Data Observer

Beatrice van Berk and Anett Friedrich*

The Harmonised BIBB/BAuA Employment Survey 2006, 2012 and 2018 (H-ETB)

https://doi.org/10.1515/jbnst-2024-0048
Received May 13, 2024; accepted May 17, 2024

Abstract: The BIBB/BAuA Employment Surveys 2006, 2012 and 2018 are large cross-sections of the working population on qualification and working conditions in Germany and many publications are based on them. The data covers a wide range of topics, including information on occupations, education, job tasks, working conditions, job satisfaction and health status. A new harmonised dataset now compiles these three surveys into one file. The harmonised BIBB/BAuA Employment Survey (H-ETB) simplifies the use of the surveys for analyses over time and analyses of the pooled data. The dataset includes 60,048 cases and all variables that were surveyed in each wave in a harmonised form. The paper outlines the subject and methodology of the BIBB/BAuA Employment Survey as well as the harmonisation.

Keywords: occupation; employment; German labour market; working conditions; health

JEL Classification: C8; I26; J31; J44; J62; J81

1 Introduction

The analysis of changes in the employment conditions and workplaces in Germany is an ongoing task within the social and economic sciences. The harmonised

Beatrice van Berk, and Anett Friedrich contributed equally to this work and share first authorship.

***Corresponding author: Anett Friedrich**, Department 1.5, Research Data Center, Federal Institute for Vocational Education and Training (Bundesinstitut für Berufsbildung, BIBB), Friedrich-Ebert-Allee 114 – 116, 53113 Bonn, Germany, E-mail: anett.friedrich@bibb.de
Beatrice van Berk, Department 1.5, Research Data Center, Federal Institute for Vocational Education and Training (Bundesinstitut für Berufsbildung, BIBB), Friedrich-Ebert-Allee 114 – 116, 53113 Bonn, Germany

BIBB/BAuA Employment Survey (H-ETB) 2006, 2012 and 2018 (https://doi.org/10.7803/501.061218.1.1.10) is a valuable source of information on the changing nature of employment in Germany. The dataset combines three waves of the BIBB/BAuA Employment Survey of the Working Population on Qualification and Working Conditions in Germany (short BIBB/BAuA Employment Survey).[1] The pooled dataset simplifies the use of the BIBB/BAuA Employment Survey for analyses over time, thereby facilitating insights into long-term trends and dynamics of the German labour force.

The BIBB/BAuA Employment Survey is a representative survey of employees in Germany, aged 15 years or above and working at least 10 h per week. The surveys were conducted by the Federal Institute for Vocational Education and Training (BIBB) in collaboration with the Federal Institute for Occupational Safety and Health (BAuA). The data covers information about employment, the workplace and socio-demographic data.

The harmonised dataset of the BIBB/BAuA Employment Surveys 2006, 2012 and 2018 is useful for two purposes. Firstly, the data facilitates time comparisons. Secondly, analysing the pooled data allows more specific analysis for smaller groups, such as certain occupations, due to the larger number of cases.

A number of publications have already used the BIBB/BAuA Employment Surveys to conduct over time analysis. For instance, Bauknecht and Wesselborg (2022) have studied the emotional exhaustion in different occupations between 2006 and 2018. Further, Hall (2021) has used the surveys from 2006 to 2018 to analyse the effects of the Bologna reform on returns on education in Germany. Meyer, Tisch and Hünefeld (2019) have conducted regression analyses to examine the relationship between the implementation of new technologies and work demands.

The pooled data has already been the basis of several publications on smaller workforce groups. For instance, Rohrbach-Schmidt (2020) has used the pooled data from 2006, 2012 and 2018 to analyse whether foreign skilled workers have similar access to licensed and more credentialed occupations, and whether they profit from these regulations in terms of similar wages in these occupations to comparable domestic skilled workers. Storm (2023) demonstrates the presence of omitted-variable bias in conventional task data derived from expert assessment by using the BIBB/BAuA Employment Survey from 2012 to 2018 at the occupation-level.

The H-ETB simplifies various further use cases, for instance, the construction of indicators at occupation level for all three years, which is more straightforward

1 The individual waves of the BIBB Employment Survey are already available via the BIBB-FDZ (see https://www.bibb.de/de/62971.php).

for researchers. The information can then be linked to longitudinal data. Additionally, the H-ETB facilitates the application of complex statistical methods, including those in the field of causal inference, due to the large number of cases available. Matching methods can be used more effectively with a greater number of cases. Therefore, we are convinced that the research potential of time analyses and analysis of pooled data has not yet been fully exhausted and can be exploited with the new dataset.

The harmonised BIBB/BAuA Employment Survey comprises 596 variables surveyed in each of the three survey waves and all 60,048 cases. The principal advantage of this dataset over using each single wave is that all the harmonisation over time has already been done. The dataset only contains variables that can be analysed in all three survey years and these variables all have the same names, labels, categories and filters. Further, the dataset facilitates the easy assessment of which time or cohort comparisons can be processed with the BIBB/BAuA Employment Survey between 2006 and 2018, as it brings together the necessary information. This combination eliminates the need to check the questionnaire for each individual wave of the survey. Instead, researchers can use the codebook that comes with the data. Taken together, using the harmonised BIBB/BAuA Employment Survey for temporal analyses reduces the effort required for data preparation and saves time.

2 Subject Areas and Survey Instruments of the BIBB/BAuA Employment Surveys

The BIBB/BAuA Employment Surveys cover a wide range of topics, enabling research across different fields. First, the BIBB/BAuA Employment Surveys concentrate on the employment and the workplace context at the time of the survey, encompassing job tasks and required knowledge. Second, they focus on information about the job holders, such as job satisfaction, workload and health. Third, the surveys collect longitudinal information on the educational and career history of the respondents. Forth, the data includes sociodemographic and company data. Table 1 provides an overview of all covered topics.

The BIBB/BAuA Employment Surveys have several strengths: a wide range of indicators reflecting the value of vocational qualification, detailed information of all vocational qualifications, well-measured and in a multitude of ways classified information on current, first and training occupations, a high sample size of about 20,000 respondents per survey, and comparability of surveys over time.

Table 1: Overview of topics.

Question numbers	Topics
F100	Current occupation
F200	Working hours and workplace
F300	Tasks, key qualifications
F400	Occupational requirements
F500	Employment relationship (firm characteristics)
F600	Physical working conditions and stress
F700	Mental working conditions and stress
F1000	Changes in the working environment in the last two years
F1100	Highest general school qualification
F1200	Vocational education and training
F1300	Further training
F1400	First job and career history
F1450	Job satisfaction
F1500	Health conditions
F1600	Sociodemographics

3 Study Design and Data Collection

3.1 Sample

The population of all BIBB/BAuA Employment Surveys includes German-speaking employees in Germany who are at least 15 years old and work at least ten hours per week. The interview is conducted only in German and consequently, the sample comprises only individuals who speak sufficiently German. The concept of an employee was operationalized to include individuals in income-related employment for a minimum of 10h per week at the time of the survey. The sample includes foreign nationals who spoke sufficient German, family workers, employees who temporarily ceased their employment for a period of less than three months and those who were engaged in paid work alongside an apprenticeship, academic programme, part of a legal clerkship or specialist training. The sample excludes volunteers, trainees, apprentices and those who are engaged at the federal volunteer service or military (Rohrbach-Schmidt and Hall 2009; Rohrbach-Schmidt and Hall 2013; Rohrbach-Schmidt and Hall 2020).

The data for all three waves was collected by Kantar Public (formerly TNS Infratest Sozialforschung) as Computer Assisted Telephone Interviews (CATI). In 2006 and 2012 the interviews were conducted exclusively via landline telephone, whereas in 2018, a dual-frame was employed, necessitating additional mobile phone

interviews. This adjustment was necessitated by changes in people's communication habits, which in recent years have resulted in a number of individuals being reachable only by mobile phone and no longer by landline.

The sampling of the landline and mobile numbers for all three waves was based on Kantar Public's in-house telephone master sample, which was designed for population representative surveys and produces unbiased samples without clustering effects. The random selection was carried out using a random digit dialling in accordance with the ADM (Arbeitsgemeinschaft Deutscher Markt- und Sozialforschungsinstitute[2]) standard, a further development of the Gabler-Häder process. The landline sample was selected via a two-stage process, necessitating the use of a Kish Selection Grid to randomly select a target person within each household. In contrast, the mobile sample was selected based on the telephone user being the target person, thus eliminating the need for further selection process (Gensicke and Tschersich 2018: 12; Gensicke, Tschersich and Hartmann 2012: 15; Hartmann 2006: 18).

3.2 Fieldwork and Response Rates

Two pretests were used to identify potential issues with the survey instrument and to assess the clarity and the length of the interview.

To ensure the quality of the interviews, the interviewers underwent a specialised training programme. Following the interview training, the interviewers conducted a test interview until they were fully familiar with all the questions.

During the fieldwork, the survey institute has implemented a series of data checks and monitoring procedures to ensure high data quality. Further, the institute has collected early and regular interim data, which subjected to rigorous scrutiny to ensure correct filtering, plausibility and usability of the open job titles. Overall, the survey institute puts great effort into their quality management, both before and during the survey, resulting in high quality data (Gensicke and Tschersich 2018: 12; Gensicke, Tschersich and Hartmann 2012: 15; Hartmann 2006: 18).

The field period starts in October and ends in March or April for each survey. The interviews last on average about 40 min. Table 2 illustrates the non-responses and response rates of each survey year. Overall about 20,000 completed interviews have been realised in each survey. It is noteworthy that the gross sample and the number of eligible addresses were considerably larger in 2018 than in 2006 and 2012. The reason for this discrepancy is the switch to a dual-frame sample. Since a greater proportion of mobile numbers are invalid, a greater number of them have to be included in the gross sample.

2 The English translation is Association of German Market and Social Research Institutes.

Table 2: Non-response and response rate.

	2006	2012	2018
Gross sample	262,876	317,980	2,661,712
Sample-neutral non-response	43.0 %	56.5 %	79.4 %
Adjusted gross sample	149,840	138,321	548,393
Not sample-neutral non-response (unknown eligibility)	5.5 %	11.6 %	7.1 %
Eligible addresses	141,573	122,257	509,346
Not sample-neutral non-response (eligible)	56.0 %	55.7 %	86.1 %
Number of interviews	62,253	54,152	60,399
Interview with not employed person	33,539	29,105	35,243
	(53.8 %)	(53.7 %)	(58.4 %)
Person not belong to target group (according screening interview)	8714 (14.0 %)	4350 (8.0 %)	5144 (8.5 %)
Completed interviews	20,000	20,036	20,012

Source: (Gensicke and Tschersich 2018: 12; Gensicke, Tschersich and Hartmann 2012: 15; Hartmann 2006 : 18).

The high of proportion of sample-neutral non-response in each year can be attributed to two key factors. Firstly, the selected telephone numbers are randomly generated, and as a result, a significant proportion of them are not real or assigned telephone numbers. Secondly, despite the maximum number of five contacts being attempted, no one could be reached.

In the event that the contact person declined to provide any further information, including details such as the number of employees in the household or age, it is not possible to identify a target person. To identify the target person, the contact person must respond to three to four questions following the screening interview, which is conducted to ensure that the target person belongs to the target population. For mobile numbers, the screening interview was conducted if the contact person was older than 15 and the main user of the mobile phone (Gensicke and Tschersich 2018: 12; Gensicke, Tschersich and Hartmann 2012: 15; Hartmann 2006: 18).

3.3 Weighting

The H-ETB dataset comprises two weighting factors and an extrapolation factor. The design weight serves to equalise the selection probability for telephones in the household, which depends on the number of lines available for calls, and the probability of selection for individuals, which depends on the number of individuals in the household in the target population. It should be noted that the procedures for creating the design weight in 2018 differ from those in 2012 and 2006 due to the dual-frame approach. The design weight from 2018 incorporates the proportion of

individuals who exclusively utilise mobile phones, i.e. mobile phone owners who do not possess a landline number. This aspect did not have to be considered in the 2006 and 2012 surveys.

The design weight constitutes the input factor for creating the adjustment weight. The adjustment weight adjusts the sample structures to those of the population by means of weighting factors (calibration). As part of the calibration process for determining the weighting factors, a systematic examination was conducted to ascertain whether and to what extent there were disproportionate dropouts. This examination was achieved by comparing the distributions of the realised net sample with that of the population. Once the characteristics relevant for an adjustment had been defined, the weighting model was implemented with multidimensional marginal distributions. The adjustment weight compensates for non-response by making a structural adjustment to the population based on the 2017/2011/2005 microcensus. As previously described, a reference structure has been created, which is slightly different in 2006, 2012 and 2018. The reference structure contains the following variables: federal state, BIK regions, administrative district (only 2012), western versus eastern Germany, household size, occupational status, gender, nationality, marital status, education, age, employment status (yes/no) (not 2018), number of landlines in the household (not 2018), number of target persons in the household (not 2018). The methodology used for the calculation of the adjustment weight in the 2018 survey differs from those utilised in the 2006 and 2012 surveys, due to the change in the screening procedure in 2018 (Gensicke and Tschersich 2018). The extrapolation factor can be used to estimate the number of employees in Germany based on the microcensus.

The use of the two weighting and the extrapolation weight is straightforward as all annual values are combined into one variable. The use of the extrapolation factor means that each value is multiplied by a factor which results in it being included in the total according to its significance for the corresponding year. Consequently, if the extrapolation factor is used in the H-ETB, the output will be the sum of the employees in Germany for the years 2006, 2012 and 2018 respectively.

We recommend that researchers use a weighting factor in the calculation of descriptive statistics. When the estimation of proportions or mean population values is required, the adjustment weight should be applied. Additional information on weighting can be found in the data and method reports as well as the field reports for each survey (Gensicke and Tschersich 2018; Gensicke, Tschersich and Hartmann 2012; Hartmann 2006; Rohrbach-Schmidt and Hall 2009; Rohrbach-Schmidt and Hall 2013; Rohrbach-Schmidt and Hall 2020).

4 Overtime Harmonisation

The H-ETB incorporates variables from the SUFs of individual waves of the BIBB/BAuA Employment Survey, which are comparable across 2006, 2012 and 2018. All variables only asked in one or two waves were excluded from the harmonised dataset unless they are the basis for the filter. Currently, the variable names are not consistently used across the three SUF, which implies additional effort for the data preparation for over time analyses. As part of the harmonisation process, we standardised all the variable names, thereby reducing the number of necessary data preparation steps.[3] The same is true for the variable and value labels of the H-ETB. Again, differences between the waves exist which are not driven by changes in the data collection. We standardised all variable and value labels based on the 2018 survey.

Although the majority of the questions remained unchanged over time, some modifications were necessary. The date of the H-ETB indicates by the variable label which questions were affected and in which survey year the question changed.

For some variables, further harmonisation steps were necessary, such as merging categories or adjusting filter guides so that they contain comparable information across all three surveys. To ensure transparency, in these cases, a copy of the original variables for each survey year was created with the endings "o06", "o12" or "o18". The missing values from the other survey years of these variables were coded with −9006, −9012 and −9018. With these variables at hand, users can decide whether they want to follow our harmonisation or whether, for their research question, another kind of harmonisation would be more appropriate.[4]

As previously mentioned, an advantage of the data are the occupational codes. However, the coding scheme changed over time. In order to facilitate the data preparation process, the number of occupational classifications in the H-ETB has been reduced. This allows for the classifications to be used in a comparable manner across all three waves. Table 3 shows which occupational classifications are included for each type of occupation.

5 Data Access and Further Documentation

The harmonised BIBB/BAuA Employment Survey 2006, 2012 and 2018 is available as a Scientific Use File (SUF) at the Research Data Centre of the Federal Institute of

3 For transparency, the BIBB-FDZ provides a table containing the variable names for the H-ETB and the variable names from the SUFs of the three surveys on its webpage.

4 Van Berk, Friedrich and Hohn (forthcoming) provide detailed information on the harmonisation process on the variable level in the respective data and methodological report.

Table 3: Occupational classifications.

Classification	Occupation			
	Current	First	Training	Father and mother
KldB 2010	x	x	x	x
BIBB definitions of occupational fields	x	x		x
KldB 1992			x	x
ISCO08	x	x	x	
European Socio-Economic Groups 2008	x			
International socio-economic index of occupational status (ISEI) 2008	x	x		x
Standard international occupational prestige scale (SIOPS) 2008	x	x		x
Blossfeld's occupational classification	x	x		x
Erikson-Goldthorpe-Portocarero scheme (EGP)	x			x
Occupational prestige scale (BAS-3, BAS-2|5)	x	x	x	x
Production and service occupations in the dual system (BBiG and HwO)			x	

Vocational Education and Training (BIBB-FDZ). The use of the data is subject to guidelines and requires application (see https://www.bibb.de/en/1400.php). The application could be submitted via e-mail to fdz@bibb.de, via fax to +49 – (0)228-1072020 or via mail to BIBB – Federal Institute for Vocational Education and Training research data centre P.O. Box 201264 53142 Bonn, Germany. After receiving the application, the BIBB-FDZ provides the data as download.

In accordance with data protection regulations, the detailed occupational codes are anonymised and excluded from the SUF. However, a separate dataset for special variables includes those variables and users can apply for it (https://www.bibb.de/dokumente/pdf/BIBB_FDZ_application_ZSV.pdf).

In addition to the data, the BIBB-FDZ provides a codebook in both German and English, which includes all variables their variable and value labels, the corresponding question, and, if applicable, the filter. The data itself is labelled in both German and English.[5]

5 In Stata one can change the language with the command 'label language eng' and for SPSS and R users an additional dataset labelled in English is available.

For university lectures and students who have not yet attained a master's degree, the BIBB-FDZ provides a Campus-File (CF; https://doi.org/10.7803/501.061218.1.8.10), which is a more anonymized version of the Scientific Use File. The CF generates identical and comparable results.

6 Outlook

The latest wave of the BIBB/BAuA Employment Surveys just has taken place. Together with the SUF for the 2024 data, just like with every new wave, also the H-ETB will be update.

In addition to potential new waves, there are also those surveys from before 2006. Harmonising those waves as well would be also a great value for the scientific community. However, these datasets are less comparable with those after 2006. Hence, more harmonisation steps are necessary which will be a future project.

Acknowledgements: The authors would like to thank Christine Hohn for her valuable help in the harmonisation of the data set. We further thank our colleagues Anja Hall and Daniela Rohrbach-Schmidt for their very helpful comments on our manuscript and the data set.

References

Bauknecht, J., and B. Wesselborg. 2022. "Psychische Erschöpfung in sozialen Interaktionsberufen von 2006 bis 2018." *Prävention und Gesundheitsförderung* 17: 328–35.

Gensicke, M., and N. Tschersich. 2018. *BIBB/BAuA-Erwerbstätigenbefragung 2018*. München: Methodenbericht. Also available at https://metadaten.bibb.de/download/pdf/978_20200213_1581598207/KANTAR_Methodenbericht_ETB1718_final.

Gensicke, M., N. Tschersich, and J. Hartmann. 2012. *BIBB/BAuA-Erwerbstätigenbefragung 2011/2012*. München: Feldbericht – Los 1. Also available at https://metadaten.bibb.de/download/pdf/126_20130801_1375355971/Feldbericht_BIBB_BAUA_2012.

Hall, A. 2021. "Alles beim Alten? Bildungserträge höherer beruflicher und akademischer Abschlüsse vor und nach Bologna." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 73: 527–53.

Hartmann, J. 2006. *BIBB/BAuA-Erwerbstätigenbefragung 2005/2006*. München: Feldbericht.

Meyer, S.-C., A. Tisch, and L. Hünefeld. 2019. "Arbeitsintensivierung und Handlungsspielraum in digitalisierten Arbeitswelten – Herausforderung für das Wohlbefinden von Beschäftigten?" *Industrielle Beziehungen. Zeitschrift für Arbeit, Organisation und Management* 26: 207–31. https://elibrary.utb.de/doi/pdf/10.3224/indbez.v26i2.06?download=true.

Rohrbach-Schmidt, D. 2020. "Licensing, Educational Credentialing, and Wages Among Foreign Skilled Workers in Germany." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72: 375–400.

Rohrbach-Schmidt, D., and A Hall. 2009. *The BIBB/IAB- and BIBB/BAuA Surveys of the Working Population on Qualification and Working Conditions in Germany*. BIBB-FDZ Data and Methodological Report. Also

available at https://metadaten.bibb.de/download/pdf/5_20180813_1534158892/FDZ_Data_Manual_ETB_2006.

Rohrbach-Schmidt, D., and A. Hall. 2013. *BIBB/BAuA Employment Survey 2012*. BIBB-FDZ Data and Methodological Report. Also available at https://metadaten.bibb.de/download/pdf/443_20200331_1585638105/FDZ_DuMB_ETB_2012_EN_6_0.

Rohrbach-Schmidt, D., and A. Hall. 2020. *BIBB/BAuA Employment Survey 2018*. BIBB-FDZ Data and Methodological Report. Also available at https://metadaten.bibb.de/download/pdf/972_20200213_1581597502/FDZ_DuMB_BIBB_BAuA_ETB_2018.

Storm, E. 2023. "On the Measurement of Tasks: Does Expert Data Get it Right?" *Journal for Labour Market Research* 57 (6): 1–24.

van Berk, B., A. Friedrich, and C. Hohn. forthcoming. *Harmonisierter Datensatz der BIBB/BAuA-Erwerbstätigenbefragung (H-ETB) 2006, 2012 und 2018. BIBB-FDZ Daten- und Methodenbericht*. Bonn: BIBB.