

Ute Hoffstätter / Andreas Sarcletti

# 19th Social Survey

Data and Methods Report on the Student Survey 2009

Version 1.0.0

## Data and Methods Report

December 2016



Author(s):  
Ute Hoffstätter  
Dr. Andreas Sarcletti

With assistance from:  
Andreas Daniel (chapter 7)

Editor:  
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)  
German Centre for Higher Education Research and Science Studies (DZHW)  
Lange Laube 12 | 30159 Hannover | Germany | [www.dzhw.eu](http://www.dzhw.eu)  
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | [info@dzhw.eu](mailto:info@dzhw.eu)

Management board:  
Prof. Dr. Monika Jungbauer-Gans  
Karen Schlüter

Chairman of the supervisory board:  
Ministerialdirigent Peter Greisler

Registration court:  
Amtsgericht Hannover | B 210251

The content of this publication is licensed under the Creative Commons License “Attribution - Non-Commercial – Share Alike 3.0” (CC-BY-NC-SA)

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



## Contents

Table of Figures .....	2
List of Tables .....	2
I Introduction .....	1
II Overview of the 19th Social Survey (2009) .....	2
III Data Use Instructions .....	4
1 Content and Structure of the Study .....	7
2 Survey Instruments .....	10
2.1 Contents of the Survey Instruments .....	10
2.2 Pre-tests .....	11
3 Population and Sample Procedure.....	12
4 Implementation of the Survey .....	13
5 Response Rate .....	14
6 Data Preparation .....	15
6.1 Data Transfer .....	15
6.2 Coding of Open Responses .....	15
6.3 Data Checking and Data Cleansing.....	16
6.4 Generation of Variables .....	18
6.5 Generation of the Data Set .....	18
6.6 Assignment of Variable Names, Variable Labels and Value Labels .....	18
6.7 Coding of Missing Values .....	19
7 Weighting .....	21
7.1 Procedure and Instructions for Use .....	21
7.2 Weighting of the Data .....	22
8 Anonymisation .....	24
9 Bibliography .....	29

## Table of Figures

Figure 1: Modes of Access and Analytical Potential.....	5
Figure 2: Responses to the 19th Social Survey (2009) Over Time .....	14
Figure 3: Mode of Access, Statistical Level of Anonymisation and Analytical Potential of the Data from the 19th Social Survey (2009).....	25

## List of Tables

Table 1: Gross and Net Samples and Response Rates of the 19th Social Survey (2009) .....	14
Table 2: Coded Topics and Code Lists Used in the 19th Social Survey (2009) .....	16
Table 3: Thematic Areas and Abbreviations of the 19th Social Survey (2009) Variable Labels.....	19
Table 4: System of the RDC-DZHW for Missing Values .....	20
Table 5: Weights provided for the 19th Social Survey (2009).....	22
Table 6: Statistical Anonymisation Measures for the Data of the 19th Social Survey (2009) .....	27

## I Introduction

The Social Survey carried out on behalf of the German National Association for Student Affairs (DSW) is a series of studies on the social and economic situation of students in Germany existing since 1951.<sup>1</sup> Since 1982<sup>2</sup> (10th Social Survey) it has been conducted by the German Centre for Higher Education Research and Science Studies (DZHW)<sup>3</sup>. Since the 6th Social Survey (1967/1968) the study has been funded by the Federal Ministry of Education and Research (BMBF). The Social Survey serves – in addition to the official higher education statistics – the national educational monitoring.

Within the scope of the BMBF funded project on the development of a Research Data Centre for Higher Education Research and Science Studies at the DZHW (RDC-DZHW), the data contained in some of the more recent surveys are subsequently processed and documented for the purpose of later use.<sup>4</sup> They are made available as *Scientific Use Files* (SUF) for secondary scientific use or as *Campus Use Files* (CUF) for teaching and training purposes via various modes of access. Along with the survey data sets, documentation material on the data sets and the implementation of the studies is provided.<sup>5</sup>

The existing data and methods report is part of the documentation on the 19th Social Survey<sup>6</sup>. Further documentation materials for the study (data set report, questionnaire, question flow diagrams etc.) can be downloaded from the Metadata Search Portal (<https://metadata.fdz.dzhw.eu>, doi: 10.21249/DZHW:ssy19:1.0.0).

Section II of the report provides an overview of the key data of the 19th Social Survey. The key information on the use of the data in this study follows in section III. Chapter 1 presents content and an appendix on the Social Surveys up to 2009<sup>7</sup> in general and on the 19th Social Survey in particular. The additional structure of the report focuses largely on the research process. The survey tool is described in Chapter 2. The survey process (sampling procedure, survey procedure, response) is detailed in Chapters 3 to 6. In Chapters 7 and 8, weighting and anonymisation practices used are presented.

<sup>1</sup> For the latest information on the Social Survey visit the project website (<http://www.sozialerhebung.de>).

<sup>2</sup> The 1st (1951) and 2nd Social Survey (1953) were both performed by Gerhard Kath of the Verband Deutscher Studentenwerke (Association of German Student Services). The 3rd (1956) to 9th Social Surveys (1979) were performed by the Deutsches Studentenwerk (DSW), also by Gerhard Kath (Managing Director of Studentenwerk Frankfurt am Main). Up until the 5th Social Survey (1963) the Social Survey was financially funded by the Federal Ministry of the Interior.

<sup>3</sup> The German Centre for Higher Education Research and Science Studies (Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH [DZHW, <http://www.dzhw.eu>]) was formed in August 2013 through a spin-off of the company HIS Hochschul-Informationen-System GmbH. The term DZHW is used throughout the following text, even if the study was performed before the spin-off took place in 2013. On all original documents in the 19th Social Survey (questionnaires, flyers etc.) as well as in the reports on the projects, the HIS Institute for Higher Education Research (HIS-HF) is described as an actor.

<sup>4</sup> As no subsequent use of the data was envisaged at the time of the survey, some of the information for the survey was not documented with a later reuse of the data in mind and some of this can no longer be reconstructed. This is indicated at the corresponding points in the text.

<sup>5</sup> Information on the available datasets and documentation is provided in the Metadata Search Portal of the RDC-DZHW (<https://metadata.fdz.dzhw.eu>).

<sup>6</sup> Only the sampling of German students and so-called 'Bildungsinländer' (students with foreign nationality who have gained their higher education entrance qualification in Germany) have been taken into account, who were surveyed using a paper questionnaire. The data that was also collected within the scope of the 19th Social Survey on the so-called 'Bildungsausländer' (international students with a higher education entrance qualification not obtained in Germany) and an online subsample have not been documented here. See Chapter 1 for details.

<sup>7</sup> Only the social surveys up to the 19th Social Survey (2009) documented here have been taken into account. The 20th Social Survey (2012) and the 21st Social Survey (2016) have also been edited for subsequent data use.

## II Overview of the 19th Social Survey (2009)

<b>Survey Series</b>	Social Survey on behalf of the German National Association for Student Affairs (DSW)
<b>Survey</b>	19th Social Survey (2009)
<b>Surveying Institution</b>	German Centre for Higher Education Research and Science Studies (DZHW)
<b>Funding</b>	Federal Ministry of Education and Research (BMBF)
<b>Project Contributors (Project Leader)</b>	Lars Borchert, <u>Wolfgang Isserstedt</u> , Maren Kandulla, Michael Leszczensky, Elke Middendorff
<b>Themes</b>	Information on studies Time budget Financing of studies and employment Living situation Need for counselling and information Study-related stays abroad
<b>Survey Design</b>	Cross-sectional survey (trend design)
<b>Population</b>	Students with German citizenship and nonmobile foreign students that have been enrolled in a state or state-recognized higher education institution in the summer semester of 2009 (not including public administration colleges, schools for distance learning and universities of the federal armed forces)
<b>Sample</b>	Simple (disproportional) stratified random sample
<b>Survey Method</b>	Standardised self-administered survey
<b>Survey Time Period</b>	May 1, 2009 to July 31, 2009
<b>Number of Cases (Data Set)</b>	n = 16,370
<b>Response Rate</b>	32.0 %
<b>Data Products and Access Methods</b>	CUF: Download SUF: Download, remote desktop, on-site
<b>DOI</b>	10.21249/DZHW:ssy19:1.0.0
<b>Further Information</b>	<a href="https://fdz.dzhw.eu">https://fdz.dzhw.eu</a>

---

**Project Publications\***

Isserstedt, Wolfgang; Middendorff, Elke; Kandulla, Maren; Borchert, Lars; Leszczensky, Michael (2010): Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2009. 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem. Edited by Bundesministerium für Bildung und Forschung (BMBF). Bonn, Berlin.

(Isserstedt, Kandulla 2010)

\* All project publications are available for download on the project website ([www.sozialerhebung.de](http://www.sozialerhebung.de)).

---

**Publications using the Data Set (selected)**

Sarcelletti, Andreas (2015): Bachelor students' transition to postgraduate studies. Do students with and without migration background have different plans? In *Beiträge zur Hochschulforschung* 37 (2), pp. 116–139

Thomsen, Stephan L. (2014): Did tuition fees in Germany constrain students' budgets? New evidence from a natural experiment. Forschungsinstitut zur Zukunft der Arbeit. Bonn (IZA Discussion Paper, 8623)

Kulik, Martina; Haaren, Friederike von; Thomsen, Stephan L. (2013): Studiengebühren: Kein Effekt auf das Ausgabeverhalten der Studierenden. Niedersächsisches Institut für Wirtschaftsforschung e. V. (NIW). Hannover (niw info spezial, 1)

---

### III Data Use Instructions

**[Data Use Requirements]** Data from the 19th Social Survey are anonymised and made available by the RDC-DZHW in accordance with the Federal Data Protection Law (cf. § 40 para. 1 and 2 BDSG) exclusively for scientific research purposes.<sup>8</sup> The RDC provides a *Scientific Use File* (SUF) for secondary academic use and a *Campus Use File* (CUF) for teaching and training purposes.

A requirement for use of the SUF is employment at a scientific institution. Students or doctoral students without such employment must be able to prove cooperation with a supervisory employee of a scientific institution. In addition, the RDC must confirm the presence of a scientific user's interest. A form for the data use contract can be downloaded from the RDC website. For use of the CUF only registration on the RDC website and agreement with the CUF user terms and conditions is required, after which the CUF will be transmitted by the RDC-DZHW.

**[Data Access]** The CUF of the 19<sup>th</sup> Social Survey can be downloaded from the RDC website and used at the local computer. The SUF is provided using three modes of access, which differ in their restrictions with respect to storage location, the opportunity for autonomous access to external data and RDC control options for restrictive data. These methods include:

- **Download:** Data are available for download – and protected against unauthorised access – on the RDC website. Users can save the data on their local computer to link with data from external sources as well as perform analysis using their own software.
- **Remote Desktop:** Data are available on a RDC terminal server. Using a secure connection between the user's local computer and the RDC terminal server, the data can be analysed using the software on the terminal server. The transfer of data to the local computer is not possible. Analysis results are made available only after a data protection clearance test by the RDC.
- **On-Site:** Data are made available for analysis at a secure computer on RDC premises and in a controlled environment (e.g. with video surveillance). As with remote desktop access the analysis results are made available only after a data protection clearance test by the RDC.

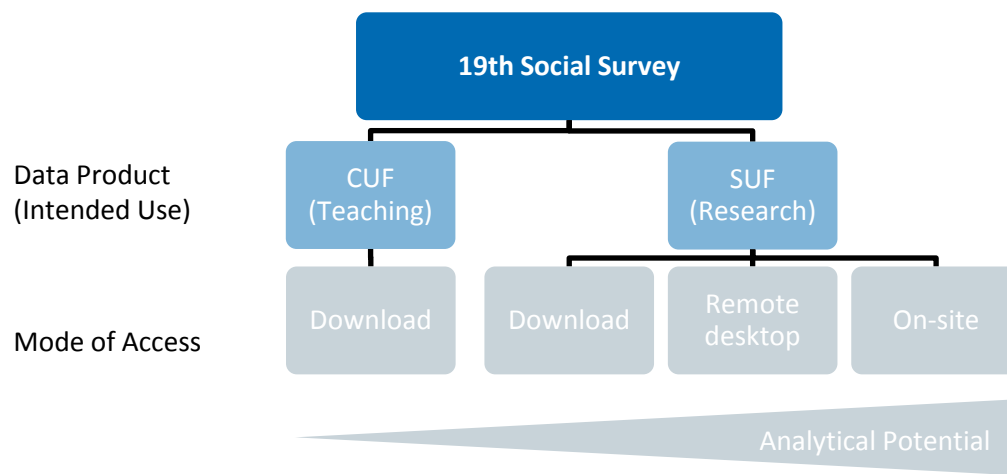
The extent of information access from the data made available differs according to the mode of access, which further impacts analytical potential (cf. Figure 1). More detailed information is made available for data users in accordance with the degree of restrictions governing the user's data access through technical and organisational measures.<sup>9</sup> Such procedures ensure the highest degree of usability, and simultaneously, the best possible data protection.

<sup>8</sup> The data protection concept of the RDC is based on the portfolio approach of Lane et al. 2008, p. 6, which the Leibniz Institute for Educational Trajectories (LifBi)(cf. Koberg 2016, pp. 699) and the RDC of the German Federal Employment Agency at the Institute for Employment Research (see Hochfellner et al. 2012, p. 9) already use as a basis. The RDC-DZHW has adapted this approach to the requirements of the particular databases and uses four categories of measures to ensure data protection, which can be combined in various ways: legal/institutional measures, informational measures, technical measures and statistical measures. For more detailed information, see the RDC data protection concept on the website.

<sup>9</sup> Cf. Chapter 8 on the various levels of anonymisation and analytical potential of the CUF and the differing SUF variants. Further information on data access and data use are available on the RDC website.



Figure 1: Modes of Access and Analytical Potential



**[Available Data]** You can access a website with key information on the study, other documentation material and an overview of the available data products via the *Digital Object Identifier* (DOI) 10.21249/DZHW:ssy19:1.0.0.

The data for the 19th Social Survey is available using any of the modes of access offered by the RDC-DZHW – each with access-specific analytical potential (see Figure 1). Data are available respectively in Stata and SPSS format using the download mode of access option. For Remote Desktop and On-Site modes of access, by default data sets are available in Stata format. Possible use of further formats or analysis software can be found on the website.

**[Charges for Data Access]** Currently SUF and CUF are available free of charge (effective June 2017). The present fees regulation can be found on the project website (<https://fdz.dzhw.eu>).

**[Responsibilities of Data Users]** Data users are required to observe the following rules<sup>10</sup>:

- **Scientific Use:** Data must be used exclusively for scientific research purposes. Commercial use is forbidden.
- **De-anonymisation forbidden:** Any attempt of re-identification for the units of analysis (e.g. persons, households, institutions) is prohibited.
- **Duty to report security loopholes:** If data users become aware of security loopholes with respect to data protection or data security, the RDC-DZHW should be informed immediately.
- **No data disclosure:** SUF may only be used by persons who have made a data use contract. CUF may only be disclosed in the context of specified teaching activities.
- **Duty to delete:** SUF downloads must be deleted after expiry of the agreed period of use (as a rule three years) from all computers, servers and data storage devices. Likewise all backup copies, modified data sets (e.g. work-, excerpt- or help-data) as well as print-outs must be destroyed.

<sup>10</sup> The data use contract regulates terms and conditions of use in detail.

- **Notification of Publications:** All types of publications that are produced using data made available by the RDC-DZHW are to be made accessible to the RDC in electronic form. A list of existing publications based on the data can be found in the Metadata Search Portal.<sup>11</sup>
- **Citation rules:** The data set used must be cited according to the following requirements in publications, other essays (e.g. final dissertations) and presentations.

#### [Citation requirements]

- **Data Set:**  
Isserstedt, W., Middendorff, E., Kandulla, M., Borchert, L. & Leszczensky, M. (2010): *19th Social Survey*. Edited by Hofstätter, U. & Sarcletti, A., doi: 10.21249/DZHW:ssy19:1.0.0, DATA SET NAME<sup>12</sup>, released 2017. Hannover: RDC-DZHW.
- **Data and Methods Report:**  
Hofstätter, U. & Sarcletti, A. (2016): *19th Social Survey: Data and Method Report on the Social Survey 2009. Version 1.0.0*. Hannover: RDC-DZHW.

In addition, the data source must be acknowledged in the text using the following formulation:

“This scientific work uses data of the 19th Social Survey (2009), conducted by the German Centre for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data was published by the Research Data Centre of the DZHW, doi:10.21249/DZHW:ssy19:1.0.0.”

---

<sup>11</sup> <https://metadata.fdz.dzhw.eu>

<sup>12</sup> Please provide the exact name of the dataset version used at this point, e.g. ssy19\_g1\_d\_1\_0\_0 for the downloadable German language SUF of the 19th Social Survey (version 1.0.0).

## 1 Content and Structure of the Study

**[Survey Series]** Since 1951<sup>13</sup> the German National Association for Student Affairs (DSW) has been conducting the Social Survey, in which students are asked standard questions about their studies, mostly at three-year intervals. The Social Survey is thus Germany's longest existing social-scientific study in the area of student research. The focus is always on the social and economic situation of the students. The students in each sampling are only interviewed once, i.e. the process takes the form of a cross-section survey.

Since the 13th Social Survey (1991), the population consists of students who are enrolled at a state or state-recognised higher education institution in the Federal Republic of Germany in the summer semester of each particular survey year. Students at colleges of public administration, schools for distance learning and universities of the German Armed Forces are not part of this population.

The first six Social Surveys (1951 to 1963) were full surveys (excluding foreign students on leave). Since the 7th Social Survey, only a sample has been extracted for each survey. The surveys have been conducted by the DZHW since the 10th Social Survey (1982). Since then, 'Bildungsausländer', i.e. international students with a higher education entrance qualification not obtained in Germany, have also been included in the population (cf. Middendorff 2016, p. 5). The Social Survey was previously limited to students of German nationality and Bildungsinländer<sup>14</sup>.<sup>15</sup> Since the 17th Social Survey (2003), the sampling of 'Bildungsausländer' has been extracted as a separate sample and the 'Bildungsausländer' have been interviewed ever since using a bilingual (German and English) questionnaire (cf. Middendorff 2016, p. 5).

**[Sample of the 19th Social Survey]** For the 19th Social Survey described here, a random sample from the groups of students with German nationality and 'Bildungsinländer' (non-mobile foreign students) was extracted, as well as a separate sample from the group containing 'Bildungsausländer'. Due to the difference in survey tools and student groups being examined, only data from the former sample has been taken into account in the hereby documented SUF (and CUF).<sup>16</sup>

The Social Survey was conducted solely using a paper questionnaire up to and including the 17th Social Survey (2003). In the 18th, and the 19th Social Survey presented here, as well as the 20th Social Survey which has already been conducted, part of the sample was used to interview the students online within the scope of a method test<sup>17</sup> (cf. Middendorff 2016, p. 12). The data in this subsample from the 19th Social Survey solely served methodological research purposes. They formed no basis for the report on the project and are not part of the SUF and CUF documented here. The SUF and CUF documented here from the 19th Social Sur-

<sup>13</sup> For a detailed overview of the history of the Social Survey see Middendorff 2016.

<sup>14</sup> 'Bildungsinländer' are students with foreign nationality, who have gained their higher education entrance qualification in Germany.

<sup>15</sup> In the 6th Social Survey, the first attempt was made to include foreign students in the study. Due to the lack of data quality, however, the data could not be analysed and the interviewing of foreign students was suspended again, up to and including the 9th Social Survey (1979) (cf. Middendorff 2016, p. 5).

<sup>16</sup> In the sample of students with German nationality and 'Bildungsinländer', one in every 27 students who matriculated in the summer semester of 2009 was included. For the sample of 'Bildungsausländer', one in every 18 students who matriculated in the summer semester of 2009 was included.

<sup>17</sup> As method tests have shown that several advantages are to be gained by converting to an online survey (cf. Middendorff 2016, p. 12), the Social Survey is to be conducted exclusively as an online survey from 2016 onwards (21st Social Survey).

vey therefore only include data on the 'Bildungsinländer' and students with German nationality who were surveyed using a paper questionnaire.

**[Analytical Potential]** The key data of the Social Survey include questions on higher education institution entrance, structural characteristics of studying and courses of study, social and economic situation (financing of studies, living costs, employment, living situation), as well as on themes relating to the 'Studentenwerke' (Student Services) and sociodemographic aspects, which are surveyed in addition to the official statistics (marital status, parenthood, social origin, migration background). Time series analyses are possible on the key element themes. The emphasis on topics surrounding the social and economic situation of the students was made based on the assumption that, along with the conditions for higher education admission and the studying conditions, the social and economic conditions are also of major importance for the progress of studies and success (cf. Apolinarski et al. 2014, p. 1). Since the 10th Social Survey, one or two additional focus areas have been included in every survey. This involves themes such as studying with a child and studying with a disability or chronic illness. The data from the Social Survey are also suitable for analyses on the time required for studying and employment. Analyses on the receipt of BAföG (student grant), student employment and study-related stays abroad are also possible. To a lesser extent, the data can also be used to perform analyses on students with children and students with a migration background.

**[Research Field]** The Social Survey is a study which is unique to student research for several reasons. Firstly, the existing time series since 1951 should be mentioned here, so that the data from the Social Survey can be used to examine questions surrounding the key elements of the study over a very long period of time. Furthermore, the survey is a unique national random sample of students, which is designed for drawing sophisticated conclusions about students in all semesters, the most diverse degree programmes, different higher education institution entrance requirements and – with the exception of colleges of public administration, schools for distance learning and universities of the German Armed Forces – all types of higher education institutions.

**[Particularities of the 19th Social Survey]** At the time of the 19th Social Survey, the structural reform to studies within the context of the Bologna Process was so far advanced that, for the first time, the sample included a significant amount (47 percent) of students from the 'new' study programmes (bachelor's degree, master's degree). In the set of data from the 18th Social Survey (2006), the percentage of students on bachelor's and master's degree programmes amounted to 13 percent (cf. Isserstedt et al. 2010, p. 6). Due to the high number of students participating in bachelor's degree programmes, the data from the 19th Social Survey (in contrast to the 18th Social Survey) can also be used to make well-founded statements about the intentions of this group to embark on master's degree courses. Another characteristic of the 19th Social Survey is the fact that around 60 percent of the students in the data set were studying at an institute of higher education that was increasing general tuition fees or contributions.<sup>18</sup> In 2009 tuition fees or study fees were raised in several western German federal states. The introduction of tuition fees or contributions in several federal states and the abolition of these in other states can be perceived as a natural experiment. It is therefore possible to perform various analyses on the effects of the introduction of tuition fees and

---

<sup>18</sup> Around one fifth of the students attending higher education institutions with general tuition fees or student contributions were exempt from paying fees/contributions (source: data from the 19th Social Survey).

student contributions. As the general tuition fees were later reabolished in all federal states, the 19th Social Survey is the only one in which a significant portion of students were affected by tuition fees.

## 2 Survey Instruments

A standardised paper questionnaire in German was used as a survey tool for the 19th Social Survey documented here.<sup>19</sup> Chapter 2.1 presents the key content from the survey tool. Chapter 2.2 describes the pre-test that was carried out for the review and improvement of the questionnaire.

### 2.1 Contents of the Survey Instruments

**[Characteristics of the Survey Series]** Just as with other survey series, the social and economic situation of the students is a focal point of the 19th Social Survey. The key themes are thereby structural characteristics of the degree courses, the financial situation of the students, their social origin and demographic information. Questions are posed on these key themes in all of the surveys. In addition, one or several focus areas (e.g. computer and internet usage, advice and information requirement) are usually covered.

The questions regarding the structural characteristics of the degree courses gather information on the current course of study (e.g. subject, envisaged completion date, higher education institution and course semesters), the educational biography (e.g. already completed higher education degrees, breaks in study, changes of subject or higher education institution; in the 19th Social Survey: questions 1 to 11) and information on the time required for studying and employment (question 13). Study-related stays abroad and their financing, obstacles in relation to study-related stays abroad (questions 45.1 to 47) and language skills (question 48) are also covered. This information is supplemented by individual perceptions of the value of studying and of the subjectively perceived workload of the course of studies (questions 12 and 14).

The financial situation of the students is the focus of every social survey. Information is recorded on the cost of living and tuition fees/contributions (questions 20 and 21). Precise information is also gathered on the level of income from personal employment, family/partner support, stipends and other sources (question 19). Detailed information on grants administered through BAföG (Federal Training Assistance Act) is also collected (questions 23 to 25). These amounts are supplemented by the subjective estimation of the financial situation (question 22) and the motives for personal employment (question 26.4).

Furthermore, the various sociodemographic (questions 27 to 33, 40 and 41) and educational-biographical characteristics (questions 15 to 18) as well as the family backgrounds of the students (questions 34 to 39) are surveyed.

**[Particularities of the 19th Social Survey]** A key issue, which is also dealt with in the questionnaires of other surveys of the series, is the students' requirement for advice and information (question 44). Information needs and the use of advisory services on various topic areas (e.g. student financing, health problems) are surveyed here in detail. Various changes to the institutional conditions were also taken into account within the survey tool of the 19th Social Survey. Against the background of the Bologna Process, information was gathered on present or intended courses of study on Masters programmes (questions 5.1 to 6). Another

---

<sup>19</sup> The questionnaire can be downloaded from the RDC website. You can also find a flow chart there for the filtering of the questionnaire.

example of current education policy-related questions is the complex issue of tuition fees (questions 21.1 to 21.3), which was dealt with in greater detail than in questionnaires from previous Social Surveys. In contrast to past Social Surveys, the 19th Social Survey asked about the nationality of the parents, thereby improving the likelihood of capturing information on the migration background (question 38). In relation to the key themes, the 19th Social Survey offers comparisons with previous social surveys. To a limited extent, there are also comparisons to be made with regards to the recurring key issue of the need for counselling and information, though only with the data in the 16th and 18th Social Surveys.

## 2.2 Pre-tests

**[Goal and Procedure]** The survey tool was tested in advance of the survey using a cognitive pre-test. The aim of this was firstly to test whether the multiple-choice questions already used by survey tools in previous social surveys would be perceived the same by the students of 2009 as was the case for the students in past social surveys. Secondly, the comprehensibility and likelihood of response for the newly used or modified questions needed to be tested, and the way students responded to certain questions (e.g. time budget, income, expenses) established. The inclusion of the new questions also led to changes in the structure and layout of the questionnaire as well as in the duration of the survey, which needed to be evaluated. A so-called 'field pre-test procedure' was used to examine these different aspects. The aim of this procedure is to ensure that persons participating in the pre-test are "monitored under conditions as similar as possible to those participating in the the actual survey" (Häder 2015, p. 396).

**[Test subjects]** 25 students with German nationality and 'Bildungsinländer' were recruited to take part in the test via public notice boards at the local university (Leibnitz University Hannover).

**[Implementation]** The pre-test took place in January 2009, four months before the start of the field survey. The test persons were asked to provide responses to the standardised questionnaire that was intended for the survey. Then group discussions were held, each with four to eight test subjects and employees from the Social Survey research group using a discussion guide. The focus of these discussions was on the new and modified questions, whereby the students had the opportunity to express their view on every question. In addition, there were questions surrounding the completion time, content and length of the questionnaire, the structure and layout, the comprehensibility of the questions, completion instructions and the completeness of the response options. On the basis of the pre-test results, the wording of various questions was refined and the order of items and response categories was revised in particular item sets and multiple response sets. The basic structure, length and layout of the questionnaires remained unchanged. The test subjects received an expense allowance of 20 euros.

## 3 Population and Sample Procedure

**[Population]** The population of the 19th Social Survey includes students with German nationality and 'Bildungsinländer'<sup>20</sup>, who were enrolled at a state or state-recognised institution of higher education in the Federal Republic of Germany in the summer semester of 2009. Students at colleges of public administration, distance learning universities and students at universities of the German Armed Forces were thereby excluded, as their study situation is fundamentally different from that of other students.<sup>21</sup> Furthermore, due to reasons of timing, no students from institutions of higher education that were only established in 2009 were taken into account (Isserstedt et al., 2010, p. 39).

**[Sampling Procedure]** Due to non-accessible student lists, it was impossible to extract a basic random sample. A random sampling of all institutions of higher education was devised, at which students belonging to the target population were enrolled, as these could only be contacted via the institutions themselves. A random sample with the selection probability of 1/27 (i.e. 1 in 27 students enrolled in the summer semester of 2009) was taken at the relevant universities from the group of students with German nationality and 'Bildungsinländer'.<sup>22</sup> At particular universities, a larger sample (i.e. with a higher selection probability) was drawn at the request of the Studentenwerke to enable evaluations also to be performed at Studentenwerk level (disproportional sample).<sup>23</sup> The disproportional sample is balanced with a corresponding design weighting (see chapter 0).

---

<sup>20</sup> 'Bildungsinländer' are students with foreign nationality, who have gained their higher education entrance qualification in Germany.

<sup>21</sup> Students at distance learning universities are very often in employment and complete their degree courses alongside this. Students at colleges of public administration are normally civil service trainees ('Beamte auf Widerruf'). Students at universities of the German Armed Forces are normally officer candidates.

<sup>22</sup> For the purposes of the internal, methodological development of the study, 20 percent of the sample students were given a link to the online version of the survey instead of the paper questionnaire. These cases are not part of the SUF and are therefore not taken into account in this documentation (see chapter 1).

<sup>23</sup> The increased samplings were carried out at the following 15 institutions: University of Giessen, Mittelhessen University of Applied Sciences and Fulda University of Applied Sciences (both 1 in 24), University of Göttingen (1 in 17), Technical University of Munich and University of Tübingen (both 1 in 16), University of Hannover (1 in 15), University of Marburg and University of Bremen (both 1 in 13), University of Kassel (1 in 12), Technical University of Darmstadt and University of Wuppertal (both 1 in 11), Munich University of Applied Sciences (1 in 10), Technical University of Braunschweig (1 in 9) and Darmstadt University of Applied Sciences (1 in 8).



## 4 Implementation of the Survey

**[Maintenance of Contacts and Addresses]** All institutions with students from the target population were contacted by the DZHW in writing and asked to take part in the 19th Social Survey. Attention was drawn to the relevance of the study and its history in a covering letter. The DZHW also communicated the criteria to the higher education institutions, which they could use to identify the target persons for the 19th Social Survey.<sup>24</sup> 218 out of 303 institutions addressed (71.9 %) agreed to participate.<sup>25</sup> As the higher education institutions are not permitted to give out the contact details of their students for data protection reasons, the survey documents were sent out by the higher education institutions themselves in the course of the social surveys. For this purpose, the DZHW determined the number of required survey documents in each case<sup>26</sup> and sent these to the participating higher education institutions, who obtained the students' addresses from their student directories and printed out two address labels for each one. One label was required for the envelope with the survey documents, a second for the reminder postcard. The higher education institutions then posted both pieces of correspondence to the target persons within specified time frames.

**[Survey Documents]** The survey documents for each person being questioned consisted of a covering letter from the German National Association for Student Affairs as the contracting authority for the study, the questionnaire, a flyer with general information on the Social Survey and sample results from previous surveys, and a prepaid envelope addressed to the DZHW for the return of the completed questionnaire. In addition, all students from the drawn sample received a time-delayed reminder postcard.

**[Fieldwork Phase]** The survey period lasted from 11 May until the end of July 2009. Due to the method of contact used via the higher education institutions, the DZHW was unable to have any direct influence on the precise delivery time of the survey documents. The reminder postcards were sent out to the students approximately two weeks after the field start. They also received a thank you note, which was sent to all people in the random sample – including those who had already taken part in the survey –, as the higher education institutions were unaware which persons had already returned a questionnaire to the DZHW.

**[Measures to Increase Response]** Along with the reminder postcard and the flyer accompanying the covering letter, the 19th Social Survey was also advertised through press and public relations activities by the German National Association for Student Affairs and the individual Studentenwerke. It was mentioned in press releases and poster advertising for the forthcoming survey. Moreover, the project website was available (<http://www.sozialerhebung.de>) as a source of information. It contained the questionnaires, photos and blocks of text for press releases as well as other information on the current study and past surveys.<sup>27</sup>

<sup>24</sup> Normally (see chapter 3) 1 in 27 students with German nationality or 'Bildungsinländer' enrolled in the summer term.

<sup>25</sup> A disproportionate number of smaller, art and music colleges and private higher education institutions gave no positive feedback. In relation to the number of students, however, almost 95 per cent of them are enrolled at the 218 (out of 303) higher education institutions who agreed to take part in the test.

<sup>26</sup> To this end, the figures contained in the official statistics on students enrolled at the higher education institutions in the previous semester (winter semester 2008/09) were combined with the respective selection probability (see chapter 3).

<sup>27</sup> The materials for the 19th Social Survey can currently be found at <http://www.sozialerhebung.de/sozialerhebung/archiv> (As of: 01.12.2016).

## 5 Response Rate

**[Response Rate]** Approx. 53,008 students were contacted by the higher education institutions. Of these approx. 53,008 students, approx. 51,143 (gross sample<sup>28</sup>) received the questionnaire.<sup>29</sup> 16,541 questionnaires were returned to the DZHW in total. Of these, 171 were eliminated as they either did not belong to the target population (non-responses neutral to the random sample) or were non-evaluable (non-responses relevant to the random sample) (see chapter 6.3). 16,370 cases consequently remained in the net sample. With regard to the gross sample, the response rate was around 32 per cent (see Table 1).

**Table 1: Gross and Net Samples and Response Rates of the 19th Social Survey (2009)**

Gross sample	51,143
Net sample	16,370
Response rate	32.0 %

**Figure 2: Responses to the 19th Social Survey (2009) Over Time**

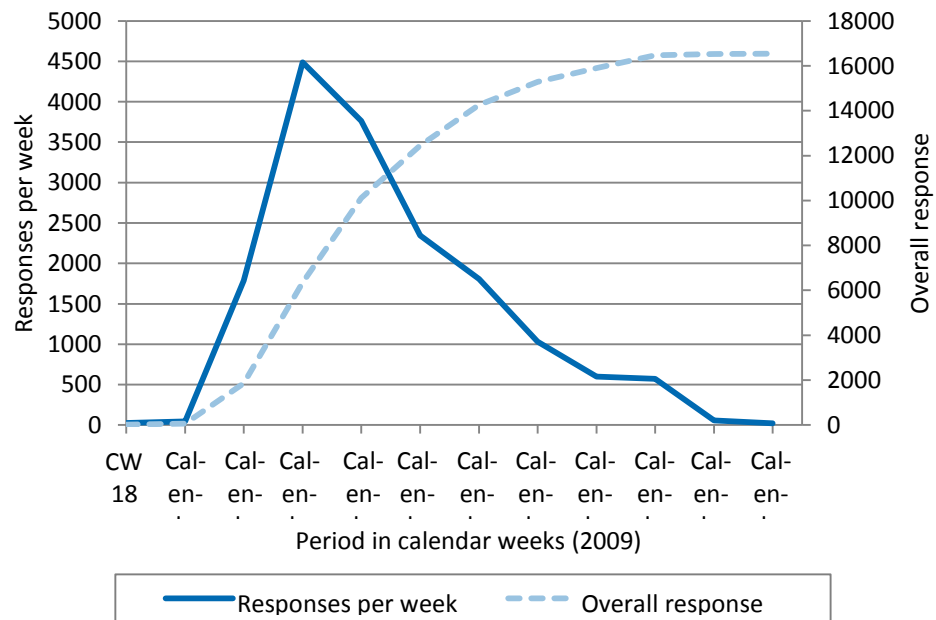


Figure 2 shows the responses to the questionnaires over the course of time. The entire response phase covered a period of twelve weeks. A large proportion of the completed questionnaires reached the DZHW in the first half of the field phase, during which the reminder postcards were also sent out.

<sup>28</sup> As an exact differentiation between the initial (unadjusted) gross sample and the adjusted gross sample is not possible, this has been limited to the description 'gross sample' in the text below.

<sup>29</sup> Due to missing or outdated postal addresses, some students could not be reached. A request was made by the higher education institutions to the DZHW for feedback on the number of undeliverable letters and the reasons for these. But this information was rarely forthcoming. The number of undeliverable mail items was recorded by 45 higher education institutions. This varied between 0 per cent and 14 per cent, with an average of 2.5 per cent.

## 6 Data Preparation

In the following sections, various steps in data preparation are described. These took place during the first and second survey waves and are identical in each wave. The procedures described in Chapters 6.1 to 6.3 had already been conducted by the primary research project. The generation of variables (chapter 6.4) was carried out by the primary project as well as the RDC during data preparation. Procedures described in Chapters 0 to 6.7 were carried out by the RDC building on the work of the primary research project. Additional procedures (e.g. weighting and anonymisation) are explained separately in Chapters 0 and 0.

### 6.1 Data Transfer

For further processing, respondent data retrieved from the paper questionnaires was transferred to a computer readable format using a code plan. Prior to this, numerical codes for the open responses were recorded on each of the paper questionnaires (see Chapter 6.2) and preliminary manual corrections were undertaken to facilitate data transfer (see Chapter 6.3).

**[Production of a Code Plan]** A code plan was produced based on the survey questionnaire. It recorded to which question or sub-question a variable is assigned, the name of the variable and the numerical codes used for the standardised answers of the respondents. To establish the order of data entry, the variables were additionally numbered.<sup>30</sup>

**[Data Entry]** For data transfer, the code plan, further instructions on data entry and the prepared paper questionnaires were given to an external service provider. Their typists manually performed the compilation of the data. The quality of the data collection was checked at random. Within the scope of the data check and data cleansing (see section 6.3), data-entry errors were detected, which were created by the skipping of variables, mostly through value range violations, and were corrected accordingly.

### 6.2 Coding of Open Responses

Before the data transfer, the (semi-) open responses were coded. Using a coding list, numerical codes were assigned to them. For each variable, various code lists were used. This was done using classification keys for official statistics (e.g. German Classification of Occupations, key lists of student and examination statistics etc.) or project-specific keys already used in prior Social Surveys. For some variables, new code lists were developed on the basis of the entries from the 19<sup>th</sup> Social Survey. For some semi-open questions, no new variables with numerical codes were created. Instead, entries were only assigned to the existing (closed) response categories. Some open questions were not coded at all. This was the case if

- they were not required for the analyses performed within the context of the 19th Social Survey.
- the only relevant issue was whether an entry was made at all.
- the questions could be roughly replaced by closed questions.

<sup>30</sup> Data were generated in a simple, column-oriented text format without a heading containing the variable names. The code plan therefore established in which order the data were to be generated so that the data belonging to a variable could be entered in the correct column.

- an encoding was not possible due to a lack of staff and time resources.<sup>31</sup>

Coding choices by the primary research project were not modified.

Coded topics and respective code lists are presented in Table 2. The data set contains solely the coded numerical variables. The open entries themselves are not contained in the data set. The values of the variables are recorded in the Data Set Report.

**Table 2: Coded Topics and Code Lists Used in the 19th Social Survey (2009)**

Topics	Coding scheme	Code List ID
Subject	Destatis Key List of Student and Examination Statistics (Winter Semester 2008/2009 and Summer Semester 2009)	cl-destatis-studienfach-2009
Degree	Project's Own Coding	cl-dzhw-14
Higher education institution	Destatis Key List of Student and Examination Statistics (Winter Semester 2008/2009 and Summer Semester 2009)	cl-destatis-hochschule-2009
Federal state	Destatis Federal State Codebook (corresponds to both first numbers of the Official Municipality Codebook, AGS)	cl-destatis-bundesland-1990
Foreign country / nationality	Destatis Nationality and Region Codebook (status: 01.01.2009) (with adjustments)	cl-destatis-ausland-2009
Reason for exemption from tuition fees	Project's Own Coding	cl-dzhw-21
Scholarship provider	Project's Own Coding	cl-dzhw-10
Other open enquiries	Assignment to given categories or project's own coding	

### 6.3 Data Checking and Data Cleansing

**[Preliminary Manual Correction]** Before the transfer of data, a preliminary manual correction was performed on the questionnaires, and if necessary, responses were adjusted.<sup>32</sup> This was intended to facilitate the capturing of data. The form, but not the content of the submitted data, was amended. For example, hardly legible data or crossings out made by the respondents were highlighted.

On the other hand, the manual inspection was also aimed at correcting initial errors or inconsistencies in the respondents' data before the software-supported correction (see below).

<sup>31</sup> The open data on the questions listed below was gathered electronically. But no coding was applied and it was simply recorded in the variable whether an open entry was made (named) or not (unnamed):

- Question 21.3 (other financial sources and tuition fees),
- Question 25 (other reason for no BAföG [application]),
- Question 26 (other activity to earn money), question 44.1 and question 44.2 (other issue, need for and use of advice) and
- Question 37 (parents' professions).

As an alternative to the parents' professions, the employment status (question 34), school (question 35) and professional qualifications (question 36) and the professional status (question 39) of the parents may be used. For study-related periods of stay abroad (question 45), no record whatsoever is made of the type of study-related foreign stay (45.1) and the other sources of finance for this (45.4) are only available in the form of 'named' versus 'unnamed'.

<sup>32</sup> The number of corrections was not recorded centrally, but simply on the paper questionnaires, and can therefore no longer be systematically reconstructed.

Various consistency rules were established in advance for this purpose, which could be checked on the paper questionnaires. If any inconsistencies were identified, the data concerned was checked for plausibility, where possible, and any missing data reconstructed from other entries in the questionnaire. Checks for content consistency were also performed. For example, a comparison of the data on the already acquired higher education degrees (question 3) with the legal status of the visited educational institutions took place.<sup>33</sup> There was also a comparison of the employment period between the time description itself and the proper classification of the activity in the corresponding period (question 26). Any identified inconsistencies were, if possible, eliminated by the comparison with other responses in the questionnaire or alternatively by assigning a corresponding missing code (see chapter 6.7).

**[Software-Assisted Correction]** Following data transference, a comprehensive test and correction of the data with the help of DZHW's own software was performed. Software-Assisted Correction was used to catch mistakes in the preliminary manual correction and data transfer. Furthermore, inconsistent responses of the respondents that could not be tested in the preliminary manual correction were identified.

To this end, the compiled questionnaire data was read in to a data bank. In addition, valid value ranges and answer combinations were defined and tested systematically. The following types of tests were carried out:

- *Test of Value Ranges:* It was tested whether the response lay in the value range defined of the respective recorded variable.
- *Test of Adherence to Filter Procedures:* Based on the defined filter procedure of the questionnaire, it was tested whether responses that would have been expected from the respondent were not (i.e. completeness test) and whether responses were made that should not have been (i.e. filter errors).<sup>34</sup>
- *Test of Variable Consistency:* The consistency of responses within a questionnaire as well as between survey waves was tested. In addition to feature combinations, which were already tested in the preliminary manual correction, more complex feature combinations could also be tested here.<sup>35</sup>

In total, hundreds of consistency rules were defined and tested. Missing, incorrect or implausible values were first tested using the paper questionnaire to determine whether the corresponding value had been falsely (or not at all) transferred. Then the correct value was inferred using other responses in the questionnaire. In case of doubt, a specific missing code was assigned (see chapter 6.7). Corrections of mistakes were documented<sup>36</sup> and checked by at least one further person.

**[Deletion of Cases]** A case was deleted if half of the questions or core questions (e.g. on course of study) were not answered or if too many inconsistencies were present. These cases were graded as not possible to evaluate. In addition, cases were identified as not belonging to

<sup>33</sup> If the respondent indicated that a higher education degree had already been obtained, this was not recorded as the obtaining of a university degree, as the vocational academies only achieved university status with the conversion into the Baden-Wuerttemberg Cooperative State University on 1 March 2009.

<sup>34</sup> The input filters for each question were recorded in the dataset report. They define for each variable, which group of respondents should respond to the related question, provided this has not been addressed to every person.

<sup>35</sup> The income and expenses and the time budget were examined in particular detail.

<sup>36</sup> The documentation of error corrections was performed in writing on the paper questionnaires and can therefore not be systematically reconstructed.

the target population.<sup>37</sup> These were likewise removed from the data set. A total of 171 cases were deleted.<sup>38</sup>

### 6.4 Generation of Variables

In addition to the variables that are generated from the coded answers of the respondents, the 19th Social Survey also generates new variables. This includes variables that were numerically coded from the originally open entries (see chapter 6.2). More frequently required variables were generated from the values of one or more sources (e.g. merging course subjects in to areas of study and subject groups or deriving the location and type of the higher education institution from the higher education institution variables). A large proportion of the generated variables were already created by the primary research group. ). The newly generated variable is identified in the data by the suffix “\_g#”. An overview of all generated variables for the 19th Social Survey as well as detailed documentation of the individual variables with information on their respective characteristics and calculation rules can be found in the Data Set Report.

Where possible, generated variables were positioned in the data set according to the respective output variable. If the output variable is no longer in the data set due to anonymisation (see Chapter 8), the generated variable occupies its place in the data set. If a variable was generated from various source variables, it was inserted after the variable to which it is thematically closest. If a clear assignment was not possible, the generated variable was inserted at the end of the data set. With minor exceptions, variables generated in the context of anonymisation measures were created by the RDC.

### 6.5 Generation of the Data Set

**[Data Structure and File Format]** The data set contains the survey data and the additional generated variables. The sequence of the variables is based on the order of the related questions in the questionnaire. The data set is available in both Stata and SPSS format (see Section III).

### 6.6 Assignment of Variable Names, Variable Labels and Value Labels

**[Variable and Value Label Assignment]** For variable and value label assignment, formulations from the questionnaire were used, or in some instances, concise formulations were chosen. As a rule, the variable labels are based on the corresponding question. Depending on the type of question, value label assignments are based on the response options or a combination of the question and response options. For generated variables based on definite classifications, value labels were adopted verbatim from the classification keys. Variable and value labels are

---

<sup>37</sup> This happened, for example, if the higher education institutions had accidentally contacted someone who could neither provide supporting evidence of German citizenship nor had gained their higher education entrance qualification in Germany. This group, the ‘Bildungsausländer’, was surveyed using a special questionnaire and a separate sample was drawn (see Chapter 1 for details). The questionnaires intended for the German students or student ‘Bildungsinländer’ were eliminated if they had been accidentally answered by ‘Bildungsausländer’.

<sup>38</sup> This corresponds to 1.0 per cent of the cases. It is not possible to reconstruct how many of these non-responses are neutral to the random sample and how many are relevant to the random sample.

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

**[Naming of Variables]** A consistent naming system was created at the RDC-DZHW for the naming of variables. With the exception of the identifier variables, the variable names are formed according to a prefix-root-suffix scheme. The variable names provide meta-information on the corresponding variable. This facilitates automated processing. For longitudinal surveys with more than one survey period (panel), the variable prefix contains the wave identifier in the form of a letter. As the 19th Social Survey is a cross-sectional survey, no prefixes are used here. The subject area in the root, to which the variable is allocated, is denoted by a three-letter English abbreviation. Table 3 presents an overview of the various thematic areas of the 19th Social Survey as well as the related abbreviations for the root of the variable names. The suffix, separated from the root by an underscore, contains various additional information such as the labelling of generated variables, panel variables and different modes of data access.

Detailed information on variable assignment for the 19th Social Survey can be found in the Data Set Report.

**Table 3: Thematic Areas and Abbreviations of the 19th Social Survey (2009) Variable Labels**

Thematic Area Abbreviation	Thematic Area (English)	Thematic Area (German)
<b>stu</b>	studies	Studium
<b>tim</b>	time usage (studies/job)	Zeitaufwand (Studium/Erwerbstätigkeit)
<b>ped</b>	prior education	Vorbildung
<b>fin</b>	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
<b>baf</b>	BAföG (German Federal Grant for Training and Education Promotion)	BAföG (Bundesausbildungsförderungsgesetz)
<b>job</b>	job (during studies)	Erwerbstätigkeit (im Studium)
<b>dem</b>	demographic information	Demographische Daten
<b>par</b>	parents	Eltern
<b>liv</b>	living (accommodation)	Wohnform
<b>nut</b>	nutrition	Ernährung
<b>cin</b>	counselling and information need	Beratungs- und Informationsbedarf
<b>abr</b>	experiences abroad	Auslandserfahrungen
<b>lan</b>	language (skills)	Sprach(kenntniss)e(n)

## 6.7 Coding of Missing Values

For coding missing values, a comprehensive system was created in the RDC, in order to guarantee unified coding for missing values across various data sets of the DZHW. Missing responses were coded using three-figure negative values. Table 4 presents an overview of the system for coding missing values. The coding for missing values used in the 19th Social Survey is highlighted.

Missing values can be assigned to four different groups. First, missing values may arise if the respondent does not answer the survey questions (i.e. non-response). Second, missing values may be assigned if questions are not relevant to the respondent, given the filter procedure (i.e. not applicable). The third group contains missing values assigned through the primary research project or the RDC in the course of the data processing (i.e. edited missing value). This includes missing variables for certain variables due to anonymisation measures (see chapter 8). The fourth group comprises missing values assigned for certain items in the context of data processing. There are no such item-specific missing values in the dataset for the 19th Social Survey.

**Table 4: System of the RDC-DZHW for Missing Values**

Value range	Code	Value label
<b>-999 to -990: Non-response</b>	-999	don't know
	<b>-998</b>	<b>no response</b>
	-997	no response (response category)
	-996	Interview cancellation
	-995	no participation (panel)
	-994	rejection
<b>-989 to -970: Not applicable</b>	<b>-989</b>	<b>missing due to filter</b>
	<b>-988</b>	<b>not applicable</b>
	-987	missing due to design (questionnaire split)
	-986	missing due to design (wave) <sup>a</sup>
	-985	missing due to design (cohort) <sup>b</sup>
<b>-969 to -950: Edited missing value</b>	-969	unknown missing value <sup>c</sup>
	-968	implausible value <sup>d</sup>
	<b>-967</b>	<b>anonymised</b>
	<b>-966</b>	<b>cannot be determined<sup>e</sup></b>
	-965	invalid multiple entry
<b>-949 to -930: Item-specific missing values<sup>f</sup></b>	<i>(not assigned in the 19th Social Survey)</i>	
<b>-929 to -920: Other missing values</b>	-929	data loss

a This value is only assigned for data sets in long format.

b This value is only assigned to pooled data sets.

c This value is assigned when no cause can be reconstructed.

d Responses which are classified as implausible due to various factors in the coding phase receive this value. An exact reconstruction may no longer be possible.

e This category is assigned when clear coding is not possible, e.g. open response which could not be encoded because it is illegible.

f The characteristics of these missing categories are, by definition, specific for every data set.



## 7 Weighting

Weighting the data adjusts for distortions in the sample due to the sample design (see Chapter 3) and compared to the defined population. A general introduction to the procedure and the presentation of the final weights follows in Section 1. The weighting procedure is also described in detail in Section 2.

### 7.1 Procedure and Instructions for Use

**[Causes of Distortions in the Samples]** Two processes are relevant with regards to the distortion of random samples:

- **Distortion due to Design:** Disproportionalities are deliberately produced to increase the number of cases in certain relevant subgroups (cf. Chapter 3).
- **Distortion through non-response:** Attrition processes (e.g. non-participation, unreachable, postal error) lead to reduced response and thus to a difference between gross and net sample (cf. Chapter 5). If these processes are non-systematic (Missing Completely at Random), they can be ignored.<sup>39</sup> However, they mostly result from a systematic process (Missing at Random, Not Missing at Random), which requires modelling.<sup>40</sup>

**[Conceptual Procedure]** In the course of the weighting procedure, disproportionalities due to design should ideally be offset. The *design weights* required for this arise directly from the sample plan with random selection processes. Related to this, an adjustment of the design weights – with the help of *non-response weights* in cross-sectional and longitudinal data sets – should be produced on the basis of information on participants and non-participants. As a last step, the non-response adjusted design weights can be calibrated using distributions of characteristics from the population (*calibration*).

Due to the random sample design (see chapter 3) of the 19th Social Survey, a design weight is created in the first stage in order to compare the unequal chances of inclusion. As there is no individual information on non-participants, no non-response adjustment of the design weight can be carried out on an individual basis. The design weight is calibrated in the final stage using a distribution of characteristics in the target population. As there is aggregated information here on participants and non-participants, a form of non-response adjustment takes place at the same time.

<sup>39</sup> Insofar as the loss of statistical test strength through the reduction of the sample is considered irrelevant.

<sup>40</sup> For the various forms of failure processes see essentially Rubin, 1976.

**Table 5: Weights provided for the 19th Social Survey (2009)**

Variable Name	Description
wgt_federal gov- ernment	Weight for analyses at federal government level
wgt_region	Weight for analyses at regional level (north, south, east and west)
wgt_east_west	Weight for analyses at eastern and western German level

**[Instructions for Use of the Weights]** The generated weights are *probability weights*, which can be incorporated in Stata with the aid of .ado-specific options.<sup>41</sup> The wgt\_federal government weight is intended for evaluations at federal government level. The wgt\_region weight has weights for the regions of northern, southern, eastern and western Germany. The wgt\_east\_west weight includes weights for the regions of eastern Germany (new federal states including Berlin) and western Germany (old federal states excluding Berlin). The weights wgt\_region and wgt\_east\_west can only be used in evaluations relating to *one of* the respective regions.<sup>42</sup> It is essential to note that weights only represent meaningful correction quantities if the analysis model contains the variables used for the weighting or in relation to them. For this reason, weights must always be used with a focus on the analysed question formulation. In the following section, the procedure for producing the weights will be presented in more detail.

## 7.2 Weighting of the Data

**[Design Weighting]** As described in Chapter 3, there is an overrepresentation of students from some of the higher education institutions due to the random sample design. As a result, these individuals demonstrate a higher selection probability within the sample. This can be corrected through the use of inverse selection probability ( $\frac{1}{\pi_i}$ ). The design weight  $wgt\_d_i$  is thus:

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

Elements that are more likely to be entered in the sample are therefore given a lower weight and vice versa.

**[Calibration of Design Weights]** As described, a non-response adjustment of the design weights was not possible at an individual level. But the following demographic characteristics from the population<sup>43</sup> existed, which were able to be used for the calibration of the weights: region, sex, subject group, type of higher education institution, people with German national-

<sup>41</sup> See also the Stata guide (Command: *help weights*).

<sup>42</sup> The regions can be selected (e.g. using the Stata command *bysort*) via the variable *stu16\_g6r* (north-south-east-west aggregation for remote desktop and on-site dataset) or *stu16\_g5* (east-west aggregation for download dataset).

<sup>43</sup> All the information that was used for the calibration of the design weights is derived from the previous semester's data (winter semester 2008/2009) from the Federal Statistical Office, as the up-to-date statistics were not yet available at the time of the weighting.

ty versus 'Bildungsinländer'.<sup>44</sup> The calibration was done at both federal government level (wgt\_federal government) and separately for each of the four regions plus eastern and western Germany (wgt\_east\_west). The respective regional weights were pooled into one variable (wgt\_region). The weights for eastern and western Germany were also pooled into one variable (wgt\_east\_west).

As the characteristic bearers in the population also contained information about non-participants, the use of redressement weights also resulted in a kind of non-response adjustment with regard to the characteristics used. The calibration was performed using the raking algorithm<sup>45</sup>.

**[Trimming of the Weights]** The initially calculated weights exhibit a small proportion of weighting factors that represent dropouts. In order to remove them, all weights were subjected to a trimming according to Potter 1990 (see also Valliant et al. 2013, 388f.). The procedure is based on the assumption that the weights conform to a probability distribution (beta distribution). All those weights that lie above the 99 percent quantile are truncated to this limit. Excess on the other side of the truncation is distributed among the remaining weights.

<sup>44</sup> The weighting was performed based on the following characteristics: Sex: male versus female; Region: North (Bremen, Hamburg, Lower Saxony, Schleswig-Holstein), South (Baden-Württemberg, Bavaria), East (Berlin, Brandenburg, Mecklenburg-Vorpommern, Saxony, Saxony-Anhalt, Thüringen), West (Hessen, North Rhine-Westphalia, Rhineland, Saarland); Type of higher education institution: university versus university of applied sciences (including teacher training colleges, theological colleges, art and music colleges); Subject group: according to the Codebook for Student and Examination Statistics (winter semester 2008/2009 and summer semester 2009).

<sup>45</sup> Raking is also known as 'iterative proportional fitting' (cf. Kolenikov 2014).

## 8 Anonymisation

**[Data Protection Legal Framework]** The Federal Data Protection Act (BDSG) applies to personal data that the DZHW collected through volunteer surveys.<sup>46</sup> Accordingly, personal data that are collected during scientific research may be processed or used exclusively for the purposes of scientific research (cf. §40 para. 1 BDSG). Moreover, personal data must be anonymised (cf. §40 para. 2 BDSG) in order to protect respondents. According to the BDSG, the procedure of anonymisation is defined as changes made to survey data so that the individual data can no longer be attributed to personal or factual circumstances – or only with unreasonable amount of time, money and labour (§3 para. 6 BDSG). Regarding the disclosure of data from scientific research projects to third parties, the data must either be *absolutely anonymised* so that no reference to the person can any longer be produced, or at least *de facto anonymised* so that the construction of a reference to a person would mean a disproportionately high expenditure and thus the likelihood of re-identification of a person is minimal.

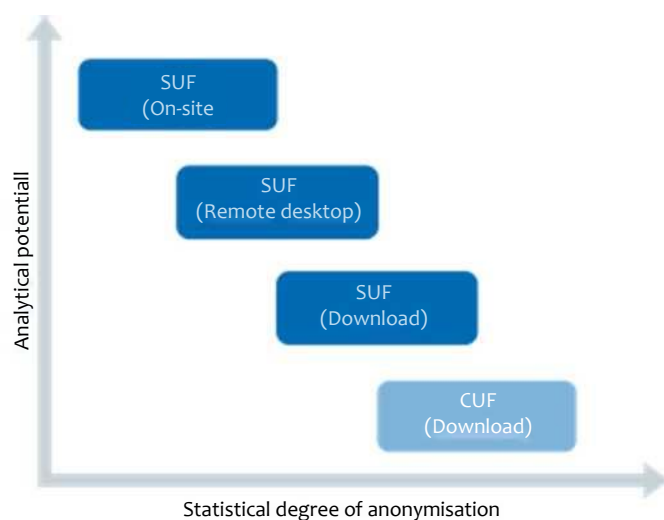
**[Data Access, Level of Anonymisation and Analytical Potential]** For the 19th Social Survey, the RDC-DZHW makes two types of data files available. Whereas SUF for scientific secondary use are effectively anonymised, CUF for teaching and exercise purposes are absolutely anonymised. The anonymity of the surveyed persons is thus protected by a combination of statistical measures and technical access barriers. The more strongly data access is technically controlled, the lower is the risk of de-anonymisation of the data, the less the data must be limited in terms of information by statistical measures and the greater their analytical potential remains.

While the CUF is directly available via download after registration, the SUF are provided using three different modes of access: download, remote desktop and on-site (for further information cf. Section III). For each mode of access a different SUF variant is made available, which is varyingly strongly anonymised and correspondingly contains less or more information. Figure 6 gives an overview of the respective level of statistical anonymisation and the related analytical potential. In the following the statistical anonymisation measures performed are explained according to data product (SUF/CUF) and mode of access.

---

<sup>46</sup> The BDSG is applicable since the DZHW GmbH is legally a public body of the federal government (cf. § 2 para. 3 BDSG). The federal government possesses an absolute majority of the shares in DZHW GmbH and the institute performs duties of public administration of the federal government in the broadest sense. For interpretation of individual legal aspects the European Data Protection Guidelines can be used as a complement.

**Figure 3: Mode of Access, Statistical Level of Anonymisation and Analytical Potential of the Data from the 19th Social Survey (2009)**



**[Statistical Anonymisation Measures]** In the course of anonymisation, all information that directly allows individuals or institutions to be identified is deleted. From these so-called *direct identifiers*, such as names, addresses or e-mail address, only the latter were collected for the purposes of the 19th Social Survey, and only if the person expressed their willingness to take part in the HISBUS Panel<sup>47</sup> on the last page of the questionnaire. The e-mail addresses were separated from the questionnaire immediately upon opening the response envelope and recorded in a separate dataset with no possibility of being back-referenced to the dataset from the 19th Social Survey. They are not included in the CUF or the various SUF versions. To further prevent any re-accessing of this information, the original identification numbers of the respondents and the (current and previous) higher education institution were removed from the dataset and replaced with a new randomly assigned identification number.

Additionally, *quasi-identifiers* were determined, i.e. information which, in combination with or by the allusion to external information, allows for indirect identification.<sup>48</sup> The following quasi-identifiers were identified for the 19th Social Survey, which are present in both external data sources<sup>49</sup> and the 19th Social Survey: name, location and type of higher education institution, subject, degree type, level of tuition fees/contributions, age and nationality. In order to prevent a definite allocation of the data from the 19th Social Survey, these key characteristics were aggregated or deleted according to data product or mode of access (see Table 6). The characteristic 'Level of tuition fees', for example, can be used unreservedly in the SUF for on-site usage. The characteristic in the remote desktop, however, is aggregated into elev-

<sup>47</sup> The HISBUS Panel is an online survey of students about important politico-educational issues, which has been regularly conducted by the DZHW since 2002. The interviewed students have willingly agreed to take part in Panel and are recruited from other DZHW surveys. More detailed information of the HISBUS Panel can be found by visiting <http://www.hisbus.de>.

<sup>48</sup> It is pointed out that the identification of a person is already made more difficult by the sample selection, since uncertainty arises whether a respondent has a unique combination of characteristics in the population.

<sup>49</sup> e.g. student and examination statistics from the Federal Statistical Office, university alumni networks or even professional networks.

en groups of 100 euro segments (0 to 100, over 100 to 200, ..., over 1,000) and into five groups in the download SUF and the CUF (0 to 250, over 250 to 500, ..., over 1,000). Open entries are also quasi-identifiers (cf. Ebel 2015, p. 3). For the large part, they were already encoded within the scope of the data processing by the primary research project and are generally<sup>50</sup> available in the CUF and in all the SUF versions. Depending on the sensitivity of the information contained therein and the mode of access, some of the codes taken from the primary research project were also aggregated. Uncoded open entries (see chapter 6.2) were deleted in the CUF and all SUF versions.

Finally it was checked whether the data contained *sensitive information*, e.g. on health, sexual orientation or political views. This information, although not suited for re-identification of individuals or institutions, can be used in case of de-anonymisation (cf. Koberg 2016, p. 694). Therefore, its protection is particularly important (cf. §3 para. 9 BDSG, Art. 8 para. 1 and 2a EG-DSRL). Health information was collected as part of the 19th Social Survey, for which no additional consent to secondary use was obtained from the respondents. These responses were therefore deleted in the CUF and all SUF versions.

To guarantee absolute anonymisation of the CUF data, more restrictive statistical anonymisation measures on the variable level in comparison to the SUF variants were performed. The following table shows an overview of the statistical anonymisation measures implemented according to data form and mode of access.

---

<sup>50</sup> One exception to this is the information on 'Scholarship provider' (question 19). This characteristic is not available in the CUF or the download SUF.

**Table 6: Statistical Anonymisation Measures for the Data of the 19th Social Survey (2009)**

Characteristic	On-Site SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF
Direct Identifiers	Deletion and assignment of a random ID	Deletion and assignment of a random ID	Deletion and assignment of a random ID	Deletion and assignment of a random ID
Study subjects	Available	Aggregation to areas of study <sup>a</sup>	Aggregation to subject groups <sup>a</sup>	Aggregation to subject groups <sup>a</sup>
Degree type	Available	Available	Aggregation <sup>b</sup>	Aggregation <sup>b</sup>
Number of semesters until change of institution/subject	Available	Available	1-10 shown individually, otherwise aggregation (more than 10)	1-10 shown individually, otherwise aggregation (more than 10)
Higher education institution	Aggregation to type of higher education institution	Aggregation to type of higher education institution	Aggregation to type of higher education institution <sup>c</sup>	Aggregation to type of higher education institution <sup>c</sup>
Location of higher education institution	Aggregation to four groups of federal states	Aggregation to four groups of federal states	Aggregation to new or old federal states	Aggregation to new or old federal states
Number of interruption semesters	Available	Available	1-4 shown individually, otherwise aggregation: 5-10, more than 10	1-4 shown individually, otherwise aggregation: 5-10, more than 10
Waiting period until commencement of studies (in months)	Available	Available	Aggregation (years): ≤1, >1 to ≤2, >2 to ≤3, >3 to ≤5, >5 to ≤10, >10	Aggregation (years): ≤1, >1 to ≤2, >2 to ≤3, >3 to ≤5, >5 to ≤10, >10
Other sources of financing	Available	Aggregation <sup>d</sup>	Aggregation <sup>e</sup>	Aggregation <sup>e</sup>
Scholarship provider	Available	Aggregation to six categories	Deletion	Deletion
Reason for exemption from tuition fees	Aggregation to four categories	Aggregation to four categories	Aggregation to four categories	Aggregation to four categories
Level of tuition fees (in EUR)	Available	Aggregation: 0 to 100, >100 to 200, ..., >1,000	Aggregation: 0 to 250, >250 to 500, ..., >1,000	Aggregation: 0 to 250, >250 to 500, ..., >1,000
Form of grant (BAföG)	Available	Deallocation	Deletion	Deletion
Age (in years)	Available	0-39 shown individually, otherwise aggregation: 40-44, 45-50, ..., ≥75	0-39 shown individually, otherwise aggregation: 40-44, 45-50, ..., ≥75	0-39 shown individually, otherwise aggregation: 40-44, 45-50, ..., ≥75
Number of children	Available	1 and 2 shown individually, oth-	1 and 2 shown individually, oth-	1 and 2 shown individually, oth-

## 8 Anonymisation

		erwise aggregation: more than 2	erwise aggregation: more than 2	erwise aggregation: more than 2
Age of youngest child (in years)	Available	Aggregation: 1 to 2, 3 to 5, 6 to 11, 12 to 17, 18 and older	Aggregation: 1 to 5, 6 to 17, 18 and older	Aggregation: 1 to 5, 6 to 17, 18 and older
Number of siblings in education	Available	1, 2 and 3 shown individually. Otherwise aggregation: 4 or more	1 and 2 shown individually. Otherwise aggregation: 3 or more	1 and 2 shown individually. Otherwise aggregation: 3 or more
other than German nationality	Available	4 states shown individually, otherwise aggregation to world regions	Aggregation to world regions	Aggregation to world regions
Previous nationality	Available	5 states shown individually, otherwise aggregation to world regions	Aggregation to world regions	Aggregation to world regions
Nationality other than German of parents	Available	4 states shown individually, otherwise aggregation to world regions	Aggregation to world regions	Aggregation to world regions
First country of study/internship abroad	Available	Available	5 states shown individually, otherwise aggregation to world regions	5 states shown individually, otherwise aggregation to world regions
Other study-related periods abroad (countries)	Available	Available	Aggregation to world regions	Aggregation to world regions
Health characteristics	Deletion	Deletion	Deletion	Deletion

<sup>a</sup> According to Destatis Key List of Student and Examination Statistics (winter semester 2008/2009 and summer semester 2009).

<sup>b</sup> 'Clerical examination' and 'other degree' pooled to 'other degree (including clerical)'.

<sup>c</sup> Only distinction between university of applied sciences and university (including teacher training colleges, theological colleges, art and music colleges).

<sup>d</sup> Shown individually: 'child benefit' and 'benefits for/due to own child'; otherwise aggregation: 'other social benefits' and 'other sources'.

<sup>e</sup> Shown individually: 'child benefit'; otherwise aggregation: 'all other sources named here'.



## 9 Bibliography

### Publication bibliography

Apolinarski, Beate; Buck, Daniel; Kandulla, Maren; Middendorff, Elke; Naumann, Heike; Poskowsky, Jonas (2014): Daten- und Methodenbericht zum Scientific Use File der 20. Sozialerhebung. Version 1.0.0. Edited by Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). Hannover (Dokumentation).

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

Häder, Michael (2015): Empirische Sozialforschung. Eine Einführung. 3<sup>rd</sup> ed. Wiesbaden: Springer Fachmedien Wiesbaden, checked on 6/12/2015.

Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum. Institut für Arbeitsmarkt- und Berufsforschung (IAB). Nürnberg (06).

Isserstedt, Wolfgang; Kandulla, Maren (2010): Internationalisierung des Studiums - Ausländische Studierende in Deutschland - Deutsche Studierende im Ausland. Ergebnisse der 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem. Edited by Bundesministerium für Bildung und Forschung (BMBF). Bonn, Berlin.

Isserstedt, Wolfgang; Middendorff, Elke; Kandulla, Maren; Borchert, Lars; Leszczensky, Michael (2010): Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2009. 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem. Edited by Bundesministerium für Bildung und Forschung (BMBF). Bonn, Berlin.

Koberg, Tobias (2016): Disclosing the National Educational Panel Study. In Hans-Peter Blossfeld, Jutta von Maurice, Michael Bayer, Jan Skopek (Eds.): *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study*. Wiesbaden: Springer VS, pp. 691–708.

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

Kulik, Martina; Haaren, Friederike von; Thomsen, Stephan L. (2013): Studiengebühren: Kein Effekt auf das Ausgabeverhalten der Studierenden. Niedersächsisches Institut für Wirtschaftsforschung e. V. (NIW). Hannover (niw info spezial, 1).

Lane, Julia; Heus, Pascal; Mulcahy, Tim (2008): Data access in a cyber world: Making use of cyberinfrastructure. In *Transactions on Data Privacy* 1 (1), pp. 2–16.

Middendorff, Elke (2016): Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen und Themen. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). Hannover.

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

Sarclotti, Andreas (2015): Bachelor students' transition to postgraduate studies. Do students with and without migration background have different plans? In *Beiträge zur Hochschulforschung* 37 (2), pp. 116–139.

Thomsen, Stephan L. (2014): Did tuition fees in Germany constrain students' budgets? New evidence from a natural experiment. Forschungsinstitut zur Zukunft der Arbeit. Bonn (IZA Discussion Paper, 8623).

Valliant, Richard; Dever, Jill A.; Kreuter, Frauke (2013): *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York.