Karsten Becker | Ulrike Schwabe | Daniel Völk | Jonas Koopmann | Frederike Gerdes |
Theresa Schommer | Thorsten Euler | Tina Oestreich | Daniel Klein

The Student Survey in Germany 2021

Data and Methods Report on the study

DZHW

German Centre for Higher Education Research
and Science Studies

|

Author(s)
Karsten Becker
Ulrike Schwabe
Daniel Völk
Jonas Koopmann
Frederike Gerdes
Theresa Schommer
Thorsten Euler
Tina Oestreich
Daniel Klein

With co-operation of
Mareike Beuße
Lennart Damaske
Marie-Christin Ehrhardt
Martina Kroher
Julia Steinkühler

Table of Contents

List of tables and figures

3

1 Content and design of the study

[Study design] " The Student Survey in Germany" (2021) is a newly designed study that integrates three previously separate student surveys (Beuße et al., 2022): the Social Survey (Sozialerhebung; Middendorff et al., 2017; Middendorff & Wallis, 2023), the Student Survey (Studierendensurvey; Multrus et al., 2017; Multrus, 2021), and the survey "best - Studying with Impairment" (best – Studieren mit einer gesundheitlichen Beeinträchtigung; Poskowsky et al., 2018; Unger et al., 2012).[1] A separate survey strand is also used to collect German data for the European education monitoring programme EUROSTUDENT (DZHW, 2018; Cuppen et al., 2023).[2]

The study "The Student Survey in Germany" (2021) is designed as a cross-sectional survey with a long-term character. This means that the cross-sectional surveys will be conducted regularly at four-year intervals. The resulting survey series will make it possible to map changes concerning the situation of students in Germany over time. The new, integrated student survey thus continues the long tradition of the Social Survey and the Student Survey as important cross-sectional long-term observational studies for describing and analysing higher education and the student population in Germany, as well as the much shorter tradition of the previous "best studies".

[Content] The survey covered a variety of topics relating to the study situation and the different life situations of students in Germany. The content programme is very diverse, as the thematic focal points of the separate, previous surveys were integrated. In the tradition of the Social Survey, the economic and social situation of students is analysed (Middendorff, 2022). This is supplemented by topics from the Student Survey (Multrus et al., 2017). For the first time the new study design enables comparisons between students with and without health impairments that have adverse effects on their studies (Steinkühler et al., 2023). Cross-connections between different thematic field can be established by combining survey contents this way. This results in a high analysis potential. More detailed explanations of the survey programme can be found in Chapter 2.

[Representativity] Not all students who were enrolled at a state, church, or private higher education institution in Germany in the summer semester 2021 received an invitation to take part in the survey or responded to it. One of the reasons for this is that a sample was drawn at several higher education institutions before the invitation was sent out. Nevertheless, to allow for representative statements about students in Germany in the summer semester 2021, the survey data was weighted in a two-step process. First, a design weight was created to consider the different inclusion probabilities of students. In the second step, a redressment weight was used to adjust the design weight to the official statistical data. The redressment weight is based on characteristics that are available in the official higher education statistics. More detailed explanations on weighting can be found in Chapter 7.

---

[1] Current information on "The Student Survey in Germany" (2021) can be found on the project website at the DZHW: https://www.dzhw.eu/forschung/projekt?pr_id=650.
[2] Every 18th survey participant was directed to a separate survey strand for the collection of German EUROSTUDENT data. The EUROSTUDENT data is also made available via the FDZ-DZHW.

[Utilisation potential] The study "The Student Survey in Germany" (2021) creates a comprehensive and unique database for research, education, and higher education policy in Germany (Beuße et al., 2022). The data can be used to answer a variety of scientifically and politically relevant questions. Based on the data from "The Student Survey in Germany" (2021), the report on the 22nd Social Survey (Kroher et al., 2023), a report on the study situation under online conditions (Multrus et al., 2023), and the report "best3 - Studying with Impairment" (Steinkühler et al., 2023) have already been produced. These are highly relevant for higher education institution monitoring. Supplementary policy papers also addressed current higher education policy issues relating to populist tendencies among students (Multrus et al., 2022), experiences of discrimination at higher education institutions (Meyer et al., 2022), the effects of inflation on the economic situation of students (Meier et al., 2023), the attractiveness of Master degree programmes (Stefani et al., 2023), and the consequences of the coronavirus pandemic for student mobility abroad (Netz & Völk, 2023).

2   Survey instruments

A standardised online questionnaire in German and English was used as the survey instrument. Section 2.1 presents the survey design, the central topic areas, and the secondarily used components of the survey instrument. Section 2.2 describes the cognitive and quantitative pretests carried out to evaluate and optimise the survey.

2.1   Survey design and survey content

[Modularised survey and split design] The integration of the previously separate courses of study (see Chapter 1) results in a variety of survey content. This conflicts with the time required for the respondents to answer all the questions. A modularised survey and split design was implemented in order to achieve a balance between the variety of topics and the survey burden (see Figure 1).

[Contents and structure of the survey programme] Figure 1 provides an overview of the survey programme of "The Student Survey in Germany" (2021). It consists of a comprehensive basic module A which addressed all students, supplemented by group-specific additional modules (e.g. for students with child(ren) or students who perceive their health impairment as having adverse effects on their studies), as well as randomly assigned thematic core modules and additional modules. This combination of a common basic set of questions, split design, and specific additional modules made it possible to cover a broad range of topics for both educational monitoring and scientific research. Specifically, the following information was collected in the modules: In module A, central basic and contextual information on the students and their course of study such as socio-demographic characteristics, study characteristics, students' previous education, and access to higher education or study planning and success were recorded. Additionally, basic information on the housing and employment situation as well as individual characteristics and the course of study of the students was collected. This module was supplemented by topic-specific core programmes on (a) the study situation and study conditions and (b) the employment, housing, and financial situation of students as well as modules for the groups (a) students with child(ren), (b) students with health impairments and (c) students with a migration background. Two additional modules focused on topics such as services for students at the higher education institutions, transitions to (further) courses of study, political orientations, and attitudes towards inequality and fairness. Secondly, the students were assigned to one of the three sub-modules "International students", "Students abroad" or "International mobility". This was based on their respective experience(s) or plans and the location where they obtained their higher education entrance qualification (in Germany or abroad).

[Randomised allocation in split design] In order to avoid self-selection of respondents into different subject areas and at the same time to achieve sufficiently large case numbers for later analyses, a survey design was developed on whose basis respondents were randomly assigned to the various questionnaire splits (B modules, D modules and E & F modules) according to predefined quotas. [3] The respondents were allocated using the two random variables "mastersplit" and "zusatzsplit" (additional split). The design stipulates that in the B module of the basic programme (Split I), half of the respondents are assigned to module B1 "Gender socialisation and diversity" and the other half to module B2 "Cultural framework conditions". In the in-depth core module (Split II), respondents were assigned to two of the three D modules, i.e. either the questions from D1 and D2, D1 and D3 *or* D2 and D3. In the additional modules E and F (Split III), the allocation was even more detailed. In some cases, random allocation to a module was combined with filter-based question guidance within the module. Respondents in module E2 "Transitions", for example, were asked follow-up questions on "Master's programme" (E2M), "Doctorate" (E2P) or "Beginning of studies" (E2S) depending on their respective study situation (see survey instrument). Allocation to the F modules was also randomised, with respondents in this case first going through the common catalogue of questions F1_1 to F1_5 (see survey instrument). They were then filtered and assigned to one of the three modules F1 "International students", F2 "Students abroad" and F3 "Mobility abroad". Except for basic module A, the number of cases available for analysis differs between the individual modules depending on the respondents assigned. This should be considered in analyses including variables from different modules.

Figure 1: The Student Survey in Germany" (2021) - Overview of the subject areas and the modularised survey structure
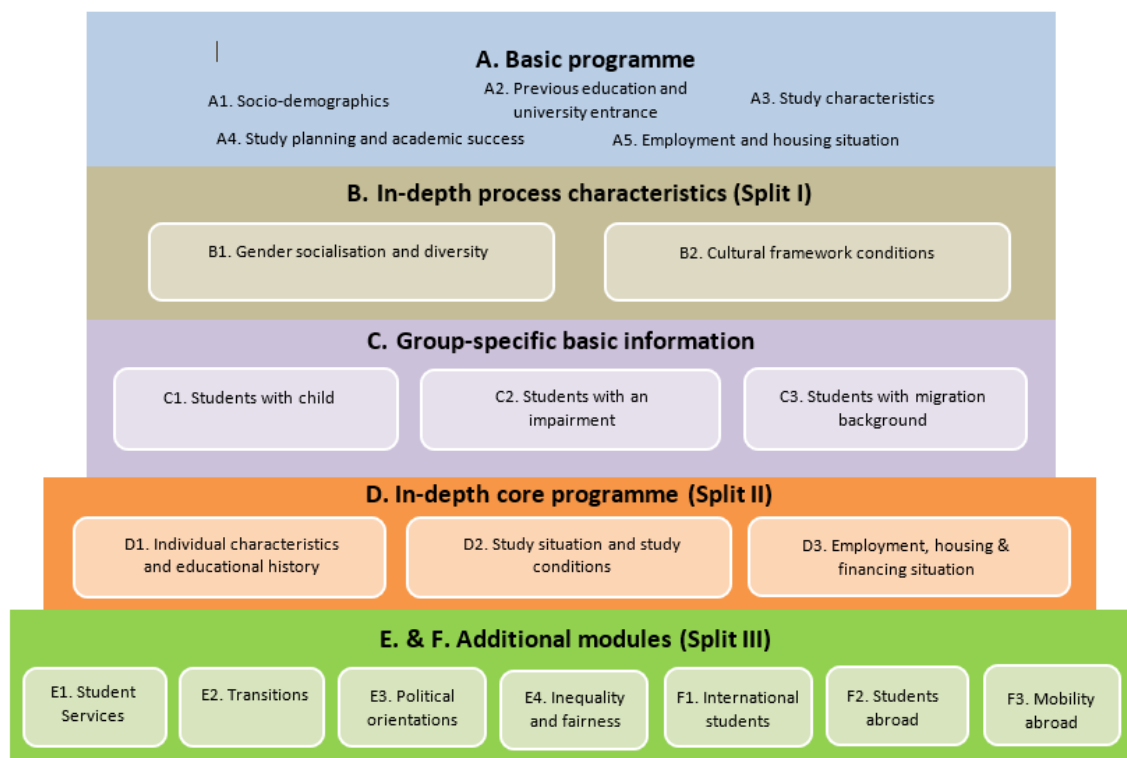
Source: "The Student survey in Germany" (2021). Own representation.

---

[3] To be able to generate a sufficiently high number of cases for individual subgroups, such as students abroad, their proportions in previous surveys of the Social Survey were taken into account in the quota system.

[Vignette studies] The study "The Student Survey in Germany" (2021) is also characterised by the fact that two vignette studies were implemented in module E2 "Transitions". The first on the Bachelor's-Master's studies transition (Stefani et al., 2023) and the second on the transition from Master's to doctoral studies.[4] Vignette studies are a research method in which the survey participants are presented with hypothetical scenarios, e.g. on transfer behaviour to the Master's degree programme (Auspurg & Hinz, 2015; Petzold & Netz, 2022). Here, variables of the outlined scenarios are systematically varied, randomly assigned to the survey participants, and the response behaviour is analysed depending on the presented variables (Stefani et al., 2023). Such a survey experiment combines the advantages of surveys with those of random experiments (Petzold & Netz, 2022). The respondents were presented with 432 different vignettes on the Bachelor's/Master's transition (108 sequences of 4 linked vignettes each) and 324 (162 sequences of 2 vignettes) on the transition from the Master's to the doctoral programme.

[Secondarily used components of the survey instrument] Established scales and questions were predominantly used in the design of the survey instruments. In Table 12 in the appendix you will find an overview of the components of the survey instruments used.

2.2   Pretests

[Aim and process] The survey instruments of "The Student Survey in Germany" (2021) were subjected to comprehensive pretests before being used in the field. Various pretest procedures were utilised, including cognitive pretests, expert pretests, and a quantitative pretest (Weichbold, 2022).

[Cognitive pretests] Cognitive pretests aim to test how respondents interpret questions and terms, what information they retrieve, how they come to a decision, and how they ultimately assign it to one of the given response categories (Lenzner et al., 2015; Prüfer & Rexroth, 2000). This active pretest procedure is used to test individual questions. By analysing which cognitive processes are activated in the respondents, information on the comprehensibility of questions and terms can be obtained, and problems identified. Students at the University of Hanover were recruited for the cognitive pretests.

[Expert pretests] In addition, the expertise of internal and external experts from the field of education and higher education research was obtained on selected instruments such as "academic and social integration" according to Tinto (1975).

[Quantitative pretest] A subsequent quantitative pretest aimed to test the preliminary survey instruments in the field under similar conditions[5] as the planned main survey. On the one hand, the implementation of the modularised split design, the duration of the survey and the general structure of the online survey were tested. For the quantitative pre-test, the questionnaire was programmed using the DZHW's own online survey software *Zofar* (Jungbauer-Gans, 2018). As this survey software was also used for the main survey it was possible to pre-test the process and functions before the survey was conducted. On the other hand, the responsive design of the survey software, i.e. the display of the questions on mobile devices with small screens, was tested and optimised.

---

[4] The data from the vignette studies are not provided in the individual data set, but in separate data sets for secondary use. The data can be merged via a person identifier (pid) in both datasets.

[5] The same respondent group (students at German higher education institutions) is surveyed using the same survey mode (standardised online survey, CAWI). In order not to influence the survey response process, the respondents were not informed that this was a pretest.

In addition, the data collected as part of the quantitative pretest was used to evaluate the instruments used regarding reliability and validity. For example, the instrument for recording the "Formal Learning Environment in Higher Education " (Schaeper & Weiß, 2016) was tested for its factorial structure and the subscales for their internal consistency. The HISBUS student panel at the DZHW was used for the quantitative pretest.[6] The fieldwork phase lasted three weeks in the winter semester 2019/2020 (07/11/2019-28/11/2019). A total of 1,779 students took part in the survey.

In addition, the paradata collected as part of the quantitative pretest were analysed. Paradata is process-generated data in online surveys, for example information on the time of participation in the survey, information on the end device used or the time spent on the page (Kreuter, 2013). The server-side and client-side paradata generated with the online survey software *Zofar* provide information on *how* people took part in the survey and thus also on data quality (Couper, 1998; Felderer et al., 2019). Analysing this information provides information about the response process but also about patterns of participation and the recruitment process. As part of the quantitative pretest conducted in the project, time stamps and page dwell times were analysed to estimate the time taken to complete the questionnaire as a whole as well as individual modules.

---

[6] The HISBUS student panel was an online access panel that was used at the DZHW from 1 August 2000 to 31 December 2020 (https://www.dzhw.eu/forschung/projekt?pr_id=273). Students were recruited from the DZHW's regular studies and surveyed several times a year on various topics. The HISBUS student panel was set up for the purpose of policy advice and student research with funding from the Federal Ministry of Education and Research (BMBF) and with methodological expertise from ZUMA Mannheim (for more information: http://www.hisbus.de/about/index_html).

3   Population, field access and sample procedure

[Population] The population of "The Student Survey in Germany" (2021) compromises all students who were enrolled at a higher education institution in Germany in the summer semester of 2021, apart from students at colleges of public administration. The population therefore includes students at public, ecclesiastical, and private higher education institutions. Unlike previous surveys conducted as part of the Social Survey (Becker et al., 2019; Middendorff et al., 2017), students at distance-learning universities are also part of the population. This also applies to students at dual universities. Apart from doctoral candidates who are also enrolled in a regular degree programme (e.g. medicine), German and non-mobile foreign ("Bildungsinländer*innen") doctoral candidates were not included. Foreign doctoral students are still included in the population, as they form a relevant group among international students (Beuße et al., 2022).

[Field access] As there is no centralised, nationwide student register with contact data in Germany, individuals had to be recruited via their higher education institutions (Beuße et al., 2022). In the first step, all higher education institutions in the defined population were invited to participate. At the level of the higher education institutions, this is therefore a complete enumeration. A total of 396 German higher education institutions were asked to participate (Beuße et al., 2022). Around 2.7 million students were enrolled at these higher education institutions at that time (Federal Statistical Office, 2022). These are students who obtained their higher education entrance qualification in Germany as well as students who obtained their higher education entrance qualification outside of it.

A total of 250 of the 396 higher education institutions decided to take part in the study. This corresponds to a participation rate at the higher education institution level of just over 63 percent. The participating higher education institutions include all 30 state universities and universities of applied sciences in the state of North Rhine-Westphalia. These higher education institutions also implemented the state's own project "State-wide student survey at universities in North Rhine-Westphalia" („Landesweite Studierendenbefragung an den Hochschulen in Nordrhein-Westfalen") and therefore participated fully in the survey (Gerdes & Völk, 2022). With around 2.3 million enrolled students, the participating higher education institutions represent 84 percent of the students in the defined population (Beuße et al., 2022; see Table 1). These form the sampling frame for the subsequent sampling at the student level.

Table 1: University participation in "The Student Survey in Germany" (2021)

Higher education institutions in the defined population	Quantity	Participation rate
Invited higher education institutions	396	
Participating higher education institutions	250	
Participation at university level		63,1 %

Source: "The Student Survey in Germany" (2021). Own calculations.

[Sample] In a second step, the current, semester-specific student lists of the participating universities were used for sampling at the student level. With a draw rate of just under 61 percent, around 1.4 million students were invited to take part (see Table 2). However, the draw rate varied between the universities depending on the chosen procedure: At all participating universities, at least every third student was randomly invited to take part in the survey. At most of the participating universities, however, a full survey was implemented. There were two reasons for this: Firstly, the aim was to generate a sufficiently high number of cases, even at smaller universities. Secondly, for organisational reasons, it was easier for some universities to carry out a full survey, i.e. to contact all students enrolled in the summer semester 2021.

Table 2: Contacting students for "The Student Survey in Germany" (2021)

Students at higher education institutions in the defined population	Quantity	Share
Students at invited higher education institutions	2.733.151	
Students at participating higher education institutions	2.295.075	
Selection framework for sampling at student level		84,0 %
Students after sampling by the participating higher education institutions	1.397.326	
Contact at student level		60,9 %

Source: "The Student Survey in Germany" (2021). Own calculations.

[Sample procedure and university cooperation] There is currently no nationwide (contact data) register for students. For this reason, the participating higher education institutions play an important gatekeeper role in contacting the students as an address intermediary between the survey project and the target group of selected students. Shortly before the start of the fieldwork phase, the contact persons or contact persons in the university administration drew the sample from the student lists available to them, unless a full survey was realised at the higher education institution in question.

Such an approach to contacting students requires the willingness to cooperate and support of the higher education institutions. The general requirement for sampling was to invite one third of the students from the defined population or to realise a full survey. In addition, the higher education institutions avoided excluding certain student groups from the sample to be drawn in advance, for example by identifying German doctoral students who are *not* enrolled in a regular degree programme. The survey instrument instead removes them at the beginning of the survey using an appropriate filter procedure(s).

4 Implementation of the survey

[Survey mode] "The Student Survey in Germany" (2021) was implemented as a standardised online survey (CAWI: Computer-Assisted Web Interview) (Wagner-Schelewsky & Hering, 2022). This survey mode enables the technical implementation of the complex survey and split design (see Chapter 2) and the realisation of a large sample size (see Chapter 3).

The survey instruments (see Chapter 2) were programmed using the DZHW's own online survey system *Zofar* and optimised for *mixed-devices* (Jungbauer-Gans, 2018; Weiß et al., 2022). This approach allowed respondents to answer the questions both on desktop PCs and on mobile devices such as tablets or smartphones. The latter are characterised above all by the fact that the screens are much smaller and touch-sensitive compared to desktop PCs (Weiß et al., 2022).

[Communication with higher education institutions] In January 2021, the management and in some cases other contact persons of all higher education institutions belonging to the population were contacted by email and informed in detail about the project (2021). All higher education institutions were asked to participate in the survey and to support the organisation of the fieldwork phase (see Chapter 3). In the period between January and May 2021, extensive communication with the higher education institutions that had not directly signalled their willingness to participate took place. The project team at the DZHW tried to convince these initially undecided higher education institutions to participate. A total of 250 higher education institutions had signed a declaration of participation at the start of the survey. With this declaration of participation, the universities also informed the project staff at the DZHW of fixed contact persons in the university administration who would implement the survey at the universities in organisational terms.

[Survey material] In order to ensure that the survey is conducted in as standardised a manner as possible, the participating higher education institutions were sent a "Manual for conducting the survey at the universities" by email. This included a detailed description of the entire survey organisation, e.g. how the random sample is drawn at the universities and how the survey links are to be sent to the students on a technical level. The higher education institutions also received the "Contact details and sample size" form with the handbook, which they filled out and sent to the DZHW before the survey began. The form was used to record updated data on the contact persons at the higher education institutions as well as the total number of students at the higher education institution and the number of students to be surveyed according to the sampling plan. The handbook contained another form, the so-called "field log", which the higher education institutions used to document information on the number of survey emails sent, including reminders. The fieldwork log was sent to the project team at DZHW at the end of the fieldwork phase and is an important prerequisite for calculating the response rates at student level (see Chapter 5).

[Technical Implementation] In order to distribute the server load that a survey of this size generates over the fieldwork phase, invitations to the survey were staggered in consultation with the participating higher education institutions. As a result, the invitations were sent out over a period of around two weeks (see Table 3).

As a first step the higher education institutions sent out the invitation letters to participate in the survey to the selected students under the guidance of the project staff at the DZHW. A personalised access link to the online survey, known as a "token", ensured that students could not participate in the survey more than once. This personalised access link also made it possible to interrupt the survey and continue it at a later time.

[Fieldwork phase] The fieldwork phase of the survey began on 4 May 2021 and ended for all higher education institutions on 19 September 2021 (see Table 3). From the start date, the higher education institutions sent out invitations to the survey to students by email over a period of two weeks, staggered by federal state.

After this initial invitation to participate in the survey, the higher education institutions were also asked to send reminders to those students who had not yet taken part in the survey or had not yet completed it. This served to motivate students to take part in the survey and thus increase the response rate (Göritz & Crutzen, 2012). At a total of three points in time, the higher education institutions were sent text templates for reminder letters as well as a daily updated list of those access links whose recipients should not be contacted again because they had already completed the questionnaire (see Table 3).

For internal organisational reasons, some higher education institutions asked to be able to carry out the fieldwork phase at a different time than planned. The documents were sent to these universities at a different time.

Table 3: Dispatch period of the survey documents

Dispatch periods[a]	
Start of fieldwork phase: dispatch of invitation letter	**04.05.2021**–18.05.2021
Dispatch 1st reminder:	approx. 2 weeks after field start
Dispatch 2nd reminder:	approx. 4 weeks after field start
Dispatch 3rd reminder:	29 June 2021 (for all higher education institutions)
End of fieldwork phase: survey is taken offline	**19.09.2021**

Source: "The student survey in Germany" (2021). Own representation.
Note:[a] = Some of the time periods were deviated from in order to take into account the special needs of the higher education institutions.

[Measures to increase response rates] The response rates for online surveys are lower than for other survey modes (Lozar Manfreda et al., 2008). For this reason, the survey was accompanied by intensive communication work. During the fieldwork phase, the higher education institutions were encouraged by the DZHW, the German National Association for Student Affairs (Deutsches Studierendenwerk; DSW), and the regional student services organisations to draw attention to the project via various communication channels and to actively promote it. For example, logos were provided for postings that the higher education institutions could publish on their social media channels (Twitter, Facebook, and Instagram). As the 2021 summer semester was still dominated by the coronavirus pandemic, only digital advertising formats were offered, and no print products were distributed. The podcast format "The podcast of the Student Survey in Germany" (Der Podcast der Studierendenbefragung in Deutschland) was used to discuss topics from the student survey with representatives from academia and politics and to publicise the survey and its relevance.[7]

Participation in the study "The Student Survey in Germany" (2021) was not incentivised. However, particularly small but important subgroups (e.g. students without a higher education entrance qualification) were offered 10-euro vouchers if they were willing to provide their contact details (email addresses) at the end of the survey. The aim of this post-paid incentivisation was to be able to address these subgroups again for subsequent DZHW surveys.

---

[7] The podcast episodes are available at https://einefueralle.podigee.io/episodes.

[Duration] The recommendation for online surveys is not to exceed a maximum survey duration of around 20 minutes, as the length can have a negative impact on participation and response behaviour (Galesic & Bosnjak, 2009). Due to the split design of the survey, the average survey duration is calculated at the level of the individual modules, which are then added together (see Table 4). On average, respondents took 32.3 minutes (median: 30.9 minutes) to complete the survey. If only those respondents who completed the questionnaire in full are considered, the mean value for this group is 43.6 minutes and the median is 39.2 minutes.

Table 4: Survey duration "The Student Survey in Germany" (2021)

Module	Mean value (in minutes)		Median (in minutes)	
	in total[a]	termina-ted[b]	in total[a]	termina-ted[b]
A. Basic module	9,1	11,6	8,3	10,0
B. In-depth process characteristics				
B1. Gender socialisation and diversity	4,2	5,0	3,7	4,3
B2. Cultural framework conditions	5,0	5,5	4,3	4,7
C. Group-specific basic information				
C1. Students with a child	2,2	2,2	1,8	1,9
C2. Students with disabilities	3,4	3,5	2,8	3,0
C3. Students with a migration background	0,4	0,4	0,3	0,3
D. In-depth core programme				
D1. Individual variables and educational progression	5,0	6,1	4,2	6,0
D2. Study situation and conditions	9,9	11,6	8,6	9,8
D3. Employment, housing, and financing situation	11,0	11,7	9,3	9,9
E. Additional module				
E1. Student Services	5,0	5,4	4,6	4,8
E2. Transitions	3,9	4,0	3,2	3,3
E3. Political orientation	4,4	4,6	3,8	3,9
E4. Inequality & Fairness	6,0	6,3	5,2	5,4
F. Additional module				
F1. International students	6,2	6,3	5,5	5,6
F2. Students abroad	4,7	4,8	4,2	4,3
F3. International mobility	2,3	2,3	1,9	1,9
Total survey	32,3	43,6	30,9	39,2

Source: "The Student Survey in Germany" (2021). Own calculations.

Note:[a] = All respondents are taken into account;[b] = Only those respondents who answered all questions of the respective module are taken into account.

5   Response Rate

[Development of response rate] During the fieldwork phase, a total of 1,397,326 invitations were sent out by the 250 participating higher education institutions (see Table 2). Figure 2 shows the gross response rate in absolute figures for every third day during the fieldwork phase and cumulatively over the entire fieldwork phase. As expected, the response rate rises sharply on the days on which invitations and reminders to participate were sent out and decreases towards the end of the fieldwork phase. In the period from May to September 2021, a total of 274,466 students[8] took part in the survey. This results in a gross response rate of 19.6 percent (see Table 5).

Figure 2: The Student Survey in Germany" (2021) - Development of the gross response rate over the entire fieldwork phase (in absolute figures)
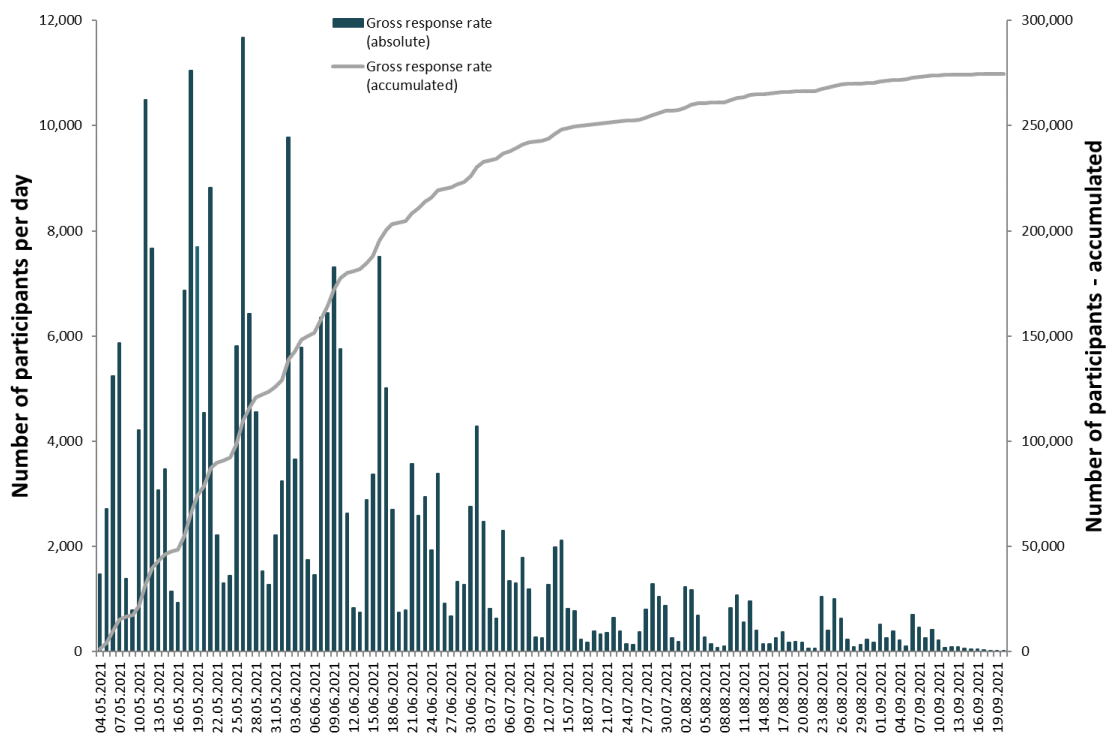
Source: "The Student Survey in Germany" (2021). Own representation.

[Response rate following AAPOR] The response rate was determined in relation to the standards of the American Association for Public Opinion Research (AAPOR 2023; see Table 5). In addition, further cases were removed to obtain an analysis data set of good quality. Of the total of 274,466 survey participants, the following were excluded from the analysis data set:

---

[8] The participation figures in absolute terms are higher than those shown in Figure 3, as this also includes respondents who only accessed the survey homepage or who did not give their informed consent.

- Respondents who provided information on less than half of the variables defined as significant (so-called non-valid cases according to AAPOR 2023 (37,089), see table 5)

- Respondents who only accessed the home page (28,667),

- Respondents who were filtered into the EUROSTUDENT survey at the beginning (13,915),

- Respondents who did not declare their informed consent (4,392),

- Respondents for whom speeding or straightlining was identified in the response behaviour (1,699),

- Respondents who made a subsequent data deletion request and test accesses (769)

After these exclusions, data from 187,935 students remained, which were defined as valid cases. This results in an adjusted response rate[9] of 13.4 percent at student level (see Table 5). The data from the 187,935 respondents can be used for analyses. Due to the split design of the survey, the number of cases in the different modules differs considerably in some cases (see chapter 2).

[Assessment of the response rate] Since "The Student Survey in Germany" was conducted for the first time in the summer semester 2021, there are no comparative figures from past surveys regarding the response rate. However, other studies can be used for comparison purposes: In the 21st Social Survey, another standardised online questionnaire was used with a similar population (Becker et al., 2019). The adjusted response rate was 16.2 percent for German students and students with a domestic education and 10.2 percent for students with a foreign education. Both paper (PAPI) and online (CAWI) questionnaires were used in the last Student Survey in the 2015/2016 winter semester (Multrus, 2021). The response rate for PAPI was 15.2 percent and 8.8 percent for CAWI (Multrus, 2021).

For the examples mentioned, there is no information in the respective method reports as to whether the AAPOR standards were used to calculate the response rates, which, together with the other exclusion criteria[10], led to a conservative calculation of the response rate in the present case. Nevertheless, cautious comparisons of the response rates are possible. Against the background of these figures and considering that the respondents were only surveyed online, were not reminded in writing, and were not (pre-paid) incentivised, the adjusted response rate of the study "The Student Survey in Germany" (2021) of 13.4 percent at student level (SSU) can be rated as very satisfactory overall.

---

[9] The "partial completers" are also included in the calculation of the adjusted sample, provided they have answered at least half of the variables defined as significant (AAPOR, 2016).
[10] The subsequent exclusion of cases that were included in the EUROSTUDENT survey and cases that were removed from the analysis data set due to speeding or straightlining must also be taken into account.

Table 5: Participating higher education institutions, sample size and response rate [11]

	Quantity	Share (in %)
Population		
Higher education institutions		
Higher education institutions in total [a]	396	
Participating higher education institutions	250	63.1 %
Students		
Total students [b]	2,733,151	
Students at participating higher education Institutions	2,295,075	84.0 %
Sample		
Students enrolled at the participating higher education institutions [c]	*1,397,326*	*60.9 %*
Response rate		
Gross response rate	*274,466*	*19.6 %*
of which …		
… non-valid cases [d]	37,089	2.7 %
… only home page accessed	28,667	2.1 %
… in the EUROSTUDENT survey [e]	13,915	1.0 %
… no consent Informed Consent	4,392	0.3 %
… Speeding & Straightlining [f]	1,699	0.1 %
… Data deletion requests & test accesses	769	0.1 %
Cases in the analysis data set (adjusted response rate)	**187,935**	**13.4 %**

Source: "The Student Survey in Germany" (2021). Own calculations.

Note:[a] All higher education institutions in Germany in the summer semester 2021 excluding colleges of public administration. Own calculations with data from the Federal Statistical Office (Federal Statistical Office, 2022);[b] All students in Germany in the summer semester 2021 excluding students at colleges of public administration. Own calculations with data from the Federal Statistical Office ;[1c] The participating higher education institutions either included a random selection of one third of the students in the sample or conducted a full survey. The sample comprises the students who were contacted. It does not include students without an email address or with an invalid email address (including "mail-delivery-failed" messages);[d] In accordance with the guidelines of the American Association for Public Opinion Research (AAPOR), respondents are considered non-valid cases if they have provided information on less than half of the variables defined as significant (AAPOR, 2023). For "The Student Survey in Germany", these are a total of 22 variables, including higher education institution, gender, desired degree, subject and type of study. [e] A randomly selected proportion of respondents (5.6 percent) were included in the survey of the international comparative study Eurostudent;[f] The implausible response behaviour of those who answered the entire questionnaire either very quickly (speeding) or with identical answers (straightlining) is excluded from the following analyses.

---

[11] The table can also be found in Kroher et al. (2023: 146).

6 Data preparation

The various steps of data preparation are described below. In detail, these are the data transfer, checking and cleaning (Section 6.1), the assignment of variable names, variable labels, and value labels (Section 6.2), the coding of missing values (Section 6.3), the coding of open responses (Section 6.4), the generation of variables (Section 6.5), and the creation of the dataset (Section 6.6). The activities described were carried out by the project team at the DZHW and at the University of Konstanz, in co-operation with the FDZ-DZHW. The data preparation steps were carried out almost exclusively with the *Stata* software. The weighting and anonymisation processes carried out as part of the data prep-aration are described in the chapters 7 and 8. The weighting was carried out by the project team, the anonymisation by the FDZ-DZHW team.

6.1 Data transfer, Data Checking and Data Cleansing

[Data transfer] The survey was conducted using the DZHW's own online survey software *Zofar.* The data from the online survey was exported directly from the survey software as a *.csv file* and further processed as a *Stata file* (.dta).

[Consistency and plausibility check] Following the data transfer*,* a comprehensive check of the data was carried out. The respective steps of this check were carried out syntax-based using Stata-Do-Files and executed in a rule-based procedure. The following types of checks were carried out:

[Checking the export data set] During the data collection, test accesses to the online survey was provided for the purpose of field checks. Project staff at the DZHW also received the emails sent by the higher education institutions via these test accesses. This gave the project team an insight into the (de facto) timing and quality of the invitation and reminder emails sent by the higher education insti-tutions. These test accesses were completely deleted from the data set after the conclusion of data collection. Furthermore, cases from the EUROSTUDENT survey or cases that (subsequently) withdrew their participation were deleted (see Chapter 5). Furthermore, any export inconsistencies (e.g. Javas-cript-based errors) or (missing) values inadequately assigned by the survey system were recoded in the raw data set.[12]

[Checking the value range] In order to reduce the number of implausible values in the data set from the outset, the questionnaire programming defined value ranges that should not be exceeded or fallen short of. If these value ranges were exceeded or undercut, the respondents were informed in red markings that their entry appeared implausible and were asked to correct the value. If no cor-rection was made after this notification, the survey could still be continued ("*soft-forcing*"). The data set was therefore checked to see whether the recorded value of a variable was within the value range defined for this variable.

---

[12] This concerns, for example, the coding of the category "please select", which was collected in the online questionnaire via a drop-down menu. Here, the export value "0" was recoded into an (item non-response) missing value.

[Test of adherence to filter procedures] Due to the modularised survey and split design, the filter procedures are complex. Therefore, on the one hand, it was checked whether information could have been expected for the respective respondent based on the filter procedure(s) provided in the questionnaire ("completeness check"), and on the other hand, whether no information should have been provided for the respective person ("filter violations") and the corresponding correct missing values were set. Due to the possibility of "going back" in the questionnaire and the possible correction of information, the relevant path for checking the course of the questionnaire and the filter procedure(s) is always the last saved path.

[Consistency check] After the filter check, a basic consistency check of the information within the questionnaire was carried out. Content-related information in the respective subject areas was specifically checked for possible contradictions.

[Definition of valid cases] The valid cases were determined based on several criteria in several steps. In a first step, incomplete questionnaires were analysed more closely. Following a guideline from the American Association for Public Opinion Research (AAPOR 2016), all cases in which at least 50 percent of the questions previously defined as significant were answered were determined to be valid. A total of 22 variables from the A module were used to define "significant". These are listed in Table 6 below.

Table 6: Variables defined as significant in the A module.

Questionnaire page	Variable description	Variable name
A_18	Higher education institution	preloadhs_id
A_17	Subject related semester	ssemfs
A_22	Satisfaction with university conditions	sperleiszufr
A_23	Subject	sfach1o2
A_25	Type of degree	sabsan
A_27	Form of study programme	sformpraes-sformsons
A_1	Place of acquisition of higher educ. entrance qual. (Germany vs. abroad)	vsbdeba
A_5	Place of birth (Germany vs. abroad)	dgebort
A_9b	Gender	demosex
A_12	Parenthood	dkinja
A_15	Disability	gbeges
A_16	Current stay abroad	imausl
A_31	Probability of successful graduation	studerfolg
A_32	Highest professional qualification Father	deltberuv
A_32	Highest professional qualification Mother	deltberum
A_34	Country of birth Father	deltgebv
A_34	Country of birth Mother	deltgebm
A_38	Highest school-leaving qualification	vsbart
A_42	Employment	eaktsens
A_51b	Intention to drop out	ssaaja
A_54	Previous study-related stays abroad	ainfaus
D1_9	Full/part-time study programme	sartzeit

Source: "The Student Survey in Germany" (2021). Own representation.

[Straightlining] Straightlining is an indicator of the survey participants' lack of motivation to give valid answers and has a negative effect on data quality (Roßmann, 2017). Following Roßmann (2017), straightlining is defined as the non-differentiation of answers in matrix questions. For example: if the same value is always given for approval items, regardless of the specific content of the item, e.g. "strongly agree" (value 5). Following the identification of the valid cases via the significant variables, the response behaviour on the relevant questionnaire pages was therefore analysed in a second step to identify these straightliners. According to the above definition, straightlining is present on the respective questionnaire page if all answers to a matrix question on a questionnaire page have the same value for a person.

According to Roßmann (2017), the questionnaire pages with at least three items with a 4, 5 or 6-point scale were considered ex-ante. Only questionnaire pages to which these criteria apply, but which ask for factual knowledge (e.g. parents' or grandparents' university degree, see question B2_10) or which were only presented to certain respondent groups (e.g. students with child(ren)), were not used for straightlining identification. One of the aims of this procedure was to reduce the probability of observing straightlining by chance (Roßmann, 2017: 87). A total of 65 questionnaire pages were included in the analysis (see Table 7).

Table 7: Overview of the straightlining-relevant questionnaire pages in "The Student Survey in Germany" (2021)

No.	Page	Scale	No.	Page	Scale
1	A_57	5-point scale with 12 items	34	E1_3	5-point scale with 11 items
2	B1_2	5-point scale with 7 items	35	E1_3a	5-point scale with 10 items
3	B1_3a	5-point scale with 6 items	36	E1_6	5-point scale with 13 items
4	B1_3b	5-point scale with 7 items	37	E1_11	5-point scale with 6 items
5	B2_6	5-point scale with 6 items	38	E2M_4	5-point scale with 4 items
6	B2_7	5-point scale with 7 items	39	E2M_5	4-point scale with 11 items
7	D1_18	5-point scale with 3 items	40	E2M_6	5-point scale with 13 items
8	D1_7	5-point scale with 12 items	41	E2M_7	5-point scale with 4 items
9	D1_8	5-point scale with 12 items	42	E2P_2	5-point scale with 18 items
10	D1_19	5-point scale with 4 items	43	E2S_2	5-point scale with 18 items
11	D1_20	5-point scale with 15 items	44	E2S_3	5-point scale with 12 items
12	D1_21	5-point scale with 7 items	45	E2S_4	5-point scale with 12 items
13	D1_24	5-point scale with 10 items	46	E2S_6	4-point scale with 7 items and 5-point scale with 7 items
14	D1_25	5-point scale with 3 items	47	E2S_7	6-point scale with 7 items
15	D1_29	5-point scale with 18 items	48	E2S_8	6-point scale with 11 items
16	D2_2	5-point scale with 9 items	49	E3_3	6-point scale with 10 items
17	D2_3	5-point scale with 10 items	50	E3_4	6-point scale with 13 items
18	D2_4	5-point scale with 6 items	51	E3_5	6-point scale with 5 items
19	D2_6a	5-point scale with 3 items	52	E3_6	5-point scale with 7 items
20	D2_7	5-point scale with 6 items	53	E4_3	5-point scale with 3 items
21	D2_9	5-point scale with 3 items	54	E4_4	5-point scale with 12 items
22	D2_11	5-point scale with 6 items	55	E4_5	6-point scale with 6 items
23	D2_12	5-point scale with 3 items	56	E4_8	5-point scale with 9 items
24	D2_13	5-point scale with 6 items	57	F1_2	5-point scale with 14 items

25	D2_16	5-point scale with 3 items	58	N_9	5-point scale with 5 items
26	D2_21	5-point scale with 11 items	59	N_9c	5-point scale with 4 items
27	D3_5	5-point scale with 5 items	60	N_11	5-point scale with 8 items
28	D3_11	5-point scale with 4 items	61	N_12	5-point scale with 9 items
29	D3_22	5-point scale with 4 items	62	CO_10	5-point scale with 7 items
30	D1_27	5-point scale with 4 items	63	N_17	5-point scale with 3 items
31	D3_25	5-point scale with 4 items	64	N_16	5-point scale with 8 items
32	E1_1	5-point scale with 5 items	65	B2_5	5-point scale with 5 items
33	E1_2	5-point scale with 13 items			

Source: "The Student Survey in Germany" (2021). Own representation.

Straightlining was indexed as a binary indicator variable for each of the questionnaire pages listed in Table 7. If straightlining was present, the indicator variable took the value "1", in all other cases the value "0".

Following on from this, a straightlining index per person i was calculated from the sum of the pages indexed with straightlining j (see Formula 1).

$$Straightlining - Index_i = \sum_{j=1}^{n_i} Straightlining_j \qquad (1)$$

Due to the modularised split design of the questionnaire, not all questions were presented to the respondents, which is why the straightlining index could assume values between $[0, 32] := \{x \in \mathbb{N} \mid 0 \leq x \leq 32\}$. As this index variable was right-skewed, the 99th percentile was 9.

Taking this into account, the straightlining rate was calculated based on the sum of the questions for which straightlining was indicated by the indicator variables ($j = 1$), and the number of question-naire pages seen or presented to respondents during the survey n (see Formula 2).

$$Straightlining - Quote_i = \frac{\sum_{j=1}^{n_i} Straightlining_j}{n_i} \qquad (2)$$

Finally, regarding the entire questionnaire, respondents were identified as straightliners if they showed straightlining in their response behaviour on more than 8 questionnaire pages (99th percen-tile, see above) and for whom straightlining was present on at least 50 percent of the questionnaire pages seen. In addition to the cases defined as invalid, the respondents defined as straightliners were also excluded from the data set.

[Speeding] In addition to straightlining, so-called speeding also has a negative impact on data qual-ity. Speeders are respondents who have gone through the questionnaire particularly or comparatively quickly. Here, too, a lack of motivation to provide valid answers is assumed. The total questionnaire duration of all respondents who completed the questionnaire was used to identify this group (see Table 4). Those respondents who completed the questionnaire faster than 99 percent of the other respondents were defined as speeders. Analogous to the calculation of the straightliners, the 99th percentile of the distribution of the speeding variable was also used here in accordance with distribu-tion theory assumptions. The respondents identified as feeders were deleted from the data set (see Table 5).

6.2 Assignment of variable names, variable labels, and value labels

[Assignment of variable names] The variable names were assigned using an internal scheme. The names reflect the respective variables or construct in short form (see also the variable questionnaire). There are a total of four different suffixes to the variable names in the data set:

- _p variables: Variables whose values have been plausibilised.

- _g variables: Variables generated during data preparation.[13][14]

- _pg variables: Variables that have been both plausibilised and generated.

- _a variables: Variables whose values have been aggregated.

[Assignment of variable and value labels] For the variable and value labels, verbalisations or short forms from the questionnaire were adopted or concise short forms of these formulations were chosen. The variable labels are generally based on the description of the content of the variables. Depending on the question type, the value labels are based on the texts of the answer options or a combination of the question text and answer option.

6.3 Coding of missing values

[Missing categories] The coding of missing values is based on the FDZ-DZHW system. Missing data are coded according to this scheme using three-digit negative values. Table 8 lists the categories that were assigned. The categories used in the study "The Student Survey in Germany" (2021) are highlighted in bold.

Table 8: System of the FDZ-DZHW for missing values

Value range	Code	value label
-999 to -990: Non-response	-999	don't know
	-998	**no answer**
	-997	no answer (response categories)
	-996	**interview break-off**
	-995	not participation (panel)
	-994	refused
-989 to -970: Not applicable	**-989**	**filter**
	-988	does not apply
	-987	**missing by design (questionnaire split)**
	-986	missing by design (wave) [a]
	-985	missing by design (cohort) [b]

---

[13] In the case of variables that were created by the FDZ-DZHW during anonymisation, the created variables are given the suffix _g# (# = number), as several new variables may have been created based on an original variable. The suffixes assigned by the primary research project were only revised by the FDZ-DZHW in exceptional cases.

[14] There are also generated additional variables without a specific ending (see also Table 13 and 14 in the appendix).

-969 to -950: Edited missing value	-969	unknown missing [c]
	-968	implausible value [d]
	-967	**anonymised**
	-966	**not determinable [e]**
	-965	invalid multiple entry
-949 to -930: Item-specific missing values	*(not assigned)*	
-929 to -920: Other missing values	-929	loss of data

[a] This value is only assigned for data records in long format.
[b] This value is only assigned in pooled data records.
[c] This value is assigned if 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 if clear coding is not possible.

[Survey break-off] To reconstruct survey break-offs, the page up to which the survey was completed by the respondents was checked. If the survey was cancelled prematurely, a missing for "interview break-off" was assigned for all subsequent pages (-996).

[Filter procedures] The online survey software *Zofar* works with output filters (*transitions*). These output filters from the programming template, the flowchart, and the *Zofar* QML file are translated page by page into input filters. The corresponding missing category for filter missings (-989) is assigned to survey participants who did not see a variable due to previously entered filter information.

[Missing by design] With the modularised survey and split design (see Chapter 2) not all respondents were presented with all questions. Depending on the randomly assigned preloads for the master and additional splits a certain path through the questionnaire was specified. Questions that were not presented due to the respective master split were assigned the missing category "missing by design (questionnaire split)" (-987).

Furthermore, additional splits were used for certain questions if questions were only to be seen by a subsample (e.g. only 12.5 percent of all respondents) or if two questions were alternatively shown to 50 percent of respondents each (*split-half design*). In these cases, too, if questions were not presented due to the respective design, the missing category for "missing due to design (questionnaire split)" (-987) is assigned.

["No answer"] After processing the above-mentioned missing categories, the remaining missings were checked again. As part of the module plausibility checks, missings were for example assigned for missing responses to multiple-choice questions.

[Alternative categories] For some questions, respondents were offered an additional alternative category alongside the response categories to avoid a previous, additional query. In this category, it was possible to express, among other things, that a fact to be answered did not apply, e.g. because no alleged grade was yet available. As a rule, these alternative categories were defined as outside the valid answer range of the questions and stored with a negative two-digit missing code with a unique label. The primary research project was responsible for determining whether this was actually a case of *missing* information.

After recoding the missings and checking the plausibility of the individual modules, the remaining missings were set to "unknown missing" (-969).

6.4   Coding of open responses

[Large number of open responses] The questionnaire of the study "The Student Survey in Germany" (2021) contains around 130 variables for which (semi-)open responses were possible. No open responses are provided for the use of secondary data in the Scientific Use File (SUF) and the Campus Use File (CUF). The project team at the DZHW and the University of Konstanz coded the (semi-)open responses and prepared them accordingly for secondary data use. The open responses were either assigned to existing response categories[15] or new variables or variable characteristics were generated. Two partially automated coding procedures, subject coding and occupation coding, are outlined below.

[Coding of subjects] The coding of the subjects was semi-automated using a *Python script* by the project team at the University of Konstanz (Weeber et al., 2024). The classification was based on the classification system for subjects, areas of study and subject groups of the Federal Statistical Office for the 2021 examination year[16] (Destatis, 2020). The special feature of the rule-based classification method used here was the requirement that no false-positive assignments should be made. Therefore, only assignments that were considered certain were made directly. For the remaining data, suggestions were made - where possible - from which coders could select the correct one or reassign it. The remaining data, for which neither a code could be directly assigned nor suggested, had to be coded on a case-by-case basis. The final classifications could then be added to the dataset as new variables and subsequently exported (for further details see Weeber et al., 2024). For secondary data users, the subjects, the classification of the areas of study or subject groups according to the Federal Statistical Office (variable *SB1_STB_p* or *FG1_STB_p*), and the project-specific classification of the areas of study used in the report of the 22nd Social Survey (variable *studfach*) are available in the dataset. In the report on the study situation under online conditions, the teaching post category was also reported (Multrus et al. 2023).

[Coding of occupations] The information on the occupations of the parents, the occupation learnt by the respondents, and the current and aspired occupation of the respondents was also recorded openly. During data preparation, this information was also coded semi-automatically. The automated coding was based on the classification of occupations of the Federal Statistical Office in the 2010 version (KldB, 2010) and according to an algorithm by Safikhani et al. (2023). All information that could not be clearly categorised was recoded by coders.

---

[15] If information was assigned to existing response categories, the consistency of the filter procedure(s) (see Chapter 6.1) may have been affected - if the response category concerned was used as a filter variable.
[16] This refers to the winter semester 2020/2021 and summer semester 2021.

6.5   Generation of variables

[Overview of generated variables] The data sets available for secondary data utilisation (partially) contain variables that were created by the project team at the DZHW and the University of Konstanz. An overview of the most important of these variables can be found in Table 13 and Table 14 in the appendix.

In the course of anonymisation for the respective access ways of the datasets, the FDZ-DZHW generated additional variables with aggregations of existing information. An overview of these can be found in Table 15 in the appendix .

6.6   Creation of the data set

[Data structure and file format] The available data sets as SUF and CUF contain the survey data as well as the additionally generated and aggregated variables. The order of the variables in the data set is based on the order of the corresponding questions in the questionnaire. The datasets are provided in *Stata* and *SPSS format* for secondary data utilisation.

7 Weighting

In this chapter section 7.1 provides a general introduction to the conceptual procedure and a presentation of the weightings created. The weighting procedure is then described in detail in section 7.2.

7.1 Conceptual approach

[Goal of weighting] The study "The Student Survey in Germany" (2021) aims to provide representative statements on the student body in Germany in the summer semester of 2021 (Beuße et al., 2022). For this purpose, it was necessary to scientifically weight the collected data (Kiesl, 2022). Weighting procedures make it possible to compensate for sample distortions, for example by gender, age or university affiliation, by giving underrepresented groups a higher weighting in the survey data. The weighting is based on the data from the official (higher education) statistics on the variables mentioned (HStatG, see Chapter 3). In addition, weighting procedures equalise the different selection probabilities of students at the participating higher education institutions (see Chapter 3). Table 9 provides an overview of the distribution of the weighting variables in the official statistics as well as in the unweighted and weighted data set.

[Conceptual approach] A multi-stage weighting procedure was carried out, which combines a design and a redressement weighting of the survey data:

(1) The design weighting initially takes into account the different inclusion probabilities of students.

(2) As no information on non-participants was available at the individual level, the next step was to create a redressement weighting based on the "Iterative Proportional Fitting Approach" (Kolenikov, 2014).

Table 9: Distribution of weighting variables, official statistics data, unweighted and weighted analysis data set (in %)

Weighting variables	Official statistics	Unweighted analysis data set	Difference	Weighted analysis data set
Type of university				
University (incl. art and music colleges)	62.4 %	60.0 %	2.4 %	62.3 %
Higher education institution for applied sciences	37.6 %	40.0 %	2.4 %	37.7 %
	100.0 %	100.0 %		100.0 %
Federal state of the place of study				
Baden-Württemberg	12.2 %	14.0 %	1.8 %	12.2 %
Bavaria	13.7 %	13.1 %	0.6 %	13.7 %
Berlin	6.9 %	4.0 %	2.9 %	6.9 %
Brandenburg	1.7 %	1.6 %	0.1 %	1.7 %
Bremen	1.3 %	0.9 %	0.4 %	1.3 %
Hamburg	4.1 %	2.0 %	2.1 %	4.1 %
Hesse	9.0 %	7.2 %	1.8 %	9.0 %
Mecklenburg-Western Pomerania	1.3 %	1.8 %	0.5 %	1.3 %
Lower Saxony	7.0 %	9.6 %	2.6 %	7.0 %
North Rhine-Westphalia	26.2 %	30.7 %	4.5 %	26.2 %
Rhineland-Palatinate	4.2 %	3.2 %	1.0 %	4.2 %
Saarland	1.0 %	1.5 %	0.5 %	1.0 %
Saxony	3.5 %	3.1 %	0.4 %	3.5 %
Saxony-Anhalt	1.8 %	2.7 %	0.9 %	1.9 %
Schleswig-Holstein	2.2 %	2.0 %	0.2 %	2.2 %
Thuringia	4.0 %	2.6 %	1.4 %	4.0 %
	100.0 %	100.0 %		100.0 %
Gender identity				
Male	50.1 %	39.1 %	11.0 %	49.6 %
Female	49.9 %	60.0 %	10.1 %	49.4 %
Diverse (non-binary)	n. a.	0.7 %	n. a.	0.7 %
Other gender identity	n. a.	0.3 %	n. a.	0.2 %
	100.0 %	100.0 %		100.0 %
Age				
Up to 19 years	6.1 %	7.3 %	1.2 %	6.1 %
20-22 years	27.1 %	35.5 %	8.4 %	27.1 %
23-25 years	27.8 %	30.4 %	2.6 %	27.8 %
26-30 years	23.0 %	17.3 %	5.7 %	23.0 %
31 years and older	16.0 %	9.5 %	6.5 %	16.0 %
	100.0 %	100.0 %		100.0 %

Continued on the next page

Continuation Table 9: Distribution of weighting variables, official statistics data, unweighted and weighted analysis data set (in %)

Weighting variables	Official statistics	Unweighted analysis data set	Difference	Weighted analysis data set
Area of study[a]				
Linguistics	6.3 %	6.0 %	0.3 %	6.0 %
Humanities (without linguistics)	4.9 %	4.6 %	0.3 %	5.2 %
Educational Science/Applied Social Science	7.4 %	11.7 %	4.3 %	8.0 %
Social/Political/Regional/Administrative Science, Psychology	8.1 %	8.7 %	0.6 %	7.4 %
Law	4.9 %	4.4 %	0.5 %	4.9 %
(International) Business Administration	10.3 %	9.5 %	0.8 %	10.0 %
Business and Economics	6.8 %	6.7 %	0.1 %	7.2 %
Mathematics, Natural Sciences	10.9 %	11.3 %	0.4 %	10.9 %
Medicine (Human and Dentistry)	4.2 %	4.8 %	0.6 %	4.2 %
Health Sciences (incl. Sport)	3.6 %	3.9 %	0.3 %	3.6 %
Agricultural, Forest and Nutritional Sciences, Veterinary Medicine	2.2 %	2.6 %	0.4 %	2.2 %
Computer Science	8.6 %	7.2 %	1.4 %	9.4 %
Engineering, Process/Electrical/Information technology	8.6 %	7.3 %	1.3 %	8.6 %
Other Engineering Studies	9.8 %	8.3 %	1.5 %	9.0 %
Art, Art Theory	3.4 %	3.0 %	0.4 %	3.4 %
	100.0 %	100.0 %		100.0 %
Place of residence at time of higher education entrance qualification				
Students from Germany	85.2 %	94.1 %	8.9 %	85.2 %
International students	14.8 %	5.9 %	8.9 %	14.8 %
	100.0 %	100.0 %		100.0 %

Source: Own calculations with data from the Federal Statistical Office (Federal Statistical Office, 2022) and with unweighted and weighted data from " The Student Survey in Germany " (2021). The analyses are based on data from 179,908 students.
Note: Due to rounding differences, summed values may deviate from 100.0 percent. n.a. = not available in the official statistics (Fachserie 11 Reihe 4.1). [a] = The areas of study shown here do not represent the areas of study of the Federal Statistical Office's higher education statistics, but rather an aggregation variant of subjects that was used in the reports of "The Student Survey in Germany" (2021). It was not used for weighting and only serves to visualise the distribution of weighted and unweighted data.

7.2   Weighting procedure and notes for application

[Weighting procedure] For the data of the study "The Student Survey in Germany" (2021), the following two weights were created in a multi-stage procedure using design and redressement weighting. They are both provided in the SUF and the CUF (see Table 10):

Table 10: Weights provided for the study "The Student Survey in Germany" (2021)

Variable name	Description
wgt_design	Design weight
wgt_cal_n	Redressement weight

[Design weighting] As described in chapter 3 students at some higher education institutions are overrepresented due to the sampling design. Consequently, these individuals have a higher probability of being selected for the sample than other students. Therefore, a design weight was first created to compensate for the different inclusion probabilities of students. This correction can be made by using the inverse probability of selection (π_i) (*Horvitz-Thompson estimator)*. The design weight wgt_d_i is thus:

$$wgt\_d_i = \frac{1}{\pi_i}$$

Elements that are more likely than others to be included in the sample, such as students whose higher education institutions have conducted a full survey, are thus given a lower weighting and vice versa.

[Calibrating the design weighting] Since no information on non-participants was available at the individual level, the next step was to create a redressement weighting based on the "Iterative Proportional Fitting Approach" (Kolenikov, 2014). For this purpose, the design weight was calibrated using the marginal distributions of demographic characteristics of the student body (age, gender, international students) known from official statistics as well as the variables subject group,[17] federal state and type of higher education institution for the summer semester 2021.

[Notes on using the weightings] The weighting created can be used in *Stata* using *ado-specific* options.[18] The weighting "*wgt_design*" only takes into account the unequal inclusion probabilities of the students. The calibrated design weight "*wgt_cal_n*" also considers the marginal distribution of the weighting characteristics in comparison with the distribution shown in the official (higher education) statistics. [19]

It is important to note that weighs only represent meaningful correction variables if the analysis model used contains the variables used for weighting (see Table 9) or is related to them. For this reason, weightings must always be used with a focus on the question to be analysed. If necessary, users are encouraged to create their own analysis-specific weightings.

---

[17] The subject groups used were not the subject groups of the classification by the Federal Statistical Office, but a project-specific variant that is strongly based on the breakdown of official statistics.

[18] Detailed explanations can be found in the Stata manual when the help weights command is entered.

[19] Due to missing information for the redressment, no calibrated weighting *"wgt_cal_n"* could be calculated for 8,027 survey participants.

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 data[20] collected by the DZHW in voluntary surveys.[21] 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 and organisational, technical, and statistical measures[22] . The combinations result in four potential access ways, which are summarised in Table 11.[23]

Table 11: Access ways in the FDZ-DZHW

	Scientific Use File (SUF): On-Site	Scientific Use File (SUF): Remote-Desktop	Scientific Use File (SUF): Download	Campus Use File (CUF): Download
Legal-organisational measures	high	moderate	moderate	low
Technical measures	high	moderate	low	very low
Statistical measures	low	moderate	high	very high

The more data access is legally, organisationally, and technically controlled and restricted, the lower the risk of de-anonymisation of the data. Following, the data must be anonymised less by means of statistical measures, i.e. have its information reduced, i.e. the greater its analysis potential remains.

All access ways are available for the data from "The Student Survey in Germany" (2021). The specific statistical anonymisation measures carried out are explained in more detail below.

---

[20] "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).

[21] 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 (Section 2 (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.

[22] The FDZ-DZHW's data protection concept is based on the portfolio approach of Lane, Heus and Mulcahy (2008), which is already used by the Leibniz Institute for Educational Trajectories (LIfBi) (cf. Koberg, 2016) and the FDZ of the German Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner, Müller, Schmucker & Roß, 2012).

[23] For more detailed information, see https://www.fdz.dzhw.eu/de/datennutzung.

[Statistical anonymisation measures] As part of the statistical anonymisation measures, all information was first checked to see whether it could be used to directly identify individuals. These *direct identifiers*, such as names, addresses and email addresses, were already recorded in a separate data set during the fieldwork phase and are therefore not included in the data per se. The original identification number was removed and replaced by 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 virtually harmless questions that could lead to identification." (Ebel & Meyermann, 2015).

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 association of the data, these key characteristics - depending on the access way – are released in an aggregated form or withheld completely. This applies both to information about the interviewee themselves and to information about third parties that would allow easier de-anonymization via contextualisation. When weighing up the quasi-identifiers in "The Student Survey in Germany" (2021), it must be considered that the combination of certain key characteristics may be unique for a respondent in the data set, but not in the general population. This results from the participation or non-participation of higher education institutions, the drawing of students to be invited from the sample population, the loss of potential respondents due to non-participation and the subsequent exclusion of persons during data preparation and response checks (see Chapters 5 and 6).

Finally, the data was checked for sensitive information, e.g. on health, sexual orientation, and political attitudes. Although this is not necessarily suitable for the re-identification of individuals or institutions, the information can be useful in the event of de-anonymisation (cf. Koberg, 2016) and is therefore particularly worthy of protection (Art. 9 GDPR, Recital 51 GDPR). The sensitive information is - depending on the access way - not released or only released in aggregated form.

An overview of which variables are contained in the data provided and how they can be accessed can be found in the respective dataset report of each data set within the data package.

The highest degree of information reduction was carried out for the "Campus Use File (CUF): Download" access way. In contrast to other data packages at the FDZ-DZHW, the CUF was not limited to a randomly selected subsample of the data to take into account the complex modularised structure of the survey and to enable sufficient analysis options for teaching and training purposes.

9 Literature

AAPOR (2016). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.
https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf
(Last retrieval: 26.03.2024)

AAPOR (2023). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. American Association for Public Opinion Research. 10th Edition. AAPOR.
https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf
(Last retrieval: 26.03.2024)

Augspurg, K. & Hinz, T. (2015). Factorial survey experiments. Sage

Becker, K., Baillet, F. & Weber, A. (2019). 21. Sozialerhebung. Daten- und Methodenbericht zu der Erhebung der wirtschaftlichen und sozialen Lage der Studierenden 2016. FDZ-DZHW. https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy21$/attachments/ssy21_MethodReport_de.pdf (Last retrieval: 32.01.2024)

Beuße, M., Kroher, M., Becker, K., Ehrhardt, M., Isleib, S., Koopmann, J., Steinkühler, J., Völk, D., Buchholz, S., Meyer, J., Multrus, F., Hinz, T., Marczuk, A., & Strauß, S. (2022). Die Studierendenbefragung in Deutschland: Eine neue, integrierte Datenbasis für Forschung, Bildungs- und Hochschulpolitik. (DZHW Brief 06|2022). DZHW.
https://doi.org/10.34878/2022.06.dzhw_brief

Cuppen, J., Muja, A., Hauschildt, K., Buck, D. & Daniel, A. (2023). Eurostudent VII. Micro Data Data and methods report SUF version 3.0. DZHW.
https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-es7$-3.1.0/attachments/es7_methodreport_en_3_0.pdf
(Last retrieval: 05.02.2024)

Couper, M. P. (1998): Measuring Survey Quality in a CASIC Environment. In American Statistical Association. *Proceedings of the Survey Research Methods Section* (pp. 41-49). American Statistical Association.

EU General Data Protection Regulation (GDPR)
https://gdpr-info.eu/ (Last retrieval: 15.02.2024)

DZHW (Eds.) (2018). Social and economic conditions of student life in Europe: Synopsis of Indicators. EUROSTUDENT VI 2016-2018. W. Bertelsmann Verlag.
https://www.eurostudent.eu/download_files/documents/EUROSTUDENT_VI_Synopsis_of_Indicators.pdf/
(Last retrieval: 22.01.2024)

Destatis (2020). Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik. WS 2020/2021 und SS 2021. Statistisches Bundesamt.
https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studenten_2_Pr%C3%BCfungen/Schluesselverzeichnis_WS_2020.pdf
Last retrieval (26.03.2024)

Ebel, T., Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. forschungs-daten bildung informiert 3. Forschungsdatenzentrum (FDZ) Bildung am DIPF

Felderer, B., Birg, A. & Kreuter, F. (2019). Paradaten. In Baur, N., Blasius, J. (Eds.). *Handbuch Methoden der empirischen Sozialforschung* (pp. 413-422). Springer VS. https://doi.org/10.1007/978-3-658-21308-4_29

Gerdes, F., & Völk, D. (2022). „Landesweite Studierendenbefragung an den Hochschulen in Nordrhein-Westfalen" im Kontext der „Studierendenbefragung in Deutschland". Ergebnisbericht. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (nicht zur Veröffentlichung vorgesehen).

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gesetz über die Statistik für das Hochschulwesen sowie für die Berufsakademien (Hochschulstatiskge-setz- HStatG). https://www.gesetze-im-internet.de/hstatg_1990/index.html (Last retrieval: 26.03.2024)

Göritz, A. S., & Crutzen, R. (2012). Reminders in Web-Based Data Collection: Increasing Response at the Price of Retention?. *American Journal of Evaluation*, 33(2), 240-250. https://doi.org/10.1177/1098214011421956

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

Jungbauer-Gans, M. (2018). Forschungsperspektiven am DZHW. In M. Fuhrmann, J. Güdler, J. Kohler, P. Pohlenz & U. Schmidt (Eds.), *Handbuch Qualität in Studium, Lehre und Forschung* (B4.19). DUZ Verlags- und Medienhaus.

Kiesl, H. (2022). Gewichtung. In Baur, N., Blasius, J. (Eds.) *Handbuch Methoden der empirischen Sozial-forschung* (pp. 473-481). Springer VS. https://doi.org/10.1007/978-3-658-37985-8_30

KldB (2010). Klassifikation der Berufe. Bundesagentur für Arbeit. https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/Klassifikation-der-Berufe-Nav.html (Last retrieval: 07.02.2024)

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 Na-tional Educational Panel Study* (pp. 691–708). Springer VS. https://doi.org/10.1007/978-3-658-11994-2

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

Kreuter, F. (2013): Improving Surveys with Paradata. Analytic Uses of Process Information. Wiley

Kroher, M., Beuße, M., Isleib, S., Becker, K., Ehrhardt, M. C., Gerdes, F., Koopmann, J., Schommer, T., Schwabe, U., Steinkühler, J., Völk, D., Peter, F., & Buchholz, S. (2023). Die Studierendenbefragung in Deutschland: 22. Sozialerhebung: Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2021. Leibniz Universität Hannover. https://doi.org/10.15488/15772

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.

Lenzner, T., Neuert, C., & Otto, W. (2015). Kognitives Pretesting. GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidlines).
https://doi.org/10.15465/gesis-sg_010

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I. & Vehovar, V. (2008). Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79-104.

Meier, D. H., Thomsen, S. L., & Kroher, M. (2023). Die Bedeutung der Inflation für die wirtschaftliche Situation von Studierenden in Deutschland im Zeitraum 2021 bis 2024: Eine Abschätzung. (DZHW Brief 01|2023). DZHW.
https://doi.org/10.34878/2023.01.dzhw_brief

Meyer, J., Strauß, S., & Hinz, T. (2022). Die Studierendenbefragung in Deutschland: Fokusanalysen zu Diskriminierungserfahrungen an Hochschulen. (DZHW Brief 08|2022). DZHW.
https://doi.org/10.34878/2022.08.dzhw_brief

Middendorff, E. (2022). Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen, Methoden, Themen und projektbezogene Publikationen (Working Paper). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-ssypool$-3.0.0/attachments/SE_History_1-21_Sept22.pdf
(Last retrieval: 26.06.2024)

Middendorff, E., Apolinarski, B., Becker, K., Bornkessel, P., Brandt, T., Heißenberg, S., & Poskowsky, J. (2017). Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2016: 21. Sozialerhebung des Deutschen Studentenwerks. Bundesministerium für Bildung und Forschung.

Middendorff, E., Wallis, M. (2023). 10.-21-Sozialerhebung 1982-2016. Daten- und Methodenbericht zum gepoolten Datensatz der zwölf Studierendenbefragungen. DZHW.

Multrus, F. (2021). Daten- und Methodenbericht des Datenpakets „Konstanzer Studierendensurvey 1983- 2016". AG Hochschulforschung, Universität Konstanz.

Multrus, F.S., Majer, T., Bargel, M., Schmidt (2017). Studiensituation und studentische Orientierungen. 13. Studierendensurvey an Universitäten und Fachhochschulen. Bundesministerium für Bildung und Forschung.

Multrus, F., Marczuk, A., Hinz, T. & Strauß, S. (2023). Studiensituation unter Onlinebedingungen. Studieren während der Corona-Pandemie 2021. Bundesministerium für Bildung und Forschung.
https://www.soziologie.uni-konstanz.de/ag-hochschulforschung/publikationen/publikationsarten/hauptberichte/
(Last retrieval: 18.01.2024)

Multrus, F., Strauß, S., & Hinz, T. (2022). Die Studierendenbefragung in Deutschland: Fokusanalysen zu populistischen Tendenzen unter Studierenden. (DZHW Brief 07|2022). DZHW.
https://doi.org/10.34878/2022.07.dzhw_brief

Netz, N., & Völk, D. (2023). Studienbezogene Auslandsmobilität und soziale Ungleichheiten im Kontext der Coronapandemie. (DZHW Brief 03|2023). DZHW.
https://doi.org/10.34878/2023.03.dzhw_brief

Petzold, K., & Netz, N. (2022). Vignettenexperimente in der Hochschul- und Wissenschaftsforschung: Konstruktion, Potenziale und Fallstricke illustriert am Beispiel einer Professor*innenbefragung. In G. Brandt & S. de Vogel (Eds.), *Survey-Methoden in der Hochschulforschung* (pp. 153-199). Springer.
https://doi.org/10.1007/978-3-658-36921-7_7

Poskowsky, J., Heißenberg, S., Zaussinger, S., & Brenner, J. (2018). Beeinträchtigt Studieren – best2: Datenerhebung zur Situation Studierender mit Behinderung und chronischer Krankheit 2016/17. Deutsches Studentenwerk. https://www.studierendenwerke.de/fileadmin/api/files/beeintraechtigt_studieren_2016_barrierefrei.pdf
(Last retrieval: 16.01.2024)

Prüfer, P., & Rexroth, M. (2000). Zwei-Phasen-Pretesting. (ZUMA-Arbeitsbericht 2000/08). Zentrum für Umfragen, Methoden und Analysen (ZUMA).
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200864
(Last retrieval: 23.01.2024).

Roßmann, J. (2017). Satisficing in Befragungen: Theorie, Messung und Erklärung. Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Springer VS
https://doi.org/10.1007/978-3-658-16668-7

Safikhani, P., Avetisyan, H., Föste-Eggers, D., & Broneske, D. (2023). Automated occupation coding with hierarchical features: A data-centric approach to classification with pre-trained language models. *Discover Artificial Intelligence* 3, 2023(6).
https://doi.org/10.1007/s44163-023-00050-y

Schaeper, H. & Weiß, T. (2016). The Conceptualization, Development, and Validation of an Instrument for Measuring the Formal Learning Environment in Higher Education. In H.-P.Blossfeld, J. von Maurice, M. Bayer & J. Skopek, (Eds.), *Methodological Issues of Longitudinal Surveys. The Example of the National Educational Panel Stud*y (pp. 267–290). Springer VS.
https://doi.org/10.1007/978-3-658-11994-2

Statistisches Bundesamt (2022). Studierende an Hochschulen – Sommersemester 2021 (Fachserie 11 Reihe 4.1). Statistisches Bundesamt (Destatis).
https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/studierende-hochschulen-ss-2110410217314.pdf?__blob=publicationFile
(Last retrieval: 10.05.2023)

Stefani, A., Hinz, T., & Strauß, S. (2023). Die Studierendenbefragung in Deutschland: Fokusanalysen zur Attraktivität von Masterstudiengängen. (DZHW Brief 02|2023). DZHW.
https://doi.org/10.34878/2023.02.dzhw_brief

Steinkühler, J., Beuße, M., Kroher, M., Gerdes, F., Schwabe, U., Koopmann, J., Becker, B., Völk, D., Schommer, T., Ehrhardt, M.-C., Sören, I., & Buchholz, S. (2023). Die Studierendenbefragung in Deutschland: best3. Studieren mit einer gesundheitlichen Beeinträchtigung. Leibniz Universität Hannover
https://doi.org/10.15488/15773

Tinto, V. (1975). Dropout from Higher Education. A Theoretical Synthesis of Recent Research. Review of Educational Research, 45(1), 89–125.
https://doi.org/10.3102/00346543045001089

Unger, M., Wejwar, P., Zaussinger, S., & Laimer, A. (2012). Beeinträchtigt Studieren: Datenerhebung zur

Situation Studierender mit Behinderung und chronischer Krankheit 2011. Deutsches Studenten-werk.

Wagner-Schelewsky, P. & Hering, L. (2022). Online-Befragung. In Baur, N., Blasius, J. (Eds.), Handbuch Methoden der empirischen Sozialforschung (pp. 1051-1065), Springer VS. https://doi.org/10.1007/978-3-658-37985-8_70

Weeber, F., Hinz, T., Meyer, J., Multrus, F. (2024). Rule-based semi-automated coding procedure to classify fields of study in surveys among university students in Germany. *Hefte zur Bildungs- und Hochschulforschung* 99. Higher Education Research Group (AG Hochschulforschung), University of Konstanz.
https://doi.org/10.48787/kops/352-2-15qrl6ayxtl7u8

Weiß, B., Silber, H., Struminskaya, B., Durrant, G. (2022). Mobile Befragungen. In Baur, N., Blasius, J. (Eds.) *Handbuch Methoden der empirischen Sozialforschung* (pp. 1067-1080), Springer VS. https://doi.org/10.1007/978-3-658-37985-8_71

Weichbold, M. (2022). Pretests. In: Baur, N. & Blasius, J. (Eds.) Handbuch Methoden der empirischen Sozialforschung. Springer VS (pp.443- 451).
https://doi.org/10.1007/978-3-658-37985-8_28

Appendix

Secondarily used instruments

Table 12: Sources for the secondarily used survey instruments

Question number	Source
Module A	
A_13	Adapted from Institut für Arbeitsmarkt- und Berufsforschung (2018) Arbeitsmarkt und Soziale Sicherung - Welle 11 (2017). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung. (DOI:10.5164/IAB.FDZD.1806.de.v1) Seniorenfragebogen PP1500
A_14	Adapted from Institut für Arbeitsmarkt- und Berufsforschung (2018) Arbeitsmarkt und Soziale Sicherung - Welle 11 (2017). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung. (DOI:10.5164/IAB.FDZD.1806.de.v1) Seniorenfragebogen zur Panelstudie „Arbeitsmarkt und Soziale Sicherung". Fragenummer PP1300
A_15	Adapted from Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2017). 20. Sozialerhebung – (2012). Hannover: FDZ-DZHW. (https://doi.org/10.21249/DZHW:ssy20:1.0.0). Datenkuratierung: Daniel, A., Sarcletti, A. & Vietgen, S. Die wirtschaftliche und soziale Situation der Studierenden Fragebogen. Questoin number 41
Module B1	
B1_2	Adapted from Berger, A. (2011). Entwicklung und Validierung eines Inventars zur Erfassung positiver und negativer Attribute des Geschlechtsrollenselbstkonzepts. Universität Potsdam, pp. 58
B1_3a	Adapted from Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus. Zeitschrift für Sozialpsychologie, 30(4), 211–228. https://doi.org/10.1024//0044-3514.30.4.211 Adapted from Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. Journal of Personality and Social Psychology, 70(3), 491–512
B1_3b	Adapted from Walter J. G. (2018). Measures of gender role attitudes under revision: The example of the German General Social Survey. Social science research, 72, 170–182. https://doi.org/10.1016/j.ssresearch.2018.02.009 Adapted from GESIS - Leibniz-Institut für Sozialwissenschaften (2013). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2012. GESIS Datenarchiv, Köln. ZA4614 Datenfile Version 1.1.1,

	https://doi.org/10.4232/1.11753. ALLBUS 2012 Fragebogendokumentation, Question numbers F020A, F020B.
Module B2	
B2_3a	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studierende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 759, Question number t34005a. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D1_18	Adapted from Weber, A.,Daniel, A., Becker, K., Bornkessel, P. (2018). Proximale Prädiktoren objektiver wie subjektiver Studienerfolgsindikatoren, In: Bornkessel, P. (Ed.) Erfolg im Studium: Konzeptionen, Befunde und Desiderate. Bielefeld: wbv Media GmbH & Co. KG. https://doi.org/10.3278/6004654w
B2_11	Adapted from Institut für Arbeitsmarkt- und Berufsforschung (2018) Arbeitsmarkt und Soziale Sicherung - Welle 11 (2017). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung. (DOI:10.5164/IAB.FDZD.1806.de.v1) Personenfragebogen zur Panelstudie Arbeitsmarkt und Soziale Sicherung. Question number PA0800
Module C1	
C1_14	Adapted from an Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2019). 18. Sozialerhebung — (2006). Hannover: FDZ-DZHW. (https://doi.org/10.21249/DZHW:ssy18:1.0.0). Datenkuratierung: Middendorff, E. & Hoffstätter, U. Die wirtschaftliche und soziale Situation der Studierenden Fragebogen. Question number 13
Modul C2	
C2_2	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question number BE6
C2_2a	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question number BE6
C2_7	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question numbers SB3, SB4
C2_8	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question numbers SB3, SB4, SB7
C2_9	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. F Question numbers SB3, SB4, SB7, SB8
C2_10	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question number SB10
C2_12	Adapted from Poskowsky, J., Heißenberg, S., Zaussinger, S., Brenner, J. (2018). *beeinträchtigt studieren — best2, Fragebogen.* Hannover, Berlin, Wien: DZHW, DSW, IHS. Question number B1
Module D1	
D1_19	Dickhäuser, O., Schöne, C., Spinath, B., Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept: Konstruktion und Überprüfung eines

	neuen Instrumentes. Zeitschrift für differentielle und diagnostische Psychologie 23 (4), pp. 393 su- 405 https://doi.org/10.1024//0170-1789.23.4.393
D1_22	Adapted from Nerdinger, F. W., Blickle, G., Schaper, N., Solga, M. (2018). Arbeits- und Organisationspsychologie. Deutschland: Springer Berlin Heidelberg. DOI https://doi.org/10.1007/978-3-662-56666-4
D1_23	Adapted from Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. British Journal of Educational Psychology, 71(1), pp. 133–149
D1_24	Adapted from Boerner, Sabine & Seeber, Günther & Keller, Helmut & Beinborn, Peter. (2005). Lernstrategien und Lernerfolg im Studium. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie. 37. 17-26. 10.1026/0049-8637.37.1.17. Adapted from Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. Zeitschrift für Differentielle und Diagnostische Psychologie, 15(4), 185–200. Adapted from Pintrich, P.R., Smith, D., García, T., McKeachie, W. (1991). Motivated Strategies for Learning Questionnaire (MSLQ) Manual. https://*doi*.org/*10.13140/RG.2.1.2547.6968*.
D2_35	Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2014). Allgemeine Selbstwirksamkeit Kurzskala (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis35
D1_29	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2017 DZHW-Studienberechtigtenpanel 2008 - Ausbildungswege von Studienberechtigten – (2007-2013). Hannover: FDZ-DZHW. (https://doi.org/10.21249/DZHW:gsl2008:1.0.0). Datenkuratierung: Daniel, A., Hoffstätter, U., Huß, B. & Scheller, P. Schule. Und dann? Erste Befragung von Schülerinnen und Schülern des Abschlussjahrgangs 2007/2008 Fragebogen. Question number 16
Module D2	
D2_2	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question numbers 53, 56
D2_3	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studierende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2047ff., Fragenummer t243402, t243401, t243403, t246421, https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved: 20.03.2024). Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question numbers17, 18, 24
D2_6a	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studierende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2392ff., Question numbers t245414, t245411, t245413

	https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsda-ten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D2_7	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studie-rende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2399ff., Question number t245421, t245424, t245422, t245423, t246404, t243411. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D2_8	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studie-rende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2405f., Question numbers t246404, t246412. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsda-ten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D2_9	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studie-rende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2387ff., Question numbers t246401, t246413, t246411. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsda-ten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D2_11	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studie-rende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2407ff., Question numbers t244403, t244411, t244402, t244413, t244401, t244412. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
D2_12	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studie-rende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 2396ff., Question numbers t244421, t244422, t244423. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsda-ten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
Module D3	
D3_9	Adapted from Engstler, H. Vogel, C., Böger, A. Franke, J., Klaus, D., Mahne, K., Simonson, J., Spuling, S., Wettstein, M., Wolff, J., Tesch-Römer, C. (2018) Deutscher Alterssurvey (DEAS): Instrumente der DEAS-Erhebung 2017, Ber-lin: Deutsches Zentrum für Altersfragen. DOI 10.5156/DEAS.2017.D.001, Question number 808
D3_23	Adapted from Ware, J., Jr., & Sherbourne, C.D. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection, in Medi-cal Care, June 1992, Vol. 30, No. 6, p.. 473-483
D1_27	Adapted from Thönnissen, C., Wilhelm, B., Alt, P., Greischel, H., Walper, S. (2019): pairfam Scales and Instruments Manual, Waves 1 to 10, Release 10.0. pp. 172 ff. GESIS Data Archive, Cologne. ZA5678 Data file Version 14.0.0, https://doi.org/10.4232/pairfam.5678.14.0.0 Adapted from Fliege, H., Rose, M., Arck, P., Levenstein, S. & Klapp, B. F. (2009). PSQ. Perceived Stress Questionnaire [Verfahrensdokumentation, PSQ20-Skalenberechnung, PSQ20-Fragebogen Englisch, Deutsch, Deutsch (letzte 2 Jahre), PSQ30-Skalenberechnung, PSQ30-Fragebogen Englisch, Französisch, Deutsch, Italienisch und Spanisch]. In Leibniz-Institut für Psycho-logie (ZPID) (Ed.), Open Test Archive. Trier: ZPID. https://doi.org/10.23668/psycharchives.12937

	Adapted from Levenstein, S., Prantera, C.,Varvo, V., Scribano, M., Berto, E., Luzi, C., Andreoli, A. (1993). Development of the Perceived Stress Questionnaire: A new tool for Psychosomatic Research. Journal of psychosomatic research. 37. pp. 19-32. https://doi.org/10.1016/0022-3999(93)90120-5.
D3_25	Adapted from Chmitorz, A., Wenzel, M., Stieglitz, R., Kunzler, A., Bagusat, C., Helmreich I, Gerlicher, A., Kampa, M., Kubiak, T., Kalisch, R., Lieb, K., Tüscher, O. (2018) Population-based validation of a German version of the Brief Resilience Scale. PLoS ONE 13(2): e0192761. https://doi.org/10.1371/journal.pone.0192761
Module E2	
E2M_2	Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey. Question number 66
E2M_4	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 67
E2S_1	Adapted from FDZ-LIfBi (2019). Codebook NEPS Startkohorte 5 — Studierende Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 12.0.0, pp. 194, Question number t320410. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/12-0-0/SC5_12-0-0_Codebook_de.pdf (last retrieved 20.03.2024).
E2S_2	Adapted from Holland, J.L. (1997) Making Vocational Choices: A Theory of Vocational Personalities and Work Environments. 3rd Edition, Psychological Assessment Resources, Odessa.
E2S_3	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 8
E2S_6	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 9
E2S_7	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question numbers59, 60
E2S_8	Adapted from Jänsch, V. K., & Bosse, E. (2018). Messinstrument für die Wahrnehmung von Studienanforderungen (MWS). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis263 Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-

	2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 60
Module E3	
E3_1	Adapted from GESIS Leibniz-Institut für Sozialwissenschaften (2019) ALLBUS Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (2018) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.13250 ALLBUS 2018 Fragebogendokumentation Material zu den Datensätzen der Studiennummern ZA5270 und ZA5271. Question number F028
E3_2	Adapted from GESIS Leibniz-Institut für Sozialwissenschaften (2019) ALLBUS Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (2018) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.13250 ALLBUS 2018 Fragebogendokumentation Material zu den Datensätzen der Studiennummern ZA5270 und ZA5271. Question number F023 Adapted from Presse- und Informationsamt der Bundesregierung, Berlin (2016). Vertrauen der Bevölkerung in die Politik (2016) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.12710 Vertrauen der Bevölkerung in die Politik. Endgültiger Fragebogen. Question number q7
E3_3	Adapted from GESIS Leibniz-Institut für Sozialwissenschaften (2019) ALLBUS Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (2018) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.13250 ALLBUS 2018 Fragebogendokumentation Material zu den Datensätzen der Studiennummern ZA5270 und ZA5271. Question number F033 Adapted from Vehrkamp, R., Merkel, W. (2019) Populismusbarometer 2018. Populistische Einstellungen bei Wählern und Nichtwählern in Deutschland 2018. Bertelsmann Stiftung, Gütersloh, p. 26 DOI 10.11586/2018059 Adapted from Mannheimer Zentrum für Europäische Sozialforschung (2015). Demokratie Audit Mannheim (2015). Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.12205 Mannheimer Demokratie Audit Fragebogen. Question number 31
E3_4	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 77
E3_6	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 64
E3_7	Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H., Lang, D., Ramm, M., Wuttke, K., Wiehn, E. R., Sandberger, J., Gawatz, R., Lensing, G., Framhein-Peisert, G., Peisert, H., Lenske, W., Behringer, F. & Gleich, M. (2021). Konstanzer Studierendensurvey 1983 - 2016. (1982-2017). Hannover: FDZ-DZHW. Datenkuratierung: Daniel, A. & Buck, D. https://doi.org/10.21249/DZHW:stsu8316:1.0.0, Codebuch, page 48
Module E4	

E4_1	Adapted from Georg, W., Multrus, F., Bargel, T., Majer, S., Schmidt, M., Simeaner, H. & Lang, D. (2022). Konstanzer Studierendensurvey 2016 (2015-2017). Hannover: FDZ-DZHW. https://doi.org/10.21249/DZHW:stsu16:1.0.0. Datenkuration: Daniel, A. Fragebogen Forschungsprojekt Studiensituation 13. Studierendensurvey, Question number 73
E4_3	Adapted from GESIS Leibniz-Institut für Sozialwissenschaften (2019) ALLBUS Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (2018) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.13250 ALLBUS 2018 Fragebogendokumentation Material zu den Datensätzen der Studiennummern ZA5270 und ZA5271. Question number F025
	Adapted from GESIS Leibniz-Institut für Sozialwissenschaften (2018) ALLBUS Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (2014) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.13141 ALLBUS 2014 Fragebogendokumentation Material zu den Datensätzen der Studiennummern ZA5240 und ZA5241. Question number F056
E4_4	Adapted from Sandberger, J. U. (1997). Aufstiegsmobilität. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis213
	Adapted from SSP Research Group (2017). International Social Survey Programme: Social Inequality IV – ISSP 2009 (2008-2012) Köln: GESIS Datenarchiv für Sozialwissenschaften. https://doi.org/10.4232/1.12777 nternational Social Survey Programme 2009 Social Inequality IV Final questionnaire. Question number Q1
E4_9	Adapted from Müller, U. & Kellmer, Ariana (2011) Diskriminierungserfahrungen von Studierenden Ergebnisse der großen UDE-Studierendenbefragung. Universität Duisburg-Essen
Module F1	
F1_1	Adapted from Petzold, K., Peter, T. The social norm to study abroad: determinants and effects. High Educ 69, 885–900 (2015). https://doi.org/10.1007/s10734-014-9811-4
F1_9	Adapted from Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2019). 21. Sozialerhebung – (2016). Hannover: FDZ-DZHW. (https://doi.org/10.21249/DZHW:ssy21:2.0.0). Datenkuratierung: Baillet, F. & Weber, A. 21. Sozialerhebung Fragebogen-Screenshots zu der Erhebung der wirtschaftlichen und sozialen Lage der Studierenden 2016: Bildungsausländer(innen). Question number B51
Module F2	
F2_8	Adapted from Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2019). 21. Sozialerhebung – (2016). Hannover: FDZ-DZHW. (https://doi.org/10.21249/DZHW:ssy21:2.0.0). Datenkuratierung: Baillet, F. & Weber, A. 21. Sozialerhebung Fragebogen-Screenshots zu der Erhebung der wirtschaftlichen und sozialen Lage der Studierenden 2016: Bildungsausländer(innen). Question number B51
Module F3	
F3_6	Adapted from Petzold, K., Peter, T. The social norm to study abroad: determinants and effects. High Educ 69, 885–900 (2015). https://doi.org/10.1007/s10734-014-9811-4
Other	

N_16   Otte, G. (2019). Weiterentwicklung der Lebensführungstypologie, Version 2019. Johannes Gutenberg-Universität Mainz

Overviews on generated variables

Description of the variables 1: Variable name 2: Variable label	Description of the generation
Basic programme	
migraerf "Immigrant background and international students"	*Formed using two auxiliary variables (migra and migra_a) that are not available in the data set.* *migra was formed as follows:* gen migra =. replace migra = 1 if vsbdeba==1 & dgebort == 1 replace migra = 3 if vsbdeba==1 & deltgebv == 2 replace migra = 3 if vsbdeba==1 & deltgebm == 2 replace migra = 2 if vsbdeba==1 & dgebort == 2 *migra_a was formed as follows:* gen migra_a = migra recode migra_a (3=2) replace migra_a =.a if vsbdeba==.a replace migra_a =.a if dgebort==.a & deltgebv==.a & deltgebm ==.a replace migra_a =.a if vsbdeba==1 & dgebort==.a & deltgebv!=2 & deltgebm !=2 *migraerf was formed as follows:* migraerf =. replace migraerf = 1 if migra_a == 1 replace migraerf = 2 if migra_a == 2 replace migraerf = 3 if intstud == 1
dnatsta_p "foreign nationality: country"	gen dnatsta_p = dnatsta replace dnatsta_p = dnatstao_g1 if dnatsta_p ==.a replace dnatsta_p = dnatstao_g2 if dnatsta_p ==.a
beeintrart "Type of disability/impairment" *Final report variable*	gen beeintrart = beeintr_art recode beeintrart (4=8) (5=4) (6=5) (7=6) (8=7) (10=8)
hs_p_g1 "Higher education institution current course of study: University"	*For the variable hs_p, a comparison was made between the higher education institution through which the students were invited to the survey and the higher education institution that the students indicated in the questionnaire.*
blandhs_preload_p "University current studies: state"	*Aggregation of the university location to federal states according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
FG1_STB_p "First subject area (subject group – Federal Statistical Office)"	*The subject information from question A_23 was merged into this variable. The subject groups are based on the variable sfach1 - coded according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*

SB1_STB_p "First subject area (study area – Federal Statistical Office)"	*The areas of study are based on the variable sfach1 - coded according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
FG2_STB_p "Second subject area (subject group – Federal Statistical Office)"	*The subject information from question A_23 was merged into this variable. The subject groups are based on the variable sfach1 - coded according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
SB2_STB_p "Second subject area (study area – Federal Statistical Office)"	*The areas of study are based on the variable sfach2 - coded according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
studfach "First subject area (subject group of the Student Survey in Germany)"	*The starting point for the variable studfach is the variable sfach1 - coded as subjects according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021. The individual subject codes were then broken down into 18 subject groups. From these 18 subject groups, the variable studfach was generated with 15 subject groups. For this purpose, the subject group "3. Sport Science" was included in the group "12. Health Science" and the subject group "4. Psychology" in "6. Social/Political/Regional and Administrative Science". The subject group "16. Electrical Engineering and Information Technology" was included in "15. Mechanical Engineering/Process Engineering (incl. Transport Engineering and Nautical Science)".*
sabsan_a "Degree pursued: type"	gen sabsan_a =. replace sabsan_a = 1 if sabsan==1 & inlist(sabslaja, 2, 3, 4, 5, 6) replace sabsan_a = 2 if sabsan==1 & sabslaja ==1 replace sabsan_a = 3 if sabsan==2 & inlist(sabslaja, 2, 3, 4, 5, 6) replace sabsan_a = 4 if sabsan==2 & sabslaja ==1 replace sabsan_a = 5 if sabsan==3 & inlist(sabslaja, 2, 3, 4, 5, 6) replace sabsan_a = 6 if sabsan==3 & sabslaja ==1 replace sabsan_a = 7 if inlist(sabsan, 4, 5, 7)
hsart_preload_p "University current studies: type of university"	*Aggregation of the university location into three types of higher education institution according to the key index of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
hs_groesse "University current studies: size university"	*Aggregation of university locations into three university sizes.*
akad_sid "Educational background: parents (non-academic vs. academic)"	gen akad_sid=. replace akad_sid=0 if (inlist(deltberuv,4,3,2,1) | inlist(deltberum,4,3,2,1)) replace akad_sid=1 if (inlist(deltberuv,7,6,5) | inlist(deltberum,7,6,5))
hszug "Type of university entrance qualification: higher educ. entrance qualification"	gen hszug = vsbart if inlist(vsbart,1,2,3) replace hszug = 4 if vsbtypone_a == 3
vsbtypone "Type of HZB: Access route"	*The single and multiple responses from the variables vsbtype1 - vsbtype15 were combined in this variable and supplemented by the open responses from vsbtype9o and vsbtype15o.*
vsbtypone_a "Type of university entrance qualification: agreggated access ways"	gen vsbtypone_a = vsbtypone replace vsbtypone_a=1 if inlist(vsbtypone, 1, 2, 3 ,5, 6, 7, 8, 9) replace vsbtypone_a=2 if inlist(vsbtypone, 4) replace vsbtypone_a=3 if inlist(vsbtypone, 10, 11) replace vsbtypone_a=4 if inlist(vsbtypone, 16,17,18,19)
berqual	*Was formed using two auxiliary variables (ausbild, fortbild) that are not available in the data set and summarizes whether*

"Professional qualification previous to current studies"	*vocational training and/or further training was completed before the course of study.*
	ausbild was formed as follows: gen ausbild = vsbtyp11_p replace ausbild = 1 if vausbja1_p == 1
	fortbild was formed as follows: gen fortbild = vsbtyp10_p replace fortbild = 1 if vausbja2_p == 1
	berqual was formed as follows: gen berqual = ausbild replace berqual = 1 if fortbild ==1
	replace ausbild=0 if vausbnein ==1 replace fortbild=0 if vausbnein ==1 replace berqual=0 if vausbnein ==1
Städte_Kat "University current studies: size location"	gen Städte_Kat=0 replace Städte_Kat=1 if Einwohnerzahl <100000 replace Städte_Kat=2 if Einwohnerzahl >=100000 & Einwohnerzahl <500000 replace Städte_Kat=3 if Einwohnerzahl >=500000 & Einwohnerzahl <1000000 replace Städte_Kat=4 if Einwohnerzahl >=1000000 & Einwohnerzahl < 10000000
B-Modules	
sexorient_a "Sexual orientation (aggregated)"	*The characteristics "gay" and "lesbian" were combined as "homosexual" and "other orientation" and "unclear" were combined as "other orientation/unclear"* gen sexorient_a = sexorient recode sexorient_a (2 3 = 2) recode sexorient_a (4 = 3) recode sexorient_a (5 6 = 4)
C-Modules	
stud_erschw "Disability that affects studies yes/no"	*Formed using two auxiliary variables (beeintr1 and studbeeintr1) that are not available in the data set.*
	beeintr1 was formed as follows: gen beeintr1 = 0 replace beeintr1 = 1 if gartmob_p==1 | gartseh_p==1 | gartohr_p==1 | gartspr_p==1 | gartpsy_p==1 | gartsom_p==1 | garttls_p==1 | gartson_p==1 | gartka==1 | gartneuro_g==1 | gartass_g==1 | gartads_g==1 | gartmig_g==1 | gartcan_g==1 tab beeintr1
	studbeeintr1 was formed as follows: gen studbeeintr1 =. replace studbeeintr1 = 0 if beeintr1==1 & beschwer_p==1 replace studbeeintr1 = 1 if beeintr1==1 & beschwer_p==2 tab studbeeintr1
	stud_erschw was formed as follows: gen stud_erschw = 0 replace stud_erschw = 1 if studbeeintr1==1 tab stud_erschw
grad_se	gen grad_se =.

"Extent of negative impact of disability/-ies on studies"	replace grad_se = 1 if gbegrmob_g==1 | gbegrseh_g==1 | gbe-grohr_g==1 | gbegrspr_g==1 | gbegrpsy_g==1 | gbe-grchron_g==1 | gbegrtls_g==1 | gbegrson_g==1 | gbegrka_g == 1
	replace grad_se = 2 if gbegrmob_g==2 | gbegrseh_g==2 | gbe-grohr_g==2 | gbegrspr_g==2 | gbegrpsy_g==2 | gbe-grchron_g==2 | gbegrtls_g==2 | gbegrson_g==2 | gbegrka_g == 2
	replace grad_se = 3 if gbegrmob_g==3 | gbegrseh_g==3 | gbe-grohr_g==3 | gbegrspr_g==3 | gbegrpsy_g==3 | gbe-grchron_g==3 | gbegrtls_g==3 | gbegrson_g==3 | gbegrka_g == 3
	replace grad_se = 4 if gbegrmob_g==4 | gbegrseh_g==4 | gbe-grohr_g==4 | gbegrspr_g==4 | gbegrpsy_g==4 | gbe-grchron_g==4 | gbegrtls_g==4 | gbegrson_g==4 | gbegrka_g == 4
	replace grad_se = 5 if gbegrmob_g==5 | gbegrseh_g==5 | gbe-grohr_g==5 | gbegrspr_g==5 | gbegrpsy_g==5 | gbe-grchron_g==5 | gbegrtls_g==5 | gbegrson_g==5 | gbegrka_g == 5
beausw_g "Officially recognized disability/impairment-disabled identification card"	*Indicates whether the student has an officially recognized disability or a disabled identification car, adjusted for implausible information.*
	gen beausw_g =. replace beausw_g = 1 if beausw_p==1 replace beausw_g = 2 if beausw_p==2 replace beausw_g = 3 if beausw_p==3 | beauswo_p<=40 replace beausw_g = 4 if beauswo_p>=50 & beauswo_p<=100
	D-Modules
stfwein_g "Reason field of study: income and career opportunities"	*The information from questions D1_7 and E2S_3 was combined in this variable.*
	gen stfwein_g =. replace stfwein_g = 1 if stfwein_p == 1 | stfweinb_p == 1 replace stfwein_g = 2 if stfwein_p == 2 | stfweinb_p == 2 replace stfwein_g = 3 if stfwein_p == 3 | stfweinb_p == 3 replace stfwein_g = 4 if stfwein_p == 4 | stfweinb_p == 4 replace stfwein_g = 5 if stfwein_p == 5 | stfweinb_p == 5
stfwart_g "Reason field of study: excellent prospect for secure employment"	*The information from questions D1_7 and E2S_3 was combined in this variable.*
	gen stfwart_g =. replace stfwart_g = 1 if stfwart_p == 1 | stfwartb_p == 1 replace stfwart_g = 2 if stfwart_p == 2 | stfwartb_p == 2 replace stfwart_g = 3 if stfwart_p == 3 | stfwartb_p == 3 replace stfwart_g = 4 if stfwart_p == 4 | stfwartb_p == 4 replace stfwart_g = 5 if stfwart_p == 5 | stfwartb_p == 5
stfwalt_g "Reason field of study: no admittance in the subject area of choice "	*The information from questions D1_7 and E2S_3 was combined in this variable.*
	gen stfwalt_g =. replace stfwalt_g = 1 if stfwalt_p == 1 | stfwaltb == 1 replace stfwalt_g = 2 if stfwalt_p == 2 | stfwaltb == 2

	replace stfwalt_g = 3 if stfwalt_p == 3 | stfwaltb == 3
	replace stfwalt_g = 4 if stfwalt_p == 4 | stfwaltb == 4
	replace stfwalt_g = 5 if stfwalt_p == 5 | stfwaltb == 5
stfwzeit_g "Reason field of study: short course of study"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwzeit_g =. replace stfwzeit_g = 1 if stfwzeit_p == 1 | stfwzeitb_p == 1 replace stfwzeit_g = 2 if stfwzeit_p == 2 | stfwzeitb_p == 2 replace stfwzeit_g = 3 if stfwzeit_p == 3 | stfwzeitb_p == 3 replace stfwzeit_g = 4 if stfwzeit_p == 4 | stfwzeitb_p == 4 replace stfwzeit_g = 5 if stfwzeit_p == 5 | stfwzeitb_p == 5
stfwfam_g "Reason field of study: parents/family advised me to do so"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwfam_g =. replace stfwfam_g = 1 if stfwfam_p == 1 | stfwfamb_p == 1 replace stfwfam_g = 2 if stfwfam_p == 2 | stfwfamb_p == 2 replace stfwfam_g = 3 if stfwfam_p == 3 | stfwfamb_p == 3 replace stfwfam_g = 4 if stfwfam_p == 4 | stfwfamb_p == 4 replace stfwfam_g = 5 if stfwfam_p == 5 | stfwfamb_p == 5
stfwber_g "Reason field of study: determination to pursue a specific career"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwber_g =. replace stfwber_g = 1 if stfwber_p == 1 | stfwberb_p == 1 replace stfwber_g = 2 if stfwber_p == 2 | stfwberb_p == 2 replace stfwber_g = 3 if stfwber_p == 3 | stfwberb_p == 3 replace stfwber_g = 4 if stfwber_p == 4 | stfwberb_p == 4 replace stfwber_g = 5 if stfwber_p == 5 | stfwberb_p == 5
stfwbega_g "Reason field of study: personal talent "	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwbega_g =. replace stfwbega_g = 1 if stfwbega_p == 1 | stfwbegab_p == 1 replace stfwbega_g = 2 if stfwbega_p == 2 | stfwbegab_p == 2 replace stfwbega_g = 3 if stfwbega_p == 3 | stfwbegab_p == 3 replace stfwbega_g = 4 if stfwbega_p == 4 | stfwbegab_p == 4 replace stfwbega_g = 5 if stfwbega_p == 5 | stfwbegab_p == 5
stfwint_g "Reason field of study: special interest in the subject"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwint_g =. replace stfwint_g = 1 if stfwint_p == 1 | stfwintb_p == 1 replace stfwint_g = 2 if stfwint_p == 2 | stfwintb_p == 2 replace stfwint_g = 3 if stfwint_p == 3 | stfwintb_p == 3 replace stfwint_g = 4 if stfwint_p == 4 | stfwintb_p == 4 replace stfwint_g = 5 if stfwint_p == 5 | stfwintb_p == 5
stfwwiss_g "Reason field of study: excellent academic education"	The information from questions D1_7 and E2S_3 was combined in this variable. gen stfwwiss_g =. replace stfwwiss_g = 1 if stfwwiss_p == 1 | stfwwissb_p == 1 replace stfwwiss_g = 2 if stfwwiss_p == 2 | stfwwissb_p == 2 replace stfwwiss_g = 3 if stfwwiss_p == 3 | stfwwissb_p == 3 replace stfwwiss_g = 4 if stfwwiss_p == 4 | stfwwissb_p == 4 replace stfwwiss_g = 5 if stfwwiss_p == 5 | stfwwissb_p == 5

stfwverf_g "Reason field of study: compatibility of family w. later employment opportunity"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwverf_g =. replace stfwverf_g = 1 if stfwverf_p == 1 | stfwverfb_p == 1 replace stfwverf_g = 2 if stfwverf_p == 2 | stfwverfb_p == 2 replace stfwverf_g = 3 if stfwverf_p == 3 | stfwverfb_p == 3 replace stfwverf_g = 4 if stfwverf_p == 4 | stfwverfb_p == 4 replace stfwverf_g = 5 if stfwverf_p == 5 | stfwverfb_p == 5
stfwverb_g "Reason field of study: compatibility of disability w. later empl. opport."	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwverb_g =. replace stfwverb_g = 1 if stfwverb_p == 1 | stfwverbb == 1 replace stfwverb_g = 2 if stfwverb_p == 2 | stfwverbb == 2 replace stfwverb_g = 3 if stfwverb_p == 3 | stfwverbb == 3 replace stfwverb_g = 4 if stfwverb_p == 4 | stfwverbb == 4 replace stfwverb_g = 5 if stfwverb_p == 5 | stfwverbb == 5
stfwand_g "Reason field of study: something else"	*The information from questions D1_7 and E2S_3 was combined in this variable.* gen stfwand_g =. replace stfwand_g = 1 if stfwand_p == 1 | stfwandb == 1 replace stfwand_g = 2 if stfwand_p == 2 | stfwandb == 2 replace stfwand_g = 3 if stfwand_p == 3 | stfwandb == 3 replace stfwand_g = 4 if stfwand_p == 4 | stfwandb == 4 replace stfwand_g = 5 if stfwand_p == 5 | stfwandb == 5
hswatt_g "Reason university choice: attractiveness of the city/surroundings"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswatt_g =. replace hswatt_g = 1 if hswatt_p==1 | hswattb_p==1 replace hswatt_g = 2 if hswatt_p==2 | hswattb_p==2 replace hswatt_g = 3 if hswatt_p==3 | hswattb_p==3 replace hswatt_g = 4 if hswatt_p==4 | hswattb_p==4 replace hswatt_g = 5 if hswatt_p==5 | hswattb_p==5
hswbarr_g "Reason university choice: barrier-free access to university"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswbarr_g =. replace hswbarr_g = 1 if hswbarr_p==1 | hswbarrb==1 replace hswbarr_g = 2 if hswbarr_p==2 | hswbarrb==2 replace hswbarr_g = 3 if hswbarr_p==3 | hswbarrb==3 replace hswbarr_g = 4 if hswbarr_p==4 | hswbarrb==4 replace hswbarr_g = 5 if hswbarr_p==5 | hswbarrb==5
hswbed_g "Reason university choice: favourable living conditions"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswbed_g =. replace hswbed_g = 1 if hswbed_p==1 | hswbedb_p==1 replace hswbed_g = 2 if hswbed_p==2 | hswbedb_p==2 replace hswbed_g = 3 if hswbed_p==3 | hswbedb_p==3 replace hswbed_g = 4 if hswbed_p==4 | hswbedb_p==4 replace hswbed_g = 5 if hswbed_p==5 | hswbedb_p==5
hswfach_g "Reason university choice: chosen area of expertise"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswfach_g =.

	replace hswfach_g = 1 if hswfach_p==1 | hswfachb_p==1
	replace hswfach_g = 2 if hswfach_p==2 | hswfachb_p==2
	replace hswfach_g = 3 if hswfach_p==3 | hswfachb_p==3
	replace hswfach_g = 4 if hswfach_p==4 | hswfachb_p==4
	replace hswfach_g = 5 if hswfach_p==5 | hswfachb_p==5
hswint_g "Reason university choice: international out-look of university"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswint_g =. replace hswint_g = 1 if hswint_p==1 | hswintb_p==1 replace hswint_g = 2 if hswint_p==2 | hswintb_p==2 replace hswint_g = 3 if hswint_p==3 | hswintb_p==3 replace hswint_g = 4 if hswint_p==4 | hswintb_p==4 replace hswint_g = 5 if hswint_p==5 | hswintb_p==5
hswkont_g "Reason university choice: friends/family where university is located"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswkont_g =. replace hswkont_g = 1 if hswkont_p==1 | hswkontb_p==1 replace hswkont_g = 2 if hswkont_p==2 | hswkontb_p==2 replace hswkont_g = 3 if hswkont_p==3 | hswkontb_p==3 replace hswkont_g = 4 if hswkont_p==4 | hswkontb_p==4 replace hswkont_g = 5 if hswkont_p==5 | hswkontb_p==5
hswrank_g "Reason university choice: high ranking of my field"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswrank_g =. replace hswrank_g = 1 if hswrank_p==1 | hswrankb==1 replace hswrank_g = 2 if hswrank_p==2 | hswrankb==2 replace hswrank_g = 3 if hswrank_p==3 | hswrankb==3 replace hswrank_g = 4 if hswrank_p==4 | hswrankb==4 replace hswrank_g = 5 if hswrank_p==5 | hswrankb==5
hswsupp_g "Reason university choice: university-specific counselling/support services"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswsupp_g =. replace hswsupp_g = 1 if hswsupp_p==1 | hswsuppb_p==1 replace hswsupp_g = 2 if hswsupp_p==2 | hswsuppb_p==2 replace hswsupp_g = 3 if hswsupp_p==3 | hswsuppb_p==3 replace hswsupp_g = 4 if hswsupp_p==4 | hswsuppb_p==4 replace hswsupp_g = 5 if hswsupp_p==5 | hswsuppb_p==5
hswtrad_g "Reason university choice: tradition and repu-tation of university"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswzul_g =. replace hswzul_g = 1 if hswzul_p==1 | hswzulb==1 replace hswzul_g = 2 if hswzul_p==2 | hswzulb==2 replace hswzul_g = 3 if hswzul_p==3 | hswzulb==3 replace hswzul_g = 4 if hswzul_p==4 | hswzulb==4 replace hswzul_g = 5 if hswzul_p==5 | hswzulb==5
hswtz_g "Reason university choice: possibility to study part-time"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswtz_g =. replace hswtz_g = 1 if hswtz_p==1 | hswtzb==1 replace hswtz_g = 2 if hswtz_p==2 | hswtzb==2 replace hswtz_g = 3 if hswtz_p==3 | hswtzb==3 replace hswtz_g = 4 if hswtz_p==4 | hswtzb==4

	replace hswtz_g = 5 if hswtz_p==5 | hswtzb==5
hswzul_g "Reason university choice: not admitted to university of choice"	*The information from questions D1_8 and E2S_4 was combined in this variable.* gen hswzul_g =. replace hswzul_g = 1 if hswzul_p==1 | hswzulb==1 replace hswzul_g = 2 if hswzul_p==2 | hswzulb==2 replace hswzul_g = 3 if hswzul_p==3 | hswzulb==3 replace hswzul_g = 4 if hswzul_p==4 | hswzulb==4 replace hswzul_g = 5 if hswzul_p==5 | hswzulb==5
hswand_g "Reason university choice: something else"	*The information from questions D1_8 and E2S_4 was combined in this variable.* replace hswand_g = 1 if hswand==1 | hswandb==1 replace hswand_g = 2 if hswand==2 | hswandb==2 replace hswand_g = 3 if hswand==3 | hswandb==3 replace hswand_g = 4 if hswand==4 | hswandb==4 replace hswand_g = 5 if hswand==5 | hswandb==5
vlfach1 "First subject, progression/changes"	*The starting point for the variable vlfach1 is the variable fach01, coded as subjects according to the key list of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
vlfach2 "Second subject, progression/changes"	*The starting point for the variable vlfach2 is the variable fach02, coded as subjects according to the key list of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
vlfach3 "Third subject, progression/changes"	*The starting point for the variable vlfach3 is the variable fach03, coded as subjects according to the key list of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
vlfach4 "Fourth subject, progression/changes"	*The starting point for the variable vlfach4 is the variable fach04, coded as subjects according to the key list of the Federal Statistical Office's student and examination statistics for the winter semester 2020/21 and summer semester 2021.*
E-modules	
No variables from these modules present	
F-Modules	
No variables from these modules present	

Description of the variables 1: Variable name 2: Variable label	Description of the generation
Basic programme	
demofam_a "Marital status – 3-way category" *Variable not available in the data set.*	egen demofam_count = anycount(demofamsin demofampar demofamehe demofamaus demofamtod), val(1) gen demofam_a =. replace demofam_a = 4 if (demofam_count>0 & demofam_count<6) replace demofam_a = 2 if demofampar==1 & demofam_count==1 replace demofam_a = 1 if demofamehe==1 & demofam_count==1 replace demofam_a = 3 if demofamsin==1 & demofam_count==1
pfleg_heads "Care expenditure" *Variable not available in the data set.*	gen pfleg_heads = pflegang1 replace pfleg_heads = 1 if pflegang1 == 1 | pflegang2 == 1 replace pfleg_heads = 0 if pflegangno == 1
pflegangno12 "Caring tasks - single choice" *Variable not available in the data set*	gen pflegangno12 = pfleg_heads replace pflegangno12 = 1 if pflegang1 ==1 replace pflegangno12 = 2 if pflegang2 == 1 replace pflegangno12 = 3 if pflegang1==1 & pflegang2 ==1
semob, seseh, seohr,sespr, sepsy, sesom, setls, seson, seneuro, seass, awads,semig, secan "Auxiliary variable impairment type" *These variables were created as auxiliary variables and are not available in the data set.*	gen semob =. replace semob = 0 if gartmob_p!=1 & beschwer_p==2 replace semob = 1 if gartmob_p==1 & beschwer_p==2 gen seseh =. replace seseh = 0 if gartseh_p!=1 & beschwer_p==2 replace seseh = 1 if gartseh_p==1 & beschwer_p==2 gen seohr =. replace seohr = 0 if gartohr_p!=1 & beschwer_p==2 replace seohr = 1 if gartohr_p==1 & beschwer_p==2 gen sespr =. replace sespr = 0 if gartspr_p!=1 & beschwer_p==2 replace sespr = 1 if gartspr_p==1 & beschwer_p==2 gen sepsy =. replace sepsy = 0 if gartpsy_p!=1 & beschwer_p==2 replace sepsy = 1 if gartpsy_p==1 & beschwer_p==2 gen sesom =. replace sesom = 0 if gartsom_p!=1 & beschwer_p==2 replace sesom = 1 if gartsom_p==1 & beschwer_p==2 gen setls =. replace setls = 0 if garttls_p!=1 & beschwer_p==2 replace setls = 1 if garttls_p==1 & beschwer_p==2 gen seson =. replace seson = 0 if gartson_p!=1 & beschwer_p==2 replace seson = 1 if gartson_p==1 & beschwer_p==2 gen seneuro =. replace seneuro = 0 if gartneuro_g!=1 & beschwer_p==2 replace seneuro = 1 if gartneuro_g==1 & beschwer_p==2 gen seass =. replace seass = 0 if gartass_g!=1 & beschwer_p==2 replace seass = 1 if gartass_g==1 & beschwer_p==2

	gen seads =.
	replace seads = 0 if gartads_g!=1 & beschwer_p==2
	replace seads = 1 if gartads_g==1 & beschwer_p==2
	gen semig =.
	replace semig = 0 if gartmig_g!=1 & beschwer_p==2
	replace semig = 1 if gartmig_g==1 & beschwer_p==2
	gen secan =.
	replace secan = 0 if gartcan_g!=1 & beschwer_p==2
	replace secan = 1 if gartcan_g==1 & beschwer_p==2
se_count "Counting variable for multiple disabilities" *This variable was created as an auxiliary variable and is not available in the data set.*	egen se_count = anycount(semob seseh seohr sespr sepsy sesom setls seson seneuro seass seads semig secan), values(1)
semob_, seseh_, seohr_, sespr_, sepsy_, sesom_, setls_, seson_, seneuro_, seass_, semig_, secan_ "Auxiliary variables for single disabilities" *These variables were created as auxiliary variables and are not available in the data set.*	gen semob_ =.
	replace semob_ = 0 if semob==0
	replace semob_ = 1 if semob==1 & se_count==1
	replace semob_ = 0 if semob==1 & se_count>=2
	gen seseh_ =.
	replace seseh_ = 0 if seseh==0
	replace seseh_ = 1 if seseh==1 & se_count==1
	replace seseh_ = 0 if seseh==1 & se_count>=2
	gen seohr_ =.
	replace seohr_ = 0 if seohr==0
	replace seohr_ = 1 if seohr==1 & se_count==1
	replace seohr_ = 0 if seohr==1 & se_count>=2
	gen sespr_ =.
	replace sespr_ = 0 if sespr==0
	replace sespr_ = 1 if sespr==1 & se_count==1
	replace sespr_ = 0 if sespr==1 & se_count>=2
	gen sepsy_ =.
	replace sepsy_ = 0 if sepsy==0
	replace sepsy_ = 1 if sepsy==1 & se_count==1
	replace sepsy_ = 0 if sepsy==1 & se_count>=2
	gen sesom_ =.
	replace sesom_ = 0 if sesom==0
	replace sesom_ = 1 if sesom==1 & se_count==1
	replace sesom_ = 0 if sesom==1 & se_count>=2
	gen setls_ =.
	replace setls_ = 0 if setls==0
	replace setls_ = 1 if setls==1 & se_count==1
	replace setls_ = 0 if setls==1 & se_count>=2
	gen seson_ =.
	replace seson_ = 0 if seson==0
	replace seson_ = 1 if seson==1 & se_count==1
	replace seson_ = 0 if seson==1 & se_count>=2
	gen seneuro_ =.
	replace seneuro_ = 0 if seneuro==0
	replace seneuro_ = 1 if seneuro==1 & se_count==1
	replace seneuro_ = 0 if seneuro==1 & se_count>=2
	gen seass_ =.
	replace seass_ = 0 if seass==0
	replace seass_ = 1 if seass==1 & se_count==1
	replace seass_ = 0 if seass==1 & se_count>=2
	gen seads_ =.
	replace seads_ = 0 if seads==0
	replace seads_ = 1 if seads==1 & se_count==1
	replace seads_ = 0 if seads==1 & se_count>=2

	gen semig_ =. replace semig_ = 0 if semig==0 replace semig_ = 1 if semig==1 & se_count==1 replace semig_ = 0 if semig==1 & se_count>=2 gen secan_ =. replace secan_ = 0 if secan==0 replace secan_ = 1 if secan==1 & se_count==1 replace secan_ = 0 if secan==1 & se_count>=2
grad1_count "Counting variable degree of disability" *This variable was created as an auxiliary variable and is not available in the data set.*	egen grad1_count = anycount(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p), values(1) egen grad2_count = anycount(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p), values(2) egen grad3_count = anycount(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p), values(3) egen grad4_count = anycount(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p), values(4) egen grad5_count = anycount(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p), values(5)
grad_countmax "Counting variable strongest disability" *This variable was created as an auxiliary variable and is not available in the data set.*	egen grad_countmax = rmax(gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p)
semehr_ "Formation of multiple disability" *This variable was created as an auxiliary variable and is not available in the data set.*	gen semehr_ =. replace semehr_ = 1 if grad_countmax==5 & grad5_count>1 replace semehr_ = 1 if grad_countmax==4 & grad4_count>1 replace semehr_ = 1 if grad_countmax==3 & grad3_count>1 replace semehr_ = 1 if grad_countmax==3 & grad2_count>1 replace semehr_ = 1 if grad_countmax==1 & grad1_count>1 replace semehr_ = 0 if grad_countmax==5 & grad5_count==1 replace semehr_ = 0 if grad_countmax==4 & grad4_count==1 replace semehr_ = 0 if grad_countmax==3 & grad3_count==1 replace semehr_ = 0 if grad_countmax==2 & grad2_count==1 replace semehr_ = 0 if grad_countmax==1 & grad1_count==1
beart_einf "Type of disability (single disability only)" *This variable was created as an auxiliary variable and is not available in the data set.*	gen beart_einf =. replace beart_einf = 1 if semob_==1 replace beart_einf = 2 if seseh_==1 replace beart_einf = 3 if seohr_==1 replace beart_einf = 4 if sespr_==1 replace beart_einf = 5 if sepsy_==1 replace beart_einf = 6 if sesom_==1 replace beart_einf = 7 if setls_==1 replace beart_einf = 8 if seson_==1 replace beart_einf = 9 if seneuro_==1 replace beart_einf = 10 if seass_==1 replace beart_einf = 11 if seads_==1 replace beart_einf = 12 if semig_==1 replace beart_einf = 13 if secan_==1 gen beart_einf1 = beart_einf recode beart_einf1 (9/13=8)
semehr_1	gen semehr_1 = semehr_ replace semehr_1 = 2 if beart_einf1<=8

"Multiple disabilities based on the extent of the disability" *This variable was created as an auxiliary variable and is not available in the data set.*	
gbegrmob_p_h, gbegrseh_p_h, gbegrohr_p_h, gbegrspr_p_h, gbegrpsy_p_h, gbegrchron_p_h, gbegrtls_p_h, gbegrson_p_h "Auxiliary variables for the disability types" *These variables were created as auxiliary variables and are not available in the data set.*	foreach var of varlist gbegrmob_p gbegrseh_p gbegrohr_p gbegrspr_p gbegrpsy_p gbegrchron_p gbegrtls_p gbegrson_p { gen `var'_h = `var' replace `var'_h =-999 if mi(`var'_h)}
beeintr_art "Type of disability with multiple disabilities" *This variable was created as an auxiliary variable and is not available in the data set.*	gen beeintr_art = beart_einf1 replace beeintr_art=1 if semob==1 & semehr_1==0 & (gbegrmob_p_h > gbegrseh_p_h & gbegrmob_p_h > gbegrohr_p_h & gbegrmob_p_h > gbegrspr_p_h & gbegrmob_p_h > gbegrpsy_p_h & gbegrmob_p_h > gbegrchron_p_h & gbegrmob_p_h > gbegrtls_p_h & gbegrmob_p_h > gbegrson_p_h) replace beeintr_art=2 if seseh==1 & semehr_1==0 & (gbegrseh_p_h > gbegrmob_p_h & gbegrseh_p_h > gbegrohr_p_h & gbegrseh_p_h > gbegrspr_p_h & gbegrseh_p_h > gbegrpsy_p_h & gbegrseh_p_h > gbegrchron_p_h & gbegrseh_p_h > gbegrtls_p_h & gbegrseh_p_h > gbegrson_p_h) replace beeintr_art=3 if seohr==1 & semehr_1==0 & (gbegrohr_p_h > gbegrmob_p_h & gbegrohr_p_h > gbegrseh_p_h & gbegrohr_p_h > gbegrspr_p_h & gbegrohr_p_h > gbegrpsy_p_h & gbegrohr_p_h > gbegrchron_p_h & gbegrohr_p_h > gbegrtls_p_h & gbegrohr_p_h > gbegrson_p_h) replace beeintr_art=5 if sepsy==1 & semehr_1==0 & (gbegrpsy_p_h > gbegrmob_p_h & gbegrpsy_p_h > gbegrseh_p_h & gbegrpsy_p_h > gbegrohr_p_h & gbegrpsy_p_h > gbegrspr_p_h & gbegrpsy_p_h > gbegrchron_p_h & gbegrpsy_p_h > gbegrtls_p_h & gbegrpsy_p_h > gbegrson_p_h) replace beeintr_art=6 if sesom==1 & semehr_1==0 & (gbegrchron_p_h > gbegrmob_p_h & gbegrchron_p_h > gbegrseh_p_h & gbegrchron_p_h > gbegrohr_p_h & gbegrchron_p_h > gbegrspr_p_h & gbegrchron_p_h > gbegrpsy_p_h & gbegrchron_p_h > gbegrtls_p_h & gbegrchron_p_h > gbegrson_p_h) replace beeintr_art=7 if setls==1 & semehr_1==0 & (gbegrtls_p_h > gbegrmob_p_h & gbegrtls_p_h > gbegrseh_p_h & gbegrtls_p_h > gbegrohr_p_h & gbegrtls_p_h > gbegrspr_p_h & gbegrtls_p_h > gbegrpsy_p_h & gbegrtls_p_h > gbegrchron_p_h & gbegrtls_p_h > gbegrson_p_h) replace beeintr_art=8 if seson==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h & gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h) replace beeintr_art=8 if seneuro==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h & gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h) replace beeintr_art=8 if seass==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h

	& gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h) replace beeintr_art=8 if seads==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h & gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h) replace beeintr_art=8 if semig==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h & gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h) replace beeintr_art=8 if secan==1 & semehr_1==0 & (gbegrson_p_h > gbegrmob_p_h & gbegrson_p_h > gbegrseh_p_h & gbegrson_p_h > gbegrohr_p_h & gbegrson_p_h > gbegrspr_p_h & gbegrson_p_h > gbegrpsy_p_h & gbegrson_p_h > gbegrchron_p_h & gbegrson_p_h > gbegrtls_p_h)
	replace beeintr_art=10 if semehr_1==1
studfach_best "Areas of study for best report" *This variable was created as an auxiliary variable and is not available in the data set.*	*The variable studfach_best was created on the basis of studfach. The group 12 "Health Science (incl. Sport Science)" was included in group 11 "Medicine (Human and Dentistry)". The variable therefore contains 14 subject groups.*
sabsan_int "Type of degree international students" *Variable not available in the data set.*	gen sabsan_int = sabsan if intstud==1 recode sabsan_int (4 5 7 8 = 99) recode sabsan_int (6 = 4) replace sabsan_int = 5 if inlist(sabsabsba,2,3) recode sabsan_int (99 = 6)
vsbtypbin "Type of HZB (binary)" *This variable was created as an auxiliary variable and is not available in the data set.*	recode vsbtypone (1 2 3 4 5 6 7 8 16 18 19 = 1) (10 11 17 = 2) (12 13 14 = .a), into (vsbtypbin)
Einwohnerzahl "Population of university town" *This variable was created as an auxiliary variable and is not available in the data set.*	*Official number of inhabitants assigned to the university locations using the ID.*
B-Modules	
No variables from these modules present	
C-Modules	
grad_se1 "Extent of the study difficulty (3-way category)" *Variable not available in the data set.*	*Aggregates the values 1 and 2 or 4 and 5 of the variable grad_se.*
bzeitj_g "Time of occurrence of the disability" *Variable not available in the data set.*	*The time of onset of impairment was determined using a combination of age and higher education institution semester.* gen alter_imma = (demoage-(ssemhs_p/2)) if demoage>14 & demoage<=99 & ssemhs_p<. recode alter_imma (-25.5/14.5=.) gen bezeitj_g =. replace bezeitj_g = 1 if bezeit_p==1 replace bezeitj_g = 2 if alter_imma>bezeitj_p & alter_imma<. & bezeitj_p<.

	replace bezeitj_g = 3 if alter_imma<bezeitj_p & alter_imma<. & bezeitj_p<.
einallse20 "Total revenue" *Variable not available in the data set.*	*The variable is not available in the SUF and was created for the SiD reports. It includes the cash and non-cash income of students.* *The variable is made up of the variable feininsg and feininsg2.*
ges_insg "Total expenditure" *Variable not available in the data set.*	*The variable is not available in the SUF and was created for the SiD reports. It includes students' own expenditure as well as expenditure by third parties.* *The variable is made up of the variable fausinsg and feininsg2.*
erw_quote1 "Employment rate as in the 21st SE" *Variable not available in the data set.*	gen erw_quote1 =. replace erw_quote1 = 0 if eaktsens==1 replace erw_quote1 = 1 if (erw_quote==1 & sform_a !=2) | (erw_quote==1 & sform_a==2 & etat_h==1)
Focustyp "Focus type SE" *Variable not available in the data set.*	*The variable is not available in the SUF and was created for the SiD reports. It comprises a specific group of students. These are unmarried first-year students who are not living with their parents or other family members, who are studying full-time and have obtained their higher education entrance qualification in Germany.* *The following variables are required to generate the focus type variable: demofam_a, wohnel bzw. wohnfam, sabsja, sartzeit, sformpraes und vsbdeba*
E-modules	
No variables from these modules present	
F-Modules	
No variables from these modules present	

Table 15: Variables generated by the Research Data Centre (FDZ-DZHW) as part of the anonymization process - available in the dataset

Description of the variables 1: Variable name 2: Variable label	Description of the generation
Basic programme	
pid "Case number (randomized)"	*Randomized identification number based on the project's internal identification number to anonymize the respondents*
hs_cluster_id "University current studies: university (randomized)"	*Randomized identification number of higher education institutions based on the university codes from hs_p_g1 to form clusters of respondents from the same university.*
dnatsta_p_g2 "Foreign nationality: country (NEPS classification)"	*Aggregation of the countries from dnatsta_p or dnatsta_p_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
dnatsta_p_g3 "Country of birth: student (world regions)"	*Aggregation of countries from dnatsta_p or dnatsta_p_g1 on the basis of continents. North and South America merged, Europe additionally differentiated by EU/non-EU.*
dgebsta_p_g2 "Country of birth: student (NEPS classification)"	*Aggregation of the countries from dgebsta_p or dgebsta_p_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
dgebsta_p_g3 "Country of birth: student"	*Aggregation of countries from dgebsta_p or dgebsta_p_g1 on the basis of continents. North and South America merged, Europe additionally differentiated by EU/non-EU.*
demosex_p_g "Gender (male/female/missing)"	*Pooling of the information "diverse", "none of the categories mentioned" and missing information from demosex_p to a category "anonymised".*
demoage_g1 "Age (in years) (top-bottom coding)"	*Pooling of age data from demoage_p for survey participants younger than 19 and older than 30.*
demoage_g2 "Age (in years) (categorized)"	*Aggregation of the age data from demoage_p into five age groups. The top and bottom age groups correspond to the combined categories from demoage_g1.*
dsibanz_g "Number of siblings (top coding)"	*Aggregation of the number of siblings from dsibanz; grouping of more than three siblings in one category.*
dsibageo1_p_g1 "First sibling: age (top-bottom coding)"	*Pooling of age data from dsibageo1_g1 for siblings younger than 19 years or older than 30 years.*
dsibageo2_p_g1 "Second sibling: age (top-bottom coding)"	*Pooling of age data from dsibageo2_g1 for siblings younger than 19 years or older than 30 years.*
dsibageo3_p_g1 "Third sibling: age (top-bottom coding)"	*Pooling of age data from dsibageo3_g1 for siblings younger than 19 years or older than 30 years.*
ssemfs_p_g "Study period: quantity subject-related semesters (top coding)"	*Pooling of quantity of subject-related semesters from ssemfs_p for respondents with 25 or more subject-related semsters.*
ssemhs_p_g "Study period: quantity university semesters (top coding)"	*Pooling of quantity of university semesters from ssemhs_p for respondents with 25 or more university semsters.*
hs_p_g2 "University current studies: university (federal state)"	*Aggregation of the location data from hs_p_g1 to 16 German federal states (and foreign countries if applicable).*
hs_p_g3 "University current studies: university (new or old federal states)"	*Aggregation of the location data from hs_p_g1 to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs_p_g4	*Aggregation of the higher education institution data from hs_p_g1 into "University" (including teacher training colleges,*

"University current studies: university (type -> 2 categories)"	*theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
deltjobvo_g2 "Occupation father (KldB2010 - 3 digit)"	*Aggregation of the coded occupational data from deltjobvo_g1 (occupational categories according to KldB2010 - 5-digit) to occupational groups according to KldB2010 (3-digit).*
deltjobvo_g3 "Occupation father (KldB2010 - 2 digit)"	*Aggregation of the coded occupational data from deltjobvo_g1 (occupational categories according to KldB2010 - 5-digit) to main occupational groups according to KldB2010 (2-digit).*
deltjobmo_g2 "Occupation mother (KldB2010 - 3 digit)"	*Aggregation of the coded occupational data from deltjobmo_g1 (occupational categories according to KldB2010 - 5-digit) to occupational groups according to KldB2010 (3-digit).*
deltjobmo_g3 "Occupation mother (KldB2010 - 2 digit)"	*Aggregation of the coded occupational data from deltjobmo_g1 (occupational categories according to KldB2010 - 5-digit) to main occupational groups according to KldB2010 (2-digit).*
deltgstav_p_g2 "Country of birth: father (NEPS classification)"	*Aggregation of the countries from deltgstav_p_g1 on the basis of the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
deltgstav_p_g3 "Country of birth: father (world regions)"	*Aggregation of countries from deltgstv_p_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
deltgstam_p_g2 "Country of birth: mother (NEPS classification)"	*Aggregation of the countries from deltgstam_p_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
deltgstam_p_g3 " Country of birth: mother (world regions)"	*Aggregation of countries from deltgstam_p_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
B-Modules	
No variables from these modules present	
C-Modules	
dkinagej_p_g "Age: youngest child (age groups)"	*Aggregation of the age data from dkinagej_p into ten age groups, based on common age limits (nursery, school, young people, adolescents, etc.).*
dkinage1_p_g "Age: only child (age groups)"	*Aggregation of the age data from dkinage1_p into ten age groups, based on common age limits (nursery, school, young people, adolescents, etc.).*
dkinage2_p_g "Age: first child (age groups)"	*Aggregation of the age data from dkinage2_p into ten age groups, orientated to common age limits (nursery, school, young people, adolescents, etc.).*
dkinage3_p_g "Age: second child (age groups)"	*Aggregation of the age data from dkinage3_p into ten age groups, orientated to common age limits (nursery, school, young people, adolescents, etc.).*
dkinage4_p_g "Age: third child (age groups)"	*Aggregation of the age data from dkinage4_p into ten age groups, based on common age limits (nursery, school, young people, adolescents, etc.).*
dkinage5_p_g "Age: fourth child (age groups)"	*Aggregation of the age data from dkinage5_p into ten age groups, based on common age limits (nursery, school, young people, adolescents, etc.).*
D-Modules	

vausberufo_p_g2 "Vocational training (KldB2010 - 3 digit)"	*Aggregation of the coded occupational details of the training occupation from vausberufo_g1 (occupational categories according to KldB2010 - 5-digit) to occupational groups according to KldB2010 (3-digit).*
vausberufo_p_g3 "Vocational training (KldB2010 - 2 digit)"	*Aggregation of the coded occupational details of the training occupation from vausberufo_g1 (occupational categories according to KldB2010 - 5-digit) to main occupational groups according to KldB2010 (2-digit).*
sabserhs1_p_g2 "Last higher education degree: university (federal states/abroad)	*Aggregation of the location data from sabserhs1_p_g1 to 16 German federal states (and foreign countries if applicable).*
sabserhs1_p_g3 "Last higher education degree: university (new or old federal states/abroad)"	*Aggregation of the location data from sabserhs1_p_g1 to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
sabserhs1_p_g4 "Last higher education degree: university (type -> 2 categories)"	*Aggregation of the higher education data from sabserhs1_p_g1 into "University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
sabserhs3_g2 "Last higher education degree: country (NEPS classification)"	*Aggregation of the countries from sabserhs3_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
sabserhs3_g3 "Last higher education degree: country (world regions)"	*Aggregation of countries from sabserhs3_g1 on the basis of continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
hs01_g2 "First study experience: university (federal states/abroad)"	*Aggregation of the location data from hs01_g1 to 16 German federal states (if necessary, additional foreign countries).*
hs01_g3 "First study experience: university (new or old federal states/abroad)"	*Aggregation of the location data from hs01_g1 to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs01_g4 "First study experience: university (type -> 2 categories)"	*Aggregation of the higher education institution data fromhs01_g1 into "University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
hs02_g2 "Second study experience: university (federal states/abroad)"	*Aggregation of the location data from hs02_g1 to 16 German federal states (if necessary, additional foreign countries).*
hs02_g3 "Second study experience: university (new or old federal states/abroad)"	*Aggregation of the location data from hs02_g1 to "old federal states" (former FRG without Berlin), "new federal states" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs02_g4 "Second study experience: university (type -> 2 categories)"	*Aggregation of the higher education institution data from hs02_g1 into "University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
hs03_g2	*Aggregation of the location data from hs03_g1 to 16 German federal states (if necessary, additional foreign countries).*

"Third study experience: university (federal states/abroad)"	
hs03_g3 "Third study experience: university (new or old federal states/abroad)"	*Aggregation of the location data from hs03_g1 to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs03_g4 "Third study experience: university (type -> 2 categories)"	*Aggregation of the higher education institution data from hs03_g1 into "University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
hs04_g2 "Fourth study experience: university (federal states/abroad)"	*Aggregation of the location data from hs04_g1 to 16 German federal states (if necessary, additional foreign countries).*
hs04_g3 "Fourth study experience: university (new or old federal states/abroad)"	*Aggregation of the location data from hs04_g1 to to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs04_g4 "Fourth study experience: university (type -> 2 categories)"	*Aggregation of the higher education institution data from hs04_g1 into "University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
hs05_g2 "Fifth study experience: university (federal states/abroad)"	*Aggregation of the location data from hs05_g1 to 16 German federal states (if necessary, additional foreign countries).*
hs05_g3 "Fifth study experience: university (new or old federal states/abroad)"	*Aggregation of the location data from hs05_g1 to "Old federal states" (former FRG without Berlin), "New federal states (incl. Berlin)" (former GDR incl. Berlin) and "abroad" (if applicable).*
hs05_g4 "Fifth study experience: university (type -> 2 categories)"	*Aggregation of higher education institution data from hs05_g1 into " University" (including teacher training colleges, theological colleges, art and music colleges) and "University of applied sciences" (including colleges of public administration, vocational-, administrative and business colleges) according to the Federal Statistical Office's key index for the winter semester 2020/21 and summer semester 2021.*
jobbbdso_g2 "Current occupation: Job title (KldB2010 - 3 digit)"	*Aggregation of the coded occupational data of the occupation of the current employment from jobbbdso_g1 (occupational categories according to KldB2010 - 5- digits) to occupational groups according to KldB2010 (3- digits).*
jobbbdso_g3 "Current occupation: Job title (KldB2010 - 2 digit)"	*Aggregation of the coded occupational details of the occupation of the current employment from jobbbdso_g1 (occupational categories according to KldB2010 - 5- digits) to main occupational groups according to KldB2010 (2 digits).*
E-modules	
berufswahl_g2 "Pursued employment: job title (KldB2010 - 3 digit)"	*Aggregation of the coded occupational details of the desired occupation from berufswahl_g1 (occupational categories according to KldB2010 - 5- digits) to occupational groups according to KldB2010 (3 digits).*
berufswahl_g3 "Pursued employment: job title (KldB2010 - 2 digit)"	*Aggregation of the coded occupational details of the desired occupation from berufswahl_g1 (occupational categories according to KldB2010 - 5- digits) to main occupational groups according to KldB2010 (2 digits).*

	F-Modules
ainflosia_p_g2 "Current stay abroad: country (NEPS classification)"	*Aggregation of the countries from ainflosia_p_g1 on the basis of the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
ainflosia_p_g3 "Current stay abroad: country (world regions)"	*Aggregation of countries from ainflosia_p_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
ainflo1_g2 "Last stay abroad: country (NEPS classification)"	*Aggregation of the countries from ainflo1_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
ainflo1_g3 "Last stay abroad: country (world regions)"	*Aggregation of countries from ainflo1_g1 based on continents. North and South America combined Europe additionally differentiated by EU/non-EU.*
ainflo2_g2 "Second to last stay abroad: country (NEPS classification)"	*Aggregation of the countries from ainflo2_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
ainflo2_g3 "Second to last stay abroad: country (world regions)"	*Aggregation of countries from ainflo2_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
ainflo3_g2 "Third to last stay abroad: country (NEPS classification)"	*Aggregation of the countries from ainflo3_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
ainflo3_g3 "Third to last stay abroad: country (world regions)"	*Aggregation of countries from ainflo3_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
aplanlando1_g2 "Planned stay abroad: country 1 (NEPS classification)"	*Aggregation of the countries from aplanlando1_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
aplanlando1_g3 "Planned stay abroad: country 1 (world regions)"	*Aggregation of countries from aplanlando1_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
aplanlando2_g2 "Planned stay abroad: country 2 (NEPS classification)"	*Aggregation of the countries from aplanlando2_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
aplanlando2_g3 "Planned stay abroad: country 2 (world regions)"	*Aggregation of countries from aplanlando2_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
aplanlando3_g2 "Planned stay abroad: country 3 (NEPS classification)"	*Aggregation of the countries from aplanlando3_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*
aplanlando3_g3 "Planned stay abroad: country 3 (world regions)"	*Aggregation of countries from aplanlando3_g1 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
hzb_bl_g2 "University entrance qualification: state (new or old federal states/abroad)"	*Aggregation of the location data from hzb_bl_g1 to "old federal states" (former FRG without Berlin), "new federal states" (former GDR incl. Berlin) and "abroad" (if applicable).*
wohnland_g2 "Residence lecture period: country (NEPS classification)"	*Aggregation of the countries from wohnland_g1 based on the NEPS anonymization classification. Including former Yugoslavia and former Soviet Union. Europe additionally differentiated by EU/non-EU.*

wohnland_g3 "Residence lecture period: country (world regions)"	*Aggregation of countries from wohnland_g3 based on continents. North and South America combined, Europe additionally differentiated by EU/non-EU.*
wohn_bl_g2 "Residence lecture period: state (new or old federal states/abroad)"	*Aggregation of the location data from wohn_bl_g1 to "old federal states" (former FRG without Berlin), "new federal states" (former GDR incl. Berlin) and "abroad" (if applicable).*