.42	0.02
2	0.98	0.08	-0.26	0.79
3	Ref.			
4	1.35	0.16	2.61	0.01
completely true	0.93	0.21	-0.34	0.74
not specified	0.39	0.23	-1.62	0.11
subject group				
Humanities	1.05	0.12	0.38	0.70
Sport	0.68	0.24	-1.10	0.27
Law, Economics and Social Sciences	0.85	0.07	-1.85	0.07
Mathematics, Natural Sciences	Ref.			
Human Medicine/Health Sciences	0.72	0.07	-3.26	0.00

Agricultural, Forestry, Nutritional Sciences, Veterinary Medicine	0.81	0.12	-1.46	0.14
Engineering	0.81	0.08	-2.15	0.03
Art, Art History	1.01	0.24	0.03	0.98
not specified	1.21	0.49	0.48	0.63
current/most recent gross income				
up to €3,200	1.03	0.08	0.37	0.710
€3,201 to €5,000	Ref.			
Over €5,000	0.93	0.07	-0.96	0.34
no income	0.88	0.11	-1.09	0.28
not specified	0.60	0.07	-4.38	0.00
Constant	1.04	0.12	0.30	0.77
N=5408; pseudo R ² =0.022				

9.2 Documentation of origin of secondarily used instruments

Due to the volume of data involved, the documentation detailing the origin of reused questions is contained in a separate file. This can be accessed [here](#).

10 Bibliography

- Beierlein, C., Kovaleva, A., László, Z., Kempler, C.J. & Rammstedt, B. (2014a). A single-item scale for measuring general life satisfaction: The Short Life Satisfaction Scale-1 (L-1). Mannheim: GESIS – Leibniz Institute for the Social Sciences (GESIS Working Papers, 2014|33).
- Beierlein, C., Kovaleva, A., Kemper, C.J. & Rammstedt, B. (2014b). General Self-Efficacy Short Scale (ASKU). A compilation of social science items and scales. Edited by GESIS – Leibniz Institute for the Social Sciences.
- Blumenstiel, J. E. & Gummer, T. (2015). Prevention, correction or both? Three approaches to reducing non-response bias using propensity scores. In: Jürgen Schupp and Christof Wolf (eds.): Non-response Bias. Quality assurance in social science surveys. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 13–44.
- Brandt, G., de Vogel, S. & Jaksztat, S. (2016). Development and testing of an instrument for assessing the learning environment during the doctoral phase. Results of the development study. Workshop report. German Centre for Higher Education and Science Studies. Hanover.
- Centres for Disease Control and Prevention (CDC) (2000). Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. Atlanta, Georgia: CDC. Available online at https://archive.cdc.gov/www_cdc_gov/hrqol/pdfs/mhd.pdf, last accessed on 7 July 2025.
- de Vogel, S., Brandt, G. & Jaksztat, S. (2017). A tool for assessing the learning environment during the doctoral phase. In: Journal of Empirical Higher Education Research (1), pp. 24–44.
- Deming, W. E., Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. In: Ann. Math. Statist. 11 (4), pp. 427–444. DOI: 10.1214/aoms/1177731829.
- Ebel, T. (2015). Recommendations on the anonymisation of quantitative data. GESIS – Leibniz Institute for the Social Sciences. Mannheim.
- Ebel, T., & Meyermann, A. (2015). Guidance on the anonymisation of quantitative data. Research data: Education Informed. Vol. 3. Frankfurt am Main: German Institute for International Educational Research. <https://doi.org/10.25656/01:21970> (Accessed on: 14 October 2024).
- Hochfellner, D., Müller, D., Schmucker, A., & Roß, E. (2012). Data protection at the Research Data Centre. (FDZ Methods Report 06/2012). Nuremberg: Research Data Centre (FDZ) of the Federal Employment Agency at the Department for Labour Market Research (IAB). https://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf (Accessed on: 15 July 2025).
- Jaksztat, S. (2017). Gender and academic productivity. In: Journal of Sociology 46 (5), pp. 347–361.
- Jaksztat, S., Brandt, G., de Vogel, S. & Briedis, K. (2017). Here to stay? The doctorate as a stepping stone to academic careers. In: WSI-Mitteilungen 70 (5), pp. 321–329.

- Kobarg, S., Wollersheim, J., Welp, I. M., Spörrle, M. (2017). Individual Ambidexterity and Performance in the Public Sector: A Multilevel Analysis. In: *International Public Management Journal* 20 (2), pp. 226–260. DOI: 10.1080/10967494.2015.1129379.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer, & J. Skopek (eds.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (p. 691–708). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-11994-2_38
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). In: *The Stata Journal* 14 (1), pp. 22–59.
- Kovaleva, A., Beierlein, C., Kemper, C. J. & Rammstedt, B. (2014). Internal-External Control Beliefs-4 (IE-4). Compilation of social science items and scales. Edited by GESIS – Leibniz Institute for the Social Sciences.
- 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.
- Lomax, N. & Norman, P. (2016). Estimating Population Attribute Values in a Table: “Get Me Started in” Iterative Proportional Fitting. In: *The Professional Geographer* 68 (3), pp. 451–461. DOI: 10.1080/00330124.2015.1099449.
- Mom, T. J. M., van den Bosch, F. A. J. & Volberda, H. W. (2009). Understanding Variation in Managers’ Ambidexterity: Investigating Direct and Interaction Effects of Formal Structural and Personal Coordination Mechanisms. In: **Organisation Science** 20 (4), pp. 812–828. DOI: 10.1287/orsc.1090.0427.
- Morgeson, F. P. & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. In: *The Journal of Applied Psychology* 91 (6), pp. 1321–1339. DOI: 10.1037/0021-9010.91.6.1321.
- National Educational Panel Study (NEPS) (2013). Cohort 6: Adults (SC6). Waves 2 and 3. Survey instruments (field version). Bamberg: LfBi.
- National Educational Panel Study (NEPS) (2018). Survey instruments (field version). NEPS Cohort 5 — Students. Higher education and labour market entry, wave 10. Bamberg: LfBi.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In: *Proceedings of the Survey Research Methods Section*, pp. 225–230.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). Big Five Inventory (BFI-10). Compilation of social science items and scales. Edited by GESIS – Leibniz Institute for the Social Sciences.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. In: *Biometrika* 70 (1), pp. 41–55.
- Rubin, D. B. (1976). Inference and missing data. In: *Biometrika* 63 (2), pp. 581–592.
- Scherer, S. & Brüderl, J. (2010). Sequence data analysis. In: Christof Wolf and Henning Best (eds.): *Handbook of Social Science Data Analysis*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 1031–1051.

- Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. In: *Strat. Mgmt. J.* 35 (2), pp. 179–203. DOI: 10.1002/smj.2099.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methods of Empirical Social Research*. 7th ed. Munich: Oldenbourg.
- Federal Statistical Office (2015): *Education and Culture. Examinations at Higher Education Institutions 2014*. Edited by the Federal Statistical Office. Federal Statistical Office. Wiesbaden.
- Stegmann, S., van Dick, R., Ullrich, J., Charalambous, J., Menzel, B., Egold, N. & Wu, T. T.-C. (2010). The Work Design Questionnaire. In: *Journal of Work and Organisational Psychology A&O* 54 (1), pp. 1–28. DOI: 10.1026/0932-4089/a000002.
- Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York.