Frank Multrus

The Student Survey

Data and Method Report on the Surveys from 1983 to 2016

I Introduction

The Student Survey is a permanent observation of the study situation and student orientations at German universities. The questionnaire was developed by the AG Hochschulforschung of the University of Konstanz and conducted for the first time in 1982. Every 2-3 years another survey was conducted, most recently the 13th wave was collected in 2015/16. The broad range of topics covers the entire higher education sector, from the time before high school to career transitions, and incorporates job-related, social, and political attitudes. In individual surveys, specific topics were included in greater depth. The survey was conducted at approximately 25 universities and universities of applied sciences selected according to theoretical characteristics. These universities and universities of applied sciences cover a broad spectrum of subjects, are distributed throughout Germany, and have a certain minimum size. The students at these universities were randomly selected for the survey, with the size of the random sample adjusted to the size of the university. Since the beginning of the surveys, the project has been funded by the BMBF.

For the transfer of the data sets to the Research Data Center for Higher Education and Science Research (FDZ-DZHW) for the purpose of making the data available for scientific analyses, the 13 data sets were processed, merged and documented. There they are made available as Scientific Use Files (SUF) for secondary scientific use. In addition to the survey datasets, documentation materials on the datasets and on the execution of the studies are also provided.

II Overview of the 13th Student Survey from WS 2015/16

Series of studies	Constance Student Survey; Study Situation and Student Orientations at Universities and Universities of Applied Sciences
Survey institution	AG Hochschulforschung, University of Konstanz
Project employees	Werner Georg, Frank Multrus, Tino Bargel, Sandra Majer, Monika Schmidt, Hans Simeaner, Doris Lang
Funding	Bundesministerium für Bildung und Forschung (BMBF)
Version of the data set	13th wave
Grundgesamtheit:	322 German Universities and Universities of Applied Sciences, which correspond to the OECD classification of level 5 ISCED (5A and 5B) in the tertiary sector of education.
Sampling procedure	Stratified random sample, strata according to theoretical specifications. Selection of universities: 16 Universities and 12 Universities of Applied Sciences according to theoretical specifications (largest location, range of subjects). Student selection: simple stratified random sample, adjusted to university size.
Survey method	standardized postal survey (paper pencil) and standardized online survey.
Survey period	End of November 2015 – End of March 2016
Response rate and evaluable cases	Gross sample: Paper: 36.902 (adjusted mailing), Online: 33.369 Net sample: Paper: 5.620, Online: 2.929

	Response rate: paper: 15,2%; online: 8,8%, combined: 13,3%
Questionnaire size	92 questions with 413 items, plus 19 additional questions with 125 items, divided into 4 additional modules
Themes	1. university access, subject choice, motives and expectations 2. study strategies, course of studies and qualifications 3. study intensity, time required and duration of study 4. study requirements, regulations and examinations 5. contacts and communication, social climate, counseling 6. subject studies, situation of teaching and quality of studies 7. student lifestyle, social situation, employment 8. difficulties, impairments and burdens 9. university policy: participation and student representation 10. internationality, study abroad and the European Higher Education Area 11. career choice, career values and fields of activity 12. career prospects and labor market reactions 13. democratic attitudes and social perceptions 14. political participation, location and goals 15. wishes and demands for university development 16. demographic and educational biographical data
DOI	
Data Products	SUF
Data Structure	Personal data records
Materials for the 13th Student Survey	• Data and method report (german) • Study overview (german) • Data set 13th survey (STATA) • Codebook for the 13th survey as coded questionnaire • Variable list for the 13th survey with labels and characteristic values • Original questionnaire for the 13th survey • Data set multi-wave file, surveys 1-13 • Code book for the multi-wave file, surveys 1-13 • Map with geographical location of all sample universities • Cover letter to students in WS 2015/16 for paper survey • Cover letter to students in WS 2015/16 for online survey • Cover letter to students in WS 2015/16 for paper and online survey (electoral group) • Welcome letter students-online in WS 2015/16 • Information letter for students in WS 2015/16 • Reminder letter for students in WS 2015/16 • Documentation of subjects in the student survey • List of publications on the Student Survey

Base publications on the 13th student survey

- Multrus F./ S. Majer / T. Bargel / M. Schmidt (2017): Studiensituation und studentische Orientierungen. 13. Studierendensurvey an Universitäten und Fachhochschulen. BMBF.
- Multrus F./ S. Majer / T. Bargel / M. Schmidt (2017): Studiensituation und studentische Orientierungen. Zusammenfassung zum 13. Studierendensurvey an Universitäten und Fachhochschulen. BMBF.
- Simeaner, H., Multrus, F., & Kolbert-Ramm, C. (2017). Datenalmanach. Studierendensurvey 1993 – 2016 nach Geschlecht. Hefte zur Bildungs- und Hochschulforschung (98), AG Hochschulforschung, Universität Konstanz, September 2017.
- Simeaner, H., Multrus, F., & Kolbert-Ramm, C. (2017). Datenalmanach. Studierendensurvey 1993 – 2016. Hefte zur Bildungs- und Hochschulforschung (97), AG Hochschulforschung, Universität Konstanz, September 2017.
- Basic enumeration Universities and technical colleges 13th survey
- Multrus F. / S. Majer (2017): Methodenbericht zum 13. Studierendensurvey. Vergleich Papier-Onlinebefragung. Werkstattbericht. Hefte zur Bildungs- und Hochschulforschung (95), AG Hochschulforschung, Universität Konstanz, September 2017.

Further publications

- Multrus, F. (2018). Bedingungen und Effekte von realisierter und geplanter Auslandsmobilität bei Studierenden. In: Becker, K. & Heißenberg, S. (Hrsg): Dimensionen studentischer Vielfalt. Empirische Befunde zu heterogenen Studien- und Lebensarrangements (S. 196-250). Bielefeld: wbv Publikationen.
- Majer, S. (2018). Studienerfolg von traditionellen und beruflich qualifizierten Studierenden. In Bornkessel, P. (Hg.), Erfolg im Studium. Konzepte, Befunde und Desiderate (S. 175-225). Bielefeld: wbv Publikationen.
- Bargel, T. (2017). Studium im Wandel. Im Zeitraum von 30 Jahren: 1983 – 2013. Studentische Erfahrungen und Evaluation. Hefte zur Bildungs- und Hochschulforschung (92), AG Hochschulforschung, Universität Konstanz, Juli 2017.
- Bargel, T. (2017). Studentische Orientierungen gegenüber Studium, Beruf und Politik im Wandel. Zeitreihe des Studierendensurveys 1983-2013. Hefte zur Bildungs- und Hochschulforschung (91), AG Hochschulforschung, Universität Konstanz, Juli 2017.
- Multus, F. (2016). Der Studierendensurvey. Konzept, Aufbau und theoretische Grundlagen. Hefte zur Bildungs- und Hochschulforschung (87), AG Hochschulforschung, Universität Konstanz, Juni 2016.
- Schmidt, M., Bargel, T., Georg, W., & Gorbachyk, A. (2016). Studienzwecke und Berufsaussichten - ein Vergleich Studierender in Kiew und München. Hefte zur Bildungs- und Hochschulforschung (88), AG Hochschulforschung, Universität Konstanz, September 2016.
- Georg/Bargel: Parental qualifications as determinants of university entrance and choice of a field of study in Germeny. European Journal of Higher education.
- Georg, W. & Bargel, T. (2016). Soziale Determinanten der Studienaufnahme und Fachwahl. In Grossmann, D. & Wolbring, T. (Hg.), Evaluation von Studium und Lehre (S. 93-121). Wiesbaden: Springer.
- Bargel, T., Multrus, F. & Majer, S. (2017). HSW-Gespräch zum Zusammenhang von Veranstaltungsbesuch und Studienqualität mit Tino Bargel, Frank Multrus und Sandra Majer. „Das Hochschulwesen“. HSW 4+5, 154-162.

III Data usage information

[Conditions of data use] The data of the Student Survey are released by the FDZ of the DZHW as Scientific Use Files (SUF) exclusively for secondary scientific use. The SUF can be requested via the FDZ website (https://fdz.dzhw.eu). A data use contract is concluded for the use of the SUF. Students or doctoral candidates who are not employed by a scientific institution must sign a data use contract together with a supervisor. In the course of concluding the contract, the FDZ also checks whether there is a scientific interest in using the data.

[Data access] The SUF is offered via two access paths, which are restrictive in different ways with regard to the storage location, the possibility of independent linking with external data, and the FDZ's control options:

- **Download:** The data is encrypted and made available for download on the FDZ website. Data users can save the data on their local computer, link it with data from external sources if desired, and analyze the data with their own software.
- **Remote-Desktop:** The data are made available on a terminal server of the FDZ. Via a specially secured connection between the local computer of the user and the terminal server of the FDZ, the data can be analyzed with the software available on the terminal server. It is not possible to transfer the data to the local computer. Analysis results are only released and made available after the FDZ has checked that they do not violate data protection laws.

[Data provision fees] SUF are currently (as of October 2021) provided free of charge. Changes or the current cost model can be viewed on the FDZ-DZHW website (https://fdz.dzhw.eu).

[Obligations of the data users] Data users are obliged to comply with the following rules:

- **Scientific use:** The data may only be used for scientific purposes. Commercial use is prohibited**.**
- **De-anonymization prohibition:** Any attempt to re-identify units of analysis (e.g., persons, households, institutions) is prohibited.
- **Requirement to report security breaches:** If data users become aware of security vulnerabilities with regard to data protection or data security, these must be reported to the FDZ-DZHW without delay.
- **No passing on of the data:** The data may only be used by the persons who have concluded a data use contract.
- **Provision/reporting of publications:** Any type of scientific work (e.g. publications, theses, dissertations) resulting from work with FDZ data must be reported to the FDZ. An electronic version of the printed version must be made available to the FDZ.
- **Citation obligation:** The data used must be cited in publications, other works (e.g. theses) and presentations according to the specifications of the FDZ.

1 Content and design of the student survey (all surveys)

[Study Series]

The development of the Student Survey goes back to work from the project "Higher Education Socialization" in the Collaborative Research Center "Educational Research" (SFB 23) at the University of Konstanz, in the context of which high school graduates, students and academics were studied. In concept, the Student Survey is understood as a long-term social observation, whereby the core of the surveys has been the measurement of performance in the higher education sector since the beginning, in order to provide a basis for higher education policy and its public debate.

Detailed information on the development, structure, and theoretical framing is compiled in issue 87 of the series "Hefte zur Bildungs- und Hochschulforschung": Multrus, F. (2016). The student survey. Concept, design, and theoretical foundations. AG Hochschulforschung, University of Konstanz.

The surveys of the Student Survey cover a period of 35 years, during which the study situation and student orientations at German higher education institutions (universities and universities of applied sciences) were surveyed and analyzed. Erhebung: WS 1982/83

1. Survey: WS 1982/83
2. Survey: WS 1984/85
3. Survey: WS 1986/87
4. Survey: WS 1989/90
5. Survey: WS 1992/93
6. Survey: WS 1994/95
7. Survey: WS 1997/98
8. Survey: WS 2000/01
9. Survey: WS 2003/04
10. Survey: WS 2006/07
11. Survey: WS 2009/10
12. Survey: WS 2012/13
13. Survey: WS 2015/16

The individual surveys are cross-sectional surveys. However, care was taken to ensure that, as far as possible, the same universities were always surveyed, so that comparable environmental conditions are available. This could be realized for the majority of the participating universities over a long period of time.

[University Sample]

The 25-30 selected sample universities were contacted in the spring of the year of the survey and invited to participate. The initial contact was made via the respective rectorate, the further contacts via the responsible staff members, with whom the procedure was coordinated by telephone.

The selection of the universities was based on certain criteria. Universities, technical universities, comprehensive universities (in the early survey phases) and universities of applied sciences were to be involved. A broad spectrum of subjects was to be offered at the universities, so that as many subjects as possible were represented. Therefore, no specialized universities were included. For the selection of the universities of applied sciences, this meant that they had to offer a range of subjects from at least three subject groups (social sciences, economics, engineering), so that differentiated comparisons with the universities would be possible. Furthermore, universities from all federal states should be involved so that statements for the whole of Germany are possible. In some defined cases,

the participating universities and universities of applied sciences should come from the same city, if possible, so that comparisons with controlled regional conditions are possible.

In the course of the surveys, the university sample was expanded, and new universities were recruited according to the same conditions. After the fall of communism, universities from the new German states were included in the university sample at the beginning of the 1990s. During the new recruitments at the turn of the millennium, newly founded universities from the 1950s were specifically selected in order to enable comparisons to universities with a long tradition. All these universities were invited back in all subsequent surveys from the time of initial contact.

Sample universities in the student survey 1982-2015

University

University of Applied Science

[Student Sample]

The students were selected at the respective universities. As a guideline, the responsible staff members received instructions from the Working Group on University Research (AGH) that only German students should be selected who were not in a vacation semester. If possible, all subjects and all semesters should be represented. The universities were then to randomly select a certain number of students from their student files from this pool. This number depended on the size of the university and was predefined by the AGH.

[Conduct of surveys]

The postal documents, which consisted of a questionnaire, cover letter and return envelopes on the one hand, and an envelope with a reminder letter on the other, were pre-packaged from the end of November and sent to the participating universities by mail. There, two sets of address labels had to be printed out for the selected students. These were attached by the universities to the questionnaire documents and the reminder letters. Then, only the questionnaire documents were mailed to the students at first. This marked the start of the field period. The cover letters to the students, the background information and the reminder letters can be seen in the documentation, as well as the distribution of the universities per survey.

The surveys were conducted in the winter semester at the end of November / beginning of December, so that the first-year students had at least 6-8 weeks of study experience. The field period was mostly extended to the end of March / beginning of April, with the reminder letter being sent to all students contacted by the universities in January.

During the field period, the questionnaires returned directly to the AG by return envelope were unpacked, subjected to an initial check, and stamped with a receipt stamp and sequence number. The collected completed questionnaires were sent in tranches to a data collection company, which transferred the written information into a digital form. This process was accompanied by several control checks.

From May / June, the digitalized raw data of the survey were available and could be processed in the AGH. Then the first analyses and the preparations for the publications took place.

[Questionnaire]

The questionnaires consisted of double-sided A3 pages that were stapled in the middle, resulting in an A4 format. In addition, they had a colored cover page. On the front cover the name of the project (research project student situation) and name and address of the implementing institution (project leader and University of Konstanz) were written. In addition, short notes on the voluntary nature and on how to answer the questionnaire were given on the cover page. Further information was sent with a cover letter until 1992.

Starting with the sixth survey in the winter semester 1994/95, the front and back of the questionnaire were additionally printed with the information about the study and a separate cover letter was dispensed with. From then on, students were also informed that they could receive a summary of the results if they enclosed a postcard with their address. This measure was intended as an incentive and was also used by many students. In addition, four different colors were chosen for the envelope design, differentiating by type of university and old or new states. This served to ensure that, in the absence of a university name, at least a rough allocation to the aggregate was possible (e.g. universities in the new federal states).

For the 12th survey in the winter semester 2012/13, the entire questionnaire was revised in design and rearranged according to suggestions from GESIS. In the process, the cover design was also redesigned and the information about the study was again included in a cover letter. In the last survey of WS 2015/16, the cover sheet was redesigned again, this time according to suggestions from the graphic design office of the University of Konstanz.

The inner part of the questionnaire comprised 20 A4 pages in the first six surveys. In the new millennium, the survey was shortened to 16 pages. For the 13th survey it was shortened again and limited to 13 pages. Up to the 11th survey, the layout of the inner questionnaire was table-like, so that the questions and the answers were each in a separate column. This layout was broken up for the 12th survey according to suggestions from GESIS. For the 13th survey, however, the table structure was again favored.

All questionnaires with cover sheets and cover letters are to be included in the documentation.

[Topics]

A wide variety of topics were collected in each survey. Certain topics appeared in all surveys, including social data and information about the studies (subject, semester, degree, etc.), but also questions about the time before the studies, the experience of the studies, the requirements, the difficulties and burdens, the wishes, the yield or the evaluations of the teaching, as well as the time after the studies with career ideas or social and political attitudes.

In addition, some surveys also focused on specific topics, some of which were maintained over several surveys. For example, questions on reunification, business start-ups, computer use or the European Higher Education Area.

In total, all subject areas can be compiled into the following list:

- Demographic and educational biographical data
- University access, subject choice, motives and expectations
- Study strategies, course of studies and qualifications
- Study intensity, time required and duration of studies
- Study requirements, regulations and examinations
- Contacts and communication, social climate, counseling
- Subject studies, situation of teaching and quality of studies
- Student lifestyle, social situation, gainful employment
- Difficulties, impairments and burdens
- University policy: participation and student representation
- Internationality, study abroad and European Higher Education Area
- Career choice, career values and fields of activity
- Career prospects and labor market reactions
- Democratic attitudes and social perceptions
- Political participation, location and goals
- Wishes and demands for university development
- New media in studies and teaching
- Computers, Internet and new media
- Acceptance of new study structures and degrees
- Perceptions about university and science
- Role and chances of university graduates
- Exchange between old and new federal states

[Publications on the surveys]

The following publication types were produced in each survey:

- Basic counts of all variables, by universities and universities of applied sciences.
- Indicator volumes for the participating universities, basic counts of designated characteristics according to the subject groups there. These volumes were sent exclusively to the participating universities and are not shown here as documentation.
- Data almanac, a basic count of all characteristics in time comparison, by type of university and subject groups.
- Data almanac, additionally separated by gender.
- Main report "Study situation and student orientations". Publication of the most important findings on a wide range of topics.
- Abridged version of the main report.
- Focus reports. In-depth analyses and findings on selected areas. These include the subject monographs that were compiled for individual subjects (medicine, law) or subject groups (humanities, engineering, natural sciences), as well as for types of degrees (bachelor's, master's). Other main topics were: Comparison of methods paper - online, study entry phase, research and practice, women in studies, change of political orientations, young academics among students.
- Issues of the series "Hefte zur Bildungs- und Hochschulforschung" (booklets on education and higher education research), which cover various topics, including documentation as well as specialized reports or methodological studies.
- In addition, there are scientific book contributions and articles as well as other trans-fer publications.

All publications are compiled on the website of the AG Hochschulforschung (https://www.soziologie.uni-konstanz.de/ag-hochschulforschung/publikationen/).

[Special features of the 13th survey]

In the 2015/16 winter semester, an online questionnaire was used in addition to the paper questionnaire. The aim was to check whether a switch to future online surveys would have an impact on student response behavior. Therefore, the questionnaire was redesigned into an online version, but the content was left identical. To ensure that a comparison between the responses of the paper and online respondents is methodologically secure, the student sample was doubled so that the same number of students would receive a paper questionnaire as received an invitation to the online survey with a link and access code.

In addition to the direct comparison of the paper and online versions, the 13th survey of the Student Survey was also used to test further possibilities of online conditions. For this purpose, a complex design was developed that divides the online sample into seven subgroups.

- One group received both the paper questionnaire and online access and could choose which format to answer.
- Two groups received the full, very long online questionnaire. For one group, motivational pages were added to the questionnaire to see if this approach could strengthen perseverance.
- Four groups received a version shortened by 40%. For these groups, extra modules were developed, which were inserted at the end of the questionnaire and contained very specific additional questions.

Independently of this, the sample universities were asked firstly whether it was possible to send a second reminder by e-mail and secondly whether it would be possible to advertise the survey on the websites and in the social media channels. Around half of the sample universities were prepared to take each of these measures, so that comparisons between the measures were possible.

One university only agreed to an online survey, so no mail was sent there, instead the invitation was sent by e-mail.

The implementation and data collection of the online survey were conducted by DZHW. After completion of the field period, the raw data were sent to AGH.

The analyses comparing the answers of the students from the paper and online group could prove that a change is possible without observing a significantly different result in the answering behavior. The different measures showed different effects. A detailed documentation of these comparisons was published in a method report and can be found in issue 95 of the series "Hefte zur Bildungs- und Hochschulforschung": Multrus, F. & S. Majer (2017). Methods report on the 13th Student Density Survey. Comparison paper-online survey. Working Group on Higher Education Research, University of Konstanz.

[Return Table]

The response rate for the individual surveys was over 40% until the mid-1990s. After that, student participation systematically declined. The individual responses and be-cleaned sample sizes are listed in the table.

Questionnaires mailed and participation in the thirteen surveys of the Student Survey by type of university (1983 - 2016).
(Figures in absolute and percentage terms)

Former Federal territory	**Total**	**Universitys**	**Universities of Applied Sciences**
WS 1982/83			
Shipping (adjusted)	18.940	16.586	2.354
Response: absolute	7.817	6.607	1.059
in percent	41,3	39,8	45,0
WS 1984/85			
Shipping (adjusted)	22.470	17.500	4.970
Response: absolute	10.038	7.663	2.324
in percent	44,7	43,8	46,8
WS 1986/87			
Shipping (adjusted)	22.400	17.400	5.000
Response: absolute	9.852	7.532	2.279
in percent	44,0	43,3	45,6
WS 1989/90			
Shipping (adjusted)	19.730	15.380	4.350
Response: absolute	8.812	6.999	1.813
in percent	44,7	45,5	41,7
Federal Republic of Germany (Old and New States)			
WS 1992/93			
Shipping (adjusted)	20.175	15.519	4.656
Response: absolute	9.240	7.192	2.048
in percent	45,8	46,3	44,0
WS 1994/95			
Shipping (adjusted)	19.644	15.051	4.593
Response: absolute	8.461	6.582	1.879
in percent	43,1	43,7	40,9

WS 1997/98			
Shipping (adjusted)	19.641	15.118	4.523
Response: absolute	7.271	5.799	1.472
in percent	37,0	38,4	32,5
WS 2000/01			
Shipping (adjusted)	23.391	18.064	5.327
Response: absolute	8.130	6.385	1.745
in percent	34,8	35,4	32,8
WS 2003/04			
Shipping (adjusted)	27.423	21.498	5.475
Response: absolute	9.975	8.307	1.668
in percent	36,4	37,8	30,5
WS 2006/07			
Shipping (adjusted)	26.064	20.844	5.220
Response: absolute	8.350	6.894	1.456
in percent	32,0	33,1	27,9
WS 2009/10			
Shipping (adjusted)	27.307	21.570	5.737
Response: absolute	7.590	6.117	1.473
in percent	27,8	28,4	25,7
WS 2012/13			
Shipping (adjusted)	26.274	19.927	6.347
Response: absolute	4.884	3.792	1.092
in percent	18,6	19,0	17,2
WS 2015/16 Paper			
Shipping (adjusted)	30.789	23.479	7.312
Response: absolute	4.959	3.885	1.074
in percent	16,1	16,5	14,7
Selected group[1] Paper			
Shipping (adjusted)	6.122	4.763	1.359
Response: absolute	661	508	153
in percent	10,8	10,7	11,3
WS 2015/16 Paper total			
Shipping (adjusted)	36.902	28.235	8.669
Response: absolute	5.620	4.393	1.227
in percent	15,2	15,6	14,2
WS 2015/16 online			
Shipping (adjusted)	27.244	21.296	5.948
Response: absolute	2.656	2.071	581
in percent	9,7	9,7	9,8
WS 2015/16 Selected group online (Long and short version)			
Shipping (adjusted)	6.122	4.763	1.359
Response: absolute	273	201	70
in percent	4,5	4,2	5,1
WS 2015/16 online total			
Shipping (adjusted)	33.369	26.062	7.307
Response: absolute	2.929	2.272	651
in percent	8,8	8,7	8,9
Selected group total (paper + online)			
Shipping (adjusted)	6122	4.763	1.359
Response: absolute	934	709	223
in percent	15,3	14,9	16,4

[1] Selection group was able to choose whether to fill in paper or online. Therefore, response calculations for the voting group, differentiated by paper or online, only show correct values in total.

WS 2015/16 total (paper + online)			
Shipping (adjusted)	64.286	49.642	14.646
Response: absolute	8.549	6.665	1.878
in percent	13,3	13,4	12,8

Source: Student Survey 1983-2016, AG Hochschulforschung, University of Konstanz

[Records and codebooks]

In each of the 13 surveys, a large proportion of variables remained unchanged. A smaller part was varied, on the one hand due to methodological developments, and on the other hand due to current developments or new focal points. In this respect, there is a large body of data that has remained stable over time and covers several to all worlds. At the same time, there are data that can only be found in a single survey. All variables that were asked in more than one wave are part of the multi-wave data set. Data that were only collected once can also only be found in the corresponding single data set.

Due to the revision of the data sets and their documentation, the variables have been recoded. Each variable now has the same name or number in each wave and in the multiwave file. For example, the variable "age" was coded as "v4" in all data files.

In the order of variable naming, the variables from the 13th wave were used first, followed systematically by the later waves. Variables that occur only once received their number only after all multi-wave variables from all surveys had been named. Therefore, a complete sequence of variable numbers is not included in each wave. In the data sets, all variables and all responses received labels.

For each data set there is a code book for the assignment of the data to the variables. For the individual waves these are the questionnaires themselves, in which the variable numbers were inserted with the corresponding items. For each wave a variable list was created, which lists all variables with additional information such as number of observations, number of different expressions, value range, mean, minimum, maximum as well as the corresponding labels for the variables. The files are labeled with the survey wave and the year of the survey.

In the 13th survey, the total sample was divided into nine groups, each with a different design (see above). Not all variables were asked in all groups. Therefore, three different colors were chosen for the variable numbers in the questionnaire codebook to indicate which variables belong to which groups. The blue variable numbers indicate the variables that are present in the short versions (shortened by 40%) (groups 5-8). These variables are of course also present in the long versions and thus in all groups. The red variable numbers indicate the variables that were only asked in the long versions (groups 0-4 and 9). The green variable numbers indicate those variables which were only additionally asked in the four modules of the short versions, i.e. which only occur in one of the four groups (5-8).

A separate codebook was created for the multi-wave file, which assigns the variable numbers to the questions and items in tabular form, indicating the characteristics and wave affiliation as well as the reference to the question numbers of the individual waves. In addition, there is an extra column for comments, which contains important information about special features and the use of the data. The sequence of the variables was determined on one hand according to content-related topics and on the other hand according to their occurrence over time. In other words, we started with those variables that were present in the 13th survey and had also been asked in earlier waves. However, older data were also included if they were thematically relevant. This is followed by the variables that were

only asked up to the 12th survey, and then the variables that were asked up to the 11th, and so on. The last entries are then those variables that were only present in the first two waves. In total, the SUF-Remote dataset of the multiwave file contains 1,207 variables with a volume of 108,969 cases. The more anonymized SUF download dataset contains one variable less.

For certain groups of variables, new variables were created, primarily in order to create aggregates. This applies, for example, to the differentiation by type of higher education institution (universities vs. universities of applied sciences) or the federal states (old vs. new federal states) of the higher education institutions. In addition, the large number of subjects was categorized into subject groups and study areas. Likewise, the data on foreign languages. Sum scores were created for a few variables. These new variables are all listed and explained in the codebook for the multiwave file. In the codebooks of the single waves, they are included in the variable lists and labeled, in the coded questionnaires they are listed before the questionnaire pages. Most of these variables were already scheduled when the files were created, which is why they are included in the running variable numbers. A few were created later, which is why they received additional identifiers to the already existing variables. For this purpose, a letter was appended to the variable number (e.g. 75b, v85a, v274k or v352i), which stands for cleaning (see below), recreation, categorization or indicator creation for the variable in question.

In the codebook of the multiple wave file, the questions, items, and answer options are presented as they appear in the original questionnaire. These details could not be transferred one-to-one to the labels in the data set, as some of them would have been too long. Therefore, short names were used for the labels to represent the most important content.

[Documentaries]

In addition to the data sets and codebooks, further documentation is available. On the one hand, this concerns original materials, such as original questionnaires, invitation letters, reminder letters and information material. On the other hand, special documentation on the student subjects of the Student Survey and their coding over time has been prepared. There, the compilations of the aggregated variables are also described, which include subject groups and fields of study. In addition, detailed references are made to the compatibility with the coding of the Federal Statistical Office.

[Value ranges]

The students' responses are all coded as numerical values. Seven-item scales are most common, with most having a value range of 0-6. Rating scales that rank from -3 to +3 are coded 1-7 but are named in the labels. This coding dates back to the early survey years and was retained here for traditional reasons. All other scales (e.g., four- or five-level) as well as nominal choice categories were given value ranges from 1 to x (=highest number of levels or choices).

For all values, the corresponding labels were included in the data sets. In addition, the questionnaire codebooks and the codebook of the multi-wave file show which numerical codes represent which answer specifications or values. Changes in value ranges over time are indicated in the documentation of the multiwave file.

Only a very few questions in the surveys had open response options. With the exception of foreign languages, these open responses are not included in the data set.

[Missing values]

Missing values are present in all data sets because not all students always answered all questions. In addition, it was determined in each survey at what level of dropouts the answers are still included. Therefore, there are also questionnaires in which, for example, only half were answered.

The newly created multi-wave file of the student survey for the FDZ contains all variables that were asked in at least two surveys. Five different missing codes were used to assign the data.

1. [.] Simple Missing: This refers to all missing values that are due to the fact that the respondents did not provide any information on the corresponding item. The standard missing code of statistical systems is used for this purpose. In STATA this is the dot ".".

2. [-9] System missing for survey waves. Not all variables were questioned again in all waves. Variables that were not queried in a wave are given the code: "-9".

3. [-10] System missing for online short survey. This missing code only exists in the 13th survey. There, all variables were used in the online long versions that were also present in the paper versions. In addition, short versions were created in four additional modules that no longer contained all variables. The variables that were not asked in these online short versions of the 13th survey are given the code "-10".

4. [-11] System Missing for Filter Questions. In the Student Survey, there are some filter questions whose answers refer to a preliminary question. Typical examples for such preliminary questions are questions about activities (e.g., stay abroad), which can be answered with "yes" or "no". In such cases, an additional date is usually asked, such as the duration of the stay. If in these cases the respondents indicate "no" in the preliminary question and the subsequent question remains unanswered, then the code "-11" is assigned in order to emphasize that this information cannot be provided at all, since the condition for this is not given. In cases where respondents select "Yes" in the preliminary question or do not provide any information on this (= missing), the simple missing code is used for non-responses to the follow-up question.

5. [-12; -13] System missing for cleanup. Some items have been cleaned in their value ranges. These cleansed variables are given the suffix b (e.g.: v66b) in the data set, while the original variables are retained (v4). Information that has been excluded due to the adjustment is given the missing code "-12" (see "Cleanup" below). The information on semesters is a special case. Here, the additional system missing code "-13" was assigned if the information for subject semesters or university semesters is greater than 30.

[Weighting]

There are no weighting variables in the data sets.

[Subject codes]

In the Student Survey, the subjects are selected from a list of subjects on the inside of the cover page of the questionnaire, where they are given a numerical code. In the data sets, the subject names of the list are set as labels. From the subjects, additional variables were created for study areas and for subject groups, including compilations as reported by the Federal Statistical Office. The coding of the subjects and their assignment to the subject groups as well as to the Federal Statistical Office labels is described in a separate documentation (see file: "Subjects"). There, special features and problems of the codings over the course of time are also discussed.

The lists of the questionnaire codebooks and those of the original questionnaires differ in some places. This is due to the fact that the code lists for the multi-wave file had to be newly created. This coding was then also transferred to the single waves, so that the same codes are always present in all files. The original questionnaires, on the other hand, have the original coding, which was not consistent over time.

[Anonymization]

Due to data protection regulations, it must not be possible to draw conclusions about individual persons from the data. Individuals could only be identified in the data sets of the student survey if there were access to the respective university's own student files and an assignment of the invited students to the questionnaires. Such an assignment of the participating students could neither be made by the universities nor by the AGH in the case of the paper surveys, since the questionnaires were already pre-packaged and the randomly selected addresses were glued onto them by the universities. Since the completed questionnaires were returned directly to the AGH by means of envelopes, the universities also no longer had access to the documents.

The 2015 online survey used access by using tokens that were randomly linked to the email invitations. The token list was managed by the DZHW, the email address list by the universities. The two lists remained separate and both lists were accessed only for linking the email invitation, with no shared file created at any time. The universities had agreed in advance to delete the e-mail lists immediately after the invitations were issued. The AGH only received the data of the online questionnaires with tokens, but no e-mail data.

It is not possible to relate the questionnaires to individuals without student data from the universities. With a great deal of effort, references to individuals could perhaps be made if very specific subsamples were used that included only a few people at certain universities.

At the same time, it should also be prevented or made significantly more difficult to determine the universities. For this purpose, the variables containing information on the universities were also anonymized.

Anonymization specifications are performed at two levels:

A. For the download files as SUF, higher anonymization requirements are specified. Therefore, some variables are excluded, and others are categorized or topcoded. In detail, these are the following changes:

1. The variable with the university identifier was removed from the data sets (v8). Instead, an aggregated variable was created that distinguishes between universities and universities of applied sciences (v9).
2. The coding of the federal states in which the universities are located was aggregated so that a distinction is only made between "old" and "new" federal states (v10).
3. The variable on disabilities and chronic diseases was removed (v18, v19).
4. The variable age was categorized (v4).
5. The variable number of children was topcoded (v7).
6. The variable state of higher education entrance qualification was categorized so that only old and new states are distinguished (v64).
7. The variables for the second and third subject were categorized into subject groups (v41, v42).
8. The variable grade in the final certificate of higher education entrance qualification was categorized (v74).

9. The variables university semester and subject semester were topcoded (v75, v76).
10. Variables on foreign language courses in the degree were categorized (v274-v277).
11. Variables on year of higher education entrance qualification were categorized (v60-v63, v66).
12. For some variables, small subgroups were shifted to the category "other". This applies to the variables on advanced courses in high school (v71-v73) as well as the subject coding of the fields of study (v40) or areas of study (v44).

B) for the SUF files used via remote access, many of these variables remain in the original. Only the following changes are made here:

1. The variable with the university identifier (v8) was retained, but the codes were assigned unsystematically as numerical codes, so that there are assignments across the surveys, but there is no assignment by name.
2. The variable on disabilities and chronic illnesses was removed (v18, v19).
3. The variable age was categorized (v4).
4. The variable number of children was topcoded (v7).
5. The variables university and semester of study were topcoded (v75, v76).

[Data cleaning]

In the data files, certain variables were cleaned up. In each case, the original variables were retained (e.g. v74) and the cleaned variables were given the suffix "b" (e.g. v74b). All these variables are also described in the codebook for the multiwave file.

The adjustment concerns all open numerical data provided by students in the 13 questionnaires. This includes data for:

- Number of years or year numbers (year of higher education entrance qualification.)
- Number of semesters (subject-related semesters, university semesters, assistant/tutor periods, standard period of study, planned subject-related semesters until gradu-ation)
- Number of months (activities prior to studies, internship, research project, interruption of studies, studies abroad, language courses abroad)
- Number of ECTS points
- Hours (time budget)

The aim of the cleaning process was to remove information that was logically/mathematically impossible or very improbable in real terms. At the same time, however, as much information as possible was to be retained. For this reason, more extreme information was allowed if it is theoretically possible and if it is consistent with other characteristics. However, this still leaves large ranges in the adjusted data.

Data that is eliminated by the purge is assigned the systemissing code -12.

In detail, the following adjustments have been made:

To adjust for year of higher education entrance qualification a match was made with adjusted age and year of survey. The adjusted range for the year of higher education entrance qualification. is 1927-2015.

For the data on the semester of study and university, a comparison was made with the age, as well as a comparison with the time span between the year of the survey and the higher education entrance qualification. In addition, it was checked whether the data on the subject semester were larger than those on the university semester. In the latter case, the university semesters are set

missing (-12). The subject semesters are also set to missing (-12) if bachelor's students without a first degree report more than 20 semesters.

For the 12 variables with monthly data on activities prior to the study, a comparison was made with age, the time span between the year of the survey and the year of the higher education entrance qualification, and the university semesters.

For the length of time for assistant and tutor activities (semesters), an adjustment was made with age and university semester. The adjusted ranges are now 0-24.

Matching was also performed for the other variables with monthly data. For most of the data, age and semester of study or university were included. For the time between first degree and master's degree, the time between HSR year and survey year was also considered. The data for the standard period of study were set to at least 2.

The weekly hours of study (time budget) were not adjusted because none of the individual variables had invalid (too large) values. Four sum scores were formed from these individual variables. However, these scores also reach invalid values (= has more than one week of hours). However, since it is not verifiable which of the summed individual variables was misnamed, none can be cleaned up. Therefore, for each of the four sum scores a new variable was defined, which indicates whether the resulting sum is realistic (value=0, if weekly hours are below 140), improbable (value=1, if weekly hours are between 140 and 168) or impossible (value=2, if possible weekly hours are exceeded).

For the information on the amounts of money available, a comparison is made between "total money" and "of which BaFöG". For BaFöG, missing (-12) is set if the value is above 940.

A second group of adjustments concerns some variables that contain "yes" / "no" answers to activities and then have queries about the duration (e.g.: internship before studies? months?). Here, the "no" or "missing" answers were set to "yes" if a time duration was mentioned afterwards.