

### How Standardized is Occupational Coding? A Comparison of Results from Different Coding Agencies in Germany

Massing, Natascha; Wasmer, Martina; Wolf, Christof; Züll, Cornelia

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Massing, N., Wasmer, M., Wolf, C., & Züll, C. (2019). How Standardized is Occupational Coding? A Comparison of Results from Different Coding Agencies in Germany. *Journal of Official Statistics*, 35(1), 167-187. <https://doi.org/10.2478/jos-2019-0008>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

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

#### Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

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

# How Standardized is Occupational Coding? A Comparison of Results from Different Coding Agencies in Germany

*Natascha Massing<sup>1</sup>, Martina Wasmer<sup>1</sup>, Christof Wolf<sup>1</sup>, and Cornelia Zuell<sup>1</sup>*

As occupational data play a crucial part in many social and economic analyses, information on the reliability of these data and, in particular on the role of coding agencies, is important. Based on our review of previous research, we develop four hypotheses, which we test using occupation-coded data from the German General Social Survey and the field test data from the German Programme for the International Assessment of Adult Competencies. Because the same data were coded by several agencies, their coding results could be directly compared. As the surveys used different instruments, and interviewer training differed, the effects of these factors could also be evaluated.

Our main findings are: the percentage of uncodeable responses is low (1.8–4.9%) but what is classified as “uncodeable” varies between coding agencies. Inter-agency coding reliability is relatively low  $\kappa$  ca. 0.5 at four-digit level, and codings sometimes differ systematically between agencies. The reliability of derived status scores is satisfactory (0.82–0.90). The previously reported negative relationship between answer length and coding reliability could be replicated and effects of interviewer training demonstrated. Finally, we discuss the importance of establishing common coding rules and present recommendations to overcome some of the problems in occupation coding.

*Key words:* Occupation coding; coding rules; ISCO.

## 1. Introduction

Occupations are an important outcome of previous life decisions and a determinant of life chances. Therefore, occupation is an important variable in social and economic research; it allows the analysis of labor market processes, social mobility, and status attainment, to mention just a few. Occupation is a complex construct that is difficult to measure and requires categorization. Most surveys aiming at collecting data on occupation include open-ended questions about the respondent’s current or last job. Hence, responses are recorded verbatim.

In order to be analyzed, these textual data on occupation must usually be coded, that is, unstandardized texts must be aggregated into pre-defined categorical systems, for example, the International Standard Classification of Occupations (ISCO) or the Standard Occupational Classification (SOC). Researchers often derive socioeconomic status scores, occupational prestige scores or social class positions (Ganzeboom and Treiman 2003, 159–193), or occupational health hazard scales (‘t Mannetje and Kromhout 2003) from these occupational codes.

<sup>1</sup> GESIS-Leibniz Institute for the Social Sciences, B2.1, 68159 Mannheim, Germany. Emails: [natascha.massing@gesis.org](mailto:natascha.massing@gesis.org), [martina.wasmer@gesis.org](mailto:martina.wasmer@gesis.org), [christof.wolf@gesis.org](mailto:christof.wolf@gesis.org), [cornelia.zuell@gesis.org](mailto:cornelia.zuell@gesis.org)

Because of the great importance of the occupation variable, it is essential to understand the coding process and to assess the quality of coding results. In the present article, we therefore address the following questions: What percentage of responses to open-ended questions about occupation are uncodeable? How reliable are occupational codings? With regard to the latter question we ask, in particular, to what extent coding results are affected by rules implemented by the coding agency (“house effects”). We explore how the application of different coding rules influences measures of socioeconomic status (SES) and occupational prestige. Finally, we analyze the relationship between the length of answers given by respondents and the reliability of coding across different agencies.

In order to answer these questions, we use data from the German General Social Survey (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS) 2010 and the German field test of the Programme for the International Assessment of Adult Competencies (PIAAC) carried out in 2010. Occupational information was coded into the 2008 version of ISCO (ISCO-08) by three agencies for ALLBUS and two agencies for PIAAC.

This article is structured as follows: We begin by briefly reviewing relevant literature on the coding of occupations and discussing key challenges in this field. Based on this presentation, we develop hypotheses that will guide our empirical analysis. After a short description of the data and the coding procedures used in our study, we present our results. We conclude with a discussion of the main findings and implications of our study.

## 2. Challenges of Occupation Coding and Previous Research

Transferring textual information on occupations into numerical information is a multifaceted and demanding task. Errors can occur at several stages and impact the quality of coding. To date, both theoretical attempts to systematize and explain factors affecting the quality (i.e., validity and reliability) of occupational coding results and empirical research on this topic have been limited. Elias (1997, 13) distinguished between 1) problems relating to the extent and quality of the *data to be coded*, 2) problems relating to the *classification itself*, and 3) problems relating to the formulation and application of *coding rules* and the *coding process*.

In interviewer-administered surveys, interviewers must ask the questions and record the information provided by respondents. The first prerequisite to obtaining valid and reliable information on occupation is the use of adequate questions. The usual recommendation is to use at least two separate questions (International Labour Office (ILO) 2012, 55f.) to obtain both information about the job title and about the main tasks and duties performed in the job. Hoffmann et al. (1995) point out that the questions should be simple, containing familiar, widely understood terms. Furthermore, they underline the importance of specific design features. For example, including instructions and examples in the questions seems to have different effects depending on the education of respondents. They also show that the size of the text fields provided for the answers affects the answer length because it serves as a cue to the level of the detail expected (Hoffmann et al. 1995). Hak and Bernts (1996) point out the fact that the interpretation of answers by the coders is a preliminary key step of the coding process. In particular, when respondents used vague or ambiguous

terms or provided contradictory information, the assignment of a specific occupational code is mainly a matter of interpretation. One can try to improve the quality of the responses by appropriate interviewer training (Billiet and Loosveldt 1988).

Indeed, several authors have reported challenges encountered during the interview process that might result in problems when coding answers to the questions about occupation. For example, Schierholz et al. (2017) reported that respondents tend to provide incomplete or contradictory information, which influenced coding results. Similarly, Geis and Hoffmeyer-Zlotnik (2000, 113) reported that typically 15% to 25% of answers cannot be completely coded to the most detailed level. Therefore, our first hypothesis is:

H1: We expect the percentage of not completely codeable answers to be around 20%.

The quality of coding is not only influenced by the question used or the interviewer training, but also depends on the precision, completeness, and clarity of the coding scheme. In this context, the descriptions of the categories play a crucial role in helping coders to find the correct codes. However, changes in occupational specialization reduce the applicability of coding schemes over time, and the typical lengthy updating phase of such classifications might be too long to capture ongoing changes in occupational structures. Coding schemes developed for cross-national comparison are even more challenging because they have to find the right balance between categories reflecting international comparative structures and national specificities. In this article, we cannot test the effect of the coding scheme itself on the quality of codings, as we use only one coding scheme in our empirical analysis.

Basic criteria reflecting the quality of coding results are inter-coder reliability – that is, the extent to which the same code is assigned to a given text by different coders – and internal validity – that is, the extent to which the most appropriate code is selected. For our data, we do not have any codings that could serve as a “gold standard” to assess validity. Therefore, we will focus on reliability, keeping in mind that high reliability does not guarantee high validity, but certainly is a necessary requirement. Several measures of reliability are available, for example, simple percent agreement, Cohen’s kappa, Scott’s pi, and Krippendorff’s alpha (see, e.g., Freelon 2010). We will report Cohen’s kappa – a coefficient measuring agreement adjusting for agreement occurring by chance. With an increasing number of categories, random agreement is negligible and Cohen’s kappa (multiplied by 100) is only marginally lower than percent agreement.

One way to increase the reliability of occupational coding is to formulate and apply rules capturing the definitions and general guidelines of an occupational classification (see, for example, Geis and Hoffmeyer-Zlotnik 2000). Usually, the occupational classification manual includes some general rules. For example, in the publication *International Standard Classification of Occupations: Structure, Group Definitions and Correspondence Tables*, which is referred to in what follows as the “ISCO manual”, ILO (2012) suggests that the following three rules (in that order) should be applied when classifying jobs with a broad range of tasks and duties: (1) If the tasks and duties require different skill levels, the job should be classified in accordance with the tasks and duties that require the *highest level of skills*. (2) If the tasks and duties are connected with different stages of the production and distribution process, tasks and duties related to the

*production process* should take priority. (3) In cases with tasks and duties at the same skill level and at the same stage of production and distribution, the job should be classified according to the predominant, that is, most time-consuming, tasks performed. Detailed explanations of specific boundaries (for example, concerning the distinction between managers and supervisors or operators of small businesses) are included in the ISCO manual. Also, recommendations are provided regarding the use of job-related information other than job title and main tasks and duties actually performed. The developers of ISCO (ILO) recommend that decisions should be based on the tasks actually performed, rather than on any other information.

However, beyond general clarifications and rules, more specific operational rules are needed to determine the most appropriate occupational code for a given description of a particular job. When automatic dictionary-based coding is applied, the program that assigns codes to text can be considered part of these rules. In addition, to ensure consistency, previous coding decisions will usually be incorporated into agencies' coding rules. Finally, recurring incomplete or ambiguous answers require rules on how to deal with these cases. As an example, [Ganzeboom and Treiman \(1996, 210\)](#) recommend providing coders with information on the numerical sizes of the specific occupational categories, and thus on the probability of category membership, in order to help them code ambiguous cases. Specifications on whether and how further job-related information (e.g., employment status or size of organization) are to be taken into account in case of doubt are also an important part of practical coding rules. This is especially relevant if important information about performed tasks is lacking in the input material. In the actual process of assigning codes to verbatim answers, the individual coders will – more or less systematically – use and “enrich” these rules in their own – more or less idiosyncratic – way. This may be a source of bias, especially when the correlated coder variance is high because of high workload per coder ([Campanelli et al. 1997, 444–445](#)).

[Belloni et al. \(2016\)](#) studied coding errors in occupational data in the Netherlands using data from the Survey on Health, Ageing and Retirement in Europe (SHARE). The authors recoded responses to open-ended questions about occupation for the Dutch sample of the SHARE data using software for semi-automatic coding: all cases above the certainty score threshold of 70 were coded automatically; all residual cases were coded manually by an expert coder. This coding was used as a benchmark and compared to the results of the coding according to the standard procedure implemented in SHARE (manual coding by trained coders). Inter-coder agreement at the one-digit ISCO level was 71% for current job and 72% for last job; at the three-digit ISCO level, it was 52% for current job and 56% for last job.

[Campanelli et al. \(1997\)](#) reported results of two similar empirical studies, in which they compared results from different coders of the Office of National Statistics (ONS) and applied manual, computer-assisted and computer-automated coding methods on data of the British Household Panel. Data were coded in accordance to the UK Standard Occupational Classification (SOC), a classification similar to ISCO. Reliability for manual codings was between 0.75 and 0.80. Validity, assessed by comparisons with expert coders, yielded agreement rates, ranging between .69 and .84. Only modest gains in reliability and validity could be obtained by using computer-assisted methods.

Several studies have dealt with the question of how the quality of coding results is influenced by the *rules applied during the coding process*. For manual coding, [Hak and Bernts \(1996\)](#) argued that the effectiveness of training in terms of inter-coder reliability is improved, not only by “communicating coding instructions to coders (*theoretical training*)”, but also by “socializing coders into practical rules” beyond the general coding instructions. Such practical coding rules are, at least partly, specific to the coding agency (or individual coders) and, as such, reflect their experience and expertise. Thus, coding rules can have a negative effect, as they can lead to what [Bushnell \(1998\)](#) referred to in a different context as the “coding system bias.” Depending on the coding agency and the rules in place, results of coding can differ systematically between agencies (“house effects”). The higher the differences between these agency-specific rules, the higher is the likelihood of house effects in occupational coding, that is, that the coding will differ systematically between agencies. Thus, although these house effects increase the reliability within one agency, they may reduce reliability between different agencies and jeopardize validity. Thus, our second hypothesis is:

H2: The coding reliability between different coding agencies is lower than the coding reliability within agencies.

In social research, occupational codes are often used to calculate occupational prestige and socioeconomic status (SES) scores ([Ganzeboom and Treiman 2003](#)). The agreement between prestige and SES scores derived from the codings of different agencies will be higher than the inter-agency agreement between the occupational codes. This is necessarily true because merging a large number of codes into a smaller number of categories results in higher agreements per se. Additionally, we argue that in the cases where the choice between different codes is hard to make, the occupations in question are more similar with respect to prestige and SES than a randomly chosen pair of occupations would be. In a study by [Maaz et al. \(2009\)](#), occupational information was coded by professional coders and lay coders (trained student assistants). The level of consistency between these codings was generally not as high as desirable. For example, for the ten one-digit codes of the International Standard Classification of Occupations of 1988 (ISCO-88), the authors reported a Cohen’s kappa ( $\kappa$ ) of around 0.67. For the 390 more detailed four-digit ISCO codes,  $\kappa$  did not exceed 0.5. When occupational codes were converted into International Socio-Economic Index (ISEI) scores, correlations were generally higher (between 0.75 and 0.85). Therefore, our third hypothesis is:

H3: Socioeconomic status and occupational prestige scores will show high levels of agreement.

Finally, we are concerned with the relation of answer length and codeability. Previous research suggests that longer text strings do not necessarily result in more reliable codings. [Conrad et al. \(2016\)](#) found that longer descriptions of occupations were less reliably coded than shorter descriptions. For coders, texts that provide too much detail or are too complex can be difficult to interpret within the framework of an occupational classification and can make it difficult to decide on a unique occupational code. The [Conrad et al. \(2016\)](#) study corroborates [Bergmann and Joye \(2005, 9f\)](#) that “. . . the more detailed the information to be sorted into occupational groupings, the less reliably individual cases are assigned to

categories.” It is also in line with the findings of [Cantor and Esposito \(1992\)](#). They found that coders who were asked to comment on interviewer recordings of occupations only rarely indicated that these should contain more specific information; some even criticized the fact that interviewers had provided too much information. Thus, our fourth hypothesis is:

H4: Long answers to open-ended questions about occupation do not result in higher coding reliability.

Before we present analyses on our hypotheses, we briefly discuss the data and approach we use in our empirical study.

### 3. Data, Classification, and Coding Procedure

#### 3.1. Data

To test our hypotheses, we use data from the German General Social Survey (ALLBUS) 2010 and the German field test of the Programme for the International Assessment of Adult Competencies (PIAAC) conducted in 2010. Both ALLBUS and PIAAC were carried out as computer-assisted face-to-face interviews with randomly selected respondents from official population registers.

We restricted the samples to respondents in gainful employment aged between 18 and 65 years. In our analysis, we considered the respondents’ current or last occupations and their parents’ occupations. We did not differentiate between these different types of occupations because coding reliability showed no substantial differences.

Respondents’ job titles and activities were measured with a two-part open-ended question in ALLBUS and a three-part open-ended question in the PIAAC field test (see [Table 1](#)). These questions were embedded in different background questionnaires that also

Table 1. Questions about current occupation used in ALLBUS and PIAAC.

ALLBUS	PIAAC
1. What work do you do in your main job? Please describe your work precisely.	1. What job are you in at the moment?
	2. Please describe this job exactly. Please give the exact title of the job. For example, rather than just saying “management assistant” give the full title “management assistant in freight forwarding”; instead of merely stating “worker”, give the exact title, for example “machine fitter”. If you are a civil servant, please give your exact grade, for example “police sergeant” or “tenured secondary school teacher”. And if you are a trainee/apprentice, then state the profession in question. <sup>1</sup>
2. Does this job, this work have a special name?	3. Does this job have a special name?

<sup>1</sup>A similar instruction was provided in ALLBUS in the general interviewer training. However, this instruction did not appear on the screen and was therefore not read out to the respondents.

included other occupation-related questions. The ALLBUS respondents were first asked about their status in employment using a very detailed classification that distinguished white-collar workers according to their tasks, civil servants according to their career paths, blue-collar workers according to their qualifications, employers by number of employees, and farmers by the size of their utilized agricultural area (for more details, see Supplemental data, Table A1). This was followed by the open-ended questions about occupation presented in Table 1. The first sub-question of ALLBUS asked for a description of work done in the job. Strictly speaking, only the second sub-question referred to the key information “job title”, although in practice, many respondents reacted to the first sub-question by naming the title of their job. After these questions, several other occupation-related questions (e.g., about supervisory status) were asked. The original question wording in German can be found in the Supplemental data (Table A2).

The question about occupation in PIAAC was similar to, but more detailed than that in ALLBUS (see Table 1). There were three sub-questions about occupation, starting with a sub-question explicitly asking for the exact job title. In contrast to ALLBUS, all sub-questions required an answer (a text entry or a click on the “don’t know” or “refused” button). Also in PIAAC, these questions were followed by several occupation-related questions (e.g., the industry that the respondent was working in, and supervisory status).

In both surveys, respondents who were currently not employed, but who had worked in the past, were asked about their last job. Furthermore, respondents were asked about their mothers’ and fathers’ occupations when the respondents were 15 (ALLBUS) or 16 (PIAAC) years old. The wording was similar to that of the questions about the respondent’s occupation, however only limited ancillary information was provided to the coders about the parents (formal qualifications and, in ALLBUS, employment status).

To carry out the survey, the PIAAC interviewers were trained very thoroughly. Their training included instruction on how to retrieve and record information on occupation; they were instructed to ask the respondents in a way that the interviewers actually understood what kind of job the respondents were doing. Furthermore, they also had to practice how to ask the occupation questions. In ALLBUS there was no such training. However, the interviewer instructions for the study (in written form) included explanations and examples to illustrate the level of detail necessary with respect to the open-ended questions on occupation.

### 3.2. The ISCO-08 Classification Scheme

ISCO-08 is the current version of the International Standard Classification of Occupations curated by the International Labour Organization, a specialized agency of the United Nations. It “provides a system for classifying and aggregating occupational information . . . [and] allows all jobs in the world to be classified into 436 unit groups” (ILO 2012, 3). For the purposes of ISCO-08, a *job* is defined as “a set of tasks and duties performed, or meant to be performed, by one person” (ILO 2012, 11). Jobs with very similar main tasks and duties are aggregated into “occupations”. Different occupations form the most detailed level of ISCO, unit groups.

Besides *job*, the second main concept underlying ISCO classification is *skill*, which is defined as “the ability to carry out the tasks and duties of a given job” (ILO 2012, 11). More

specifically, in ISCO, occupations are categorized according to the typically required *skill level* and *skill specialization*. Skill level is differentiated into four ordered groups from low to high educational qualifications and accompanying levels of literacy and numeracy. As can be seen in [Table 2](#), Major Group 9: Elementary Occupations is characterized by Skill Level 1; Major Group 3: Technicians and Associate Professionals by Skill Level 3; Major Group 1: Managers by Skill Levels 3 and 4; Major Group 2: Professionals by Skill Level 4; and all other major groups by Skill Level 2. Major Group 0: Armed Forces Occupations is an exception, in that it can contain a broad range of skill levels.

Within the major groups, occupations are grouped by the type of skill specialization, that is, “the field of knowledge required; the tools and machinery used; the materials worked on or with; and the kinds of goods and services produced” ([ILO 2012, 11](#)). It is important to note that the concept of skill level and skill specialization refers to the requirements of jobs and occupations, rather than the skills or education of a specific job incumbent.

Overall, ISCO follows a hierarchical structure where an increasing level of detail is expressed as one- to four-digit codes. As mentioned above, unit groups, which are denoted by four-digit codes, constitute the most detailed level of the classification. These unit groups are aggregated into 130 minor groups, expressed as three-digit codes, which in turn are grouped into 43 sub-major groups, denoted by two-digit codes. At the highest level, the classification comprises 10 major groups (see [Table 2](#)).

The structure of the classification is illustrated in [Table 3](#). For example, Unit Group 3112: Civil Engineering Technicians is part of Minor Group 311: Physical and Engineering Science Technicians, which in turn is part of Sub-major Group 31: Science and Engineering Associate Professionals, which belongs to Major Group 3: Technicians and Associate Professionals.

Table 2. Structure of ISCO–08, number of categories, and skill level.

ISCO-08 Major group		Sub-major groups	Minor groups	Unit groups	Skill level
1	Managers	4	11	31	3 + 4
2	Professionals	6	27	92	4
3	Technicians and Associate professionals	5	20	84	3
4	Clerical support workers	4	8	29	2
5	Services and sales workers	4	13	40	2
6	Skilled agricultural, forestry and fishery workers	3	9	18	2
7	Craft and related trades workers	5	14	66	2
8	Plant and machine operators, and assemblers	3	14	40	2
9	Elementary occupations	6	11	33	1
10	Armed forces occupations	3	3	3	1 + 2 + 4
Total		43	130	436	

Source: ([ILO 2012, 14, 22](#)).

Table 3. Extract from ISCO-08.

3	Technicians and associate professionals
31	Science and engineering associate professionals
311	Physical and engineering science technicians
3111	Chemical and physical science technicians
3112	Civil engineering technicians
3113	Electrical engineering technicians
...	...
3119	Physical and engineering science technicians not elsewhere classified
312	Mining, manufacturing and construction supervisors
...	...
32	Health associate professionals
321	Medical and pharmaceutical technicians
...	...
324	Veterinary technicians and assistants
3240	Veterinary technicians and assistants
325	Other health associate professionals
...	...

### 3.3. The Coding Process

For our study, a number of agencies were contracted to independently code all the occupational data. These agencies are experienced in occupational coding and offer this service commercially; all agencies offer their service roughly in the same price range. All agencies trained their coders according to their own procedures. Most of the coders were experienced in coding occupations, however, we do not have detailed knowledge about their background. The agencies were provided with respondents’ answers to open-ended questions about occupation, which had been recorded verbatim by the interviewers. Furthermore, ancillary variables were provided for use by coders during the coding process, including age, gender, formal education, status in employment, public sector employment, supervisory tasks, and industry. Three agencies (labeled here as Agency A, Agency B, and Agency C) coded the responses to the ALLBUS open-ended questions about occupation; two agencies (Agency A and Agency B) coded the PIAAC data. Whereas Agency A was the same agency in both cases, Agency B was not. Agency B refers, in fact, to two agencies: The first agency coded the PIAAC occupation data; its successor coded the ALLBUS data. As the first agency handed over all relevant material to the second agency, and the second agency also adopted the first agency’s coding procedures, they are treated as one unit here.

The agencies were not given any further instructions on how to code. Rather, they could apply their typical coding procedures. They could use either semi-automatic or manual coding, and they were allowed to use any coding tools they had available. Furthermore, the agencies themselves decided on the coding strategy to be adopted, that is, 1) the number of coders; 2) the additional material provided (e.g., the complete ISCO manual or sample lists of possible codings); and 3) the organization of the coding process (e.g., coding the major group first and coding the other digits in a second step).

For the coding process, Agency A used a software program that provided the text of the responses. There was no automatic coding involved in the process. However, the coding software offered coding suggestions via templates. These suggestions are based on the

ISCO group labels (e.g., coding digit 1, the suggestions are based on the ISCO major groups, coding digit 4, the suggestions are unit group labels). The main objective of employing this software was to reduce the complexity of the coding process for the coders by using a hierarchical coding strategy, providing the same information to all coders. The coders were not given the ISCO manual as a reference. The coding instructions, ancillary information, conventions, and known problems with suggested solutions were part of an interface in the software.

Agency B used, as a first step, automatic coding based on an extensive dictionary. About 45 to 55% of the answers can usually be coded this way (Geis and Hoffmeyer-Zlotnik 2000, 127). Responses that were not automatically codeable were listed alphabetically, and alphabetically ordered blocks were randomly assigned to coders. In the case of the coding of the PIAAC data by Agency B, an expert coder reviewed the codes assigned, and if systematic errors were observed, coders were given additional training. Several rules were specified for the coders, for example: 1) If two different codes from different major groups are plausible, assign the occupation with a lesser degree of professionalization. 2) If two occupations were mentioned by a respondent, code the first-mentioned occupation. 3) If two different job titles were mentioned, code the more concrete title. The first rule seems to contradict the ILO rule cited earlier, whereby, in the case of several plausible alternatives, the occupation with the highest skill level should be coded. However, we are not sure how “professionalization” was operationalized. To our knowledge, this agency used auxiliary information on employment status in ambiguous cases. We do not know how the two other agencies handled this status variable during coding.

The coding process at Agency C involved two steps: first, occupations were coded according to the five-digit German Classification of Occupations (KldB) 2010 (Bundesagentur für Arbeit 2011). This was initially done automatically, using a dictionary with around 100,000 entries. This coding process was supported by taking into account the ancillary information provided (for example, industry or occupational status). The remaining responses were then manually coded. Several rules were applied, for example: 1) If two occupations were reported, code the first-mentioned occupation. 2) If the open-ended response is inadequate to determine the occupational code, use the ancillary information. 3) If two different codes are equally plausible, assign the code that occurs more frequently in practice. As a second step, the codes of the German Classification of Occupations (KldB) 2010 were automatically mapped to ISCO-08 codes using correspondence tables provided by the German Employment Agency (Bundesagentur für Arbeit 2011).

To control the coding quality within the agencies, and to determine whether low coding reliabilities were due to house effects, we asked the two agencies coding the PIAAC data to code a subsample of answers a second time, assigning them to different coders.

#### **4. Results**

In the following sections, we describe our results in the order of our four hypotheses. First, we present results on uncodeable or not completely codeable answers, followed by the description of inter-agency and inter-coder reliability and coding differences. After that, we present consequences for socioeconomic status and prestige scores and finally, we report differences in answer length and reliability.

4.1. *Uncodeable or Not Completely Codeable Answers*

The number of uncodeable answers was low; to a large extent answers could be coded at the four-digit level (between 78.0% and 97.8%). As can be seen in Table 4, the percentages of answers that could not be assigned to at least a major group by the three agencies that coded the ALLBUS data were 3.4%, 3.0%, and 2.2% respectively. For PIAAC, the respective percentages were 4.9% and 1.8%. These findings are in line with the findings of Hoffmeyer-Zlotnik et al. (2004), who reported 3% of not codeable answers in a survey of the German Environment Agency in 1999. Looking closer at the answers that could not be coded, we found large differences between the agencies: For ALLBUS, 274 answers were classified as uncodeable by one of the agencies, 51 answers were classified as uncodeable by two agencies, but only 28 answers were classified as uncodeable by all three agencies. A clearer pattern could be observed for PIAAC, with Agency A showing a greater tendency to classify a given answer as uncodeable. Agency A could not code 134 of the cases that Agency B coded, whereas Agency B was unable to code only eight cases that Agency A coded. If we look at answers that were coded, but not to the most detailed four-digit level, we find that these sum up to 16% and 19% for ALLBUS and 11% and 9% for PIAAC. These percentages were, thus, lower than we expected (see hypothesis H1). Agencies A and B seem to follow the recommendation of the ILO (2012, 56) to code vague responses to the most detailed level still supported by the information provided. In addition, we found surprisingly large differences with respect to which answers were classified as uncodeable. Here, the different rules of the agencies become obvious. To give an example, one agency classified the response “housewife and typist” as uncodeable (housewife), whereas the others assigned the code for typist (4131). Agency C is a special case in this regard: because they first coded the answers into the Germany Classification of Occupations (KldB) and then recoded these codes to ISCO-08 using the official cross-walk, they used the most probable ISCO unit group when answers were not completely codeable or when the KldB code was not unambiguously transferable.

4.2. *Inter-Agency and Inter-Coder Reliability and Coding Differences*

For the analysis of the reliability we included all codings of the three agencies. If an agency marked an answer as uncodeable or not codeable on the four-digit level, whereas the other one coded the same answer with a four-digit code, this could be handled as a mismatch in the sense of coding reliability. As can be seen in Table 5, coding agreement

Table 4. *Number of codeable digits (in percent).*

	ALLBUS			PIAAC	
	Agency A	Agency B	Agency C	Agency A	Agency B
Not codeable	3.4	3.0	2.2	4.9	1.8
1 digit coded	3.7	4.4	0.0	1.4	2.3
2 digits coded	8.2	7.5	0.0	6.0	1.6
3 digits coded	3.7	7.3	0.0	3.4	4.9
4 digits coded	81.0	78.0	97.8	84.3	89.3

ALLBUS: N = 5,130; PIAAC: N = 4,159.

Table 5. Inter-agency reliability (Cohen's kappa).

	ALLBUS			PIAAC
	Agencies C & B	Agencies A & C	Agencies A & B	Agencies A & B
1 digit	0.683	0.685	0.722	0.760
2 digits	0.644	0.632	0.674	0.715
3 digits	0.572	0.528	0.574	0.630
4 digits	0.506	0.475	0.508	0.566

ALLBUS: N = 5,130; PIAAC: N = 4,159.

between the different agencies was far from perfect. We only report  $\kappa$ , because checking the proportion of agreement showed that the differences between these measures are negligible. For ALLBUS, in a comparison of Agency A and Agency B, the inter-agency reliability of assigning an open-ended answer to the same major group (i.e., the same first digit) was 0.72. With  $\kappa = 0.68$ , the inter-agency coding reliabilities for major groups between Agency A and Agency C and between Agency B and Agency C were even lower. The results are similar for PIAAC: The inter-agency reliability between Agency A and Agency B was around 0.76 for the major groups. As there are only ten major groups, these inter-agency reliabilities are not satisfactory.

Because of the hierarchical structure of ISCO, once the first digit differs, all more detailed codes differ. Thus, reliabilities for sub-major, minor, and unit groups were even lower. In the end, the assigned four-digit ISCO codes differed in nearly half of the cases for ALLBUS ( $\kappa = 0.48$  to  $0.51$ ), and inter-agency agreement was only slightly higher for PIAAC ( $\kappa = 0.57$ ). This low level of reliability casts serious doubt on the usability of these data for further analysis.

As Table 5 shows, the largest information loss occurred at the level of the major group. Taking the ALLBUS codings as an example, a glance at the distribution of the major groups (Figure 1) reveals systematic differences resulting from the codings of the three agencies involved. In particular, frequencies of Major Groups 3, 5, 8, and 9 differ considerably.

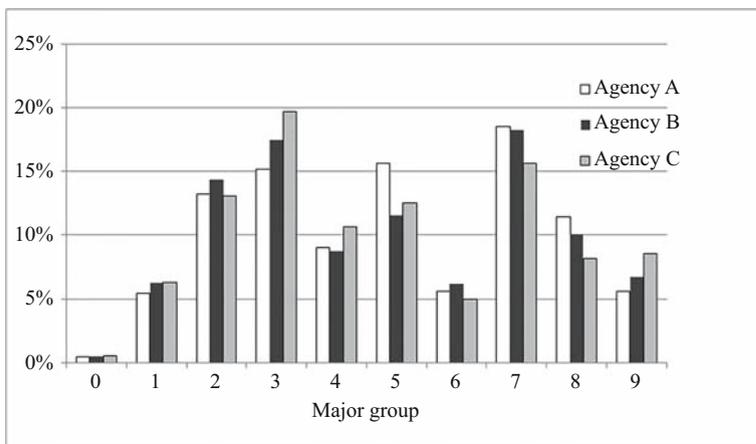


Fig. 1. Percentage distribution of major groups in ALLBUS by coding agency.

The difference is most extreme in Major Group 9, where the number of elementary occupations coded by Agency C was 1.5 times higher than that coded by Agency A.

In what follows, we briefly explore some of the systematic discrepancies we observed in the assignment of the first digit by the different agencies (see supplementary material, Figures A1a and A1b, for relative frequencies of the observed differences on Major Group level; for the most frequent unit group combinations among these first digit discrepancies, see Supplemental data, Table A3). Because it is impossible to list all the differences we found, we concentrate on those differences observed most often and on those that have a large effect on the prestige and SES scores derived from the codes. We identified six such systematic differences:

The first difference related to the handling of self-employed respondents (see Supplemental data, Table A4a). For ALLBUS and PIAAC, Agency B coded a large group of these respondents into Major Group 1: Managers, whereas the other agencies often assigned them to Major Group 5: Services and Sales Workers. To give some examples: if a respondent reported that he or she was a self-employed hairdresser, Agency B classified the occupation into Unit Group 1120: Managing Directors and Chief Executives, whereas the other agencies assigned it to Unit Group 5141: Hairdressers (see Supplemental data, Table A4a). A similar problem was caused by the answer “innkeeper”. Agency B classified it into Unit Group 1412: Restaurant Managers; the other agencies assigned it to Sub-major Group 51: Personal Service Workers. ISCO defines “Operators of small cafés, restaurants and bars to whom the management and supervision of staff is not a significant component of the work are classified in Unit Group 5120 . . .” (ILO 2012, 238). The rules are clearly defined but the answer of the respondents are often not detailed enough to decide what the main task is.

A second frequently occurring difference between the codings for ALLBUS is related to the classification of manual workers. Depending on the agency, these workers were coded either into Major Group 6: Skilled Agricultural, Forestry and Fishery Workers; Major Group 7: Craft and Related Trades Workers; Major Group 8: Plant and Machine Operators and Assemblers; or Major Group 9: Elementary Occupations. Agency C assigned blue-collar workers (e.g., chemical workers, metal workers, or railroad workers) much more often to Major Group 9, whereas, whenever possible, the other two agencies assigned them to a major group other than Major Group 9 (see Supplemental data, Table A4b). Differences arose because many of the responses were not unambiguously codeable. Major Group 6, 7, or 8 would be appropriate if more complex tasks were performed, whereas Major Group 9 should be reserved for those occupations involving only simple and routine tasks.

The third typical systematic coding difference which occurred both in the ALLBUS and the PIAAC codings related to the assignment of responses to Major Groups 3 and 4. Major Group 3 comprises technicians and associate professionals who “perform mostly technical and related tasks connected with research and the application of scientific or artistic concepts and operational methods, and government or business regulations” (ILO 2012, 169). Major Group 4 comprises clerical support workers who “record, organize, store, compute and retrieve information, and perform a number of clerical duties in connection with money-handling operations, travel arrangements, requests for information, and appointments” (ILO 2012, 219). Whereas occupations in Major Group 3 typically require

completion of upper secondary education and possibly a higher education degree (Skill Level 3), most occupations in Major Group 4 require only completion of lower secondary education and possibly vocational training (Skill Level 2). Frequent and typical answers that caused the coding problem in question were “secretary”, “clerk” (*Sachbearbeiter*), and “accountant” (see a list of most frequently observed unit group combinations among first-digit discrepancies in Supplemental data, Table A3). To give an example, secretaries may, on the one hand, be what are referred to in ISCO-08 as “administrative and specialized secretaries”, who “provide organizational, communication and documentation support services, utilizing specialized knowledge of business activity of the organization in which they are employed” (ILO 2012, 202). In this case, they should be classified into Major Group 3. On the other hand, secretaries may be “secretaries (general)”, who “transcribe correspondence and other documents, check and format documents prepared by other staff, deal with incoming and outgoing mail” and so on, in which case they should be classified into Major Group 4 (ILO 2012, 221). Without more detailed information about the respondents’ tasks and duties, the occupations cannot be coded with certainty. However, the coding of these answers is of great relevance because of their high frequency and the low inter-agency agreement in relation to these two major groups (around 70%, depending on the coding agencies compared).

The fourth systematic coding difference occurred – again for both surveys – when assigning responses to Major Group 7: Craft and Related Trades Workers, and Major Group 8: Plant and Machine Operators and Assemblers. The main difference between these two major groups is the operation or use of machines at work. Although in the modern industrial world, most crafts require the use of machines, it is not clear from the respondents’ answers whether or not the work was done predominantly with or without machines. Thus, the occupation “metal worker” may be classified into Sub-major Group 72: Metal, Machinery and Related Trades Workers or into Minor Group 812: Metal Processing Plant Operators, depending on the stage of development of the industry in the national context.

The fifth systematic coding difference occurred when coding occupations such as educators (*Erzieher*) working in early childhood education and care or youth welfare services. Agency A assigned these occupations mostly to Major Group 5: Services and Sales Workers, Unit Group 5311: Child Care Workers; Agency B assigned them to Major Group 2: Professionals, Unit Group 2342: Early Childhood Educators; and Agency C assigned them to Major Group 3: Technicians and Associate Professionals, Unit Group 3412: Social Work Associate Professionals. At least part of the confusion probably stemmed from the difference between the educational requirements for these occupations in Germany and elsewhere. Whereas in Germany these occupations usually require qualifications corresponding to Skill Level 3, most other countries appear to require Skill Level 4. The ISCO manual is quite clear on this point, stating that “occupations that require the performance of similar tasks should be classified in the same group” and that primary and pre-primary teachers “should all be classified in Major Group 2” (ILO 2012, 28). However, this innovation introduced for ISCO-08 does not seem to have been acknowledged by all coding agencies.

The sixth coding problem was caused by a special national situation in Germany, where – as in some other countries – one type of vocational education and training is “dual

vocational training”, which comprises theoretical and practical elements. Most careers in crafts start with such a training program. After successful completion of this program, further training is possible in order to obtain the qualification “Meister” (master craftsman or engineering technologist). This type of career is very common in Germany. When asked what their occupation was, many respondents answered that they were, for example, a master carpenter or a master electrician. Although no additional information was available on the tasks performed by these master craftsmen, Agency C coded such answers into Unit Group 3122: Manufacturing Supervisors, whereas the other two assigned a code from Major Group 7: Craft and Related Trades Workers that described the craft the respondent was working in. However, the required skill level in Major Group 7 is lower than that in Major Group 3. The reason for the different handling of this type of occupation may result from the coding in KldB and transferring the codes to ISCO: in KldB 2010 all job titles containing the word “Meister” are coded as supervisory tasks, which results in the 3122-coding. Once again, the difference in coding shows that agencies create specific rules to handle national specificities that are not reflected in ISCO. In this case, Agency C applied a standardized rule using the public and official German classification as their standard. As there is no concerted effort to standardize these rules, agency-specific approaches result in different codings and lower inter-agency reliability.

The reliabilities reported in [Table 5](#) reflect not only differences between the agencies’ coding rules, but also differences between coders. To disentangle house effects from coder effects, we asked the two agencies that coded the PIAAC data to code a subsample of the occupational information twice, assigning the answers to different coders. We interpret the difference between the inter-coder and inter-agency reliabilities as a house effect that reflects differences in the coding procedures and rules. Inter-coder reliability ( $\kappa$ ) of the four-digit codings of our PIAAC subsample was 0.84 between the two coders deployed by Agency B, and between 0.68 and 0.74 at Agency A, where coding was carried out by five coders. As these reliabilities were considerably higher than those between the agencies (0.57), this result supports our assumption that the coding procedures and rules applied by a coding agency affect ISCO coding to a considerable extent. This is in line with hypothesis H2.

#### 4.3. Consequences for Socioeconomic Status and Prestige Scores

As mentioned previously, occupational prestige or SES scores are often derived from ISCO codes. Therefore, we analyzed the extent to which the reliability of these scores was affected by the (comparatively low) reliability of the ISCO codes. Our analysis focused on the International Socio-Economic Index (ISEI) and the Standard International Occupational Prestige Scale (SIOPS), both of which were developed for cross-national research ([Ganzeboom and Treiman 2003](#)). Using the syntax proposed by [Ganzeboom and Treiman \(2012\)](#), we calculated these scores separately for each of the coding results of the agencies. In less than 5% of the coded occupations, it was not possible to assign an SES or occupational prestige score. Although the inter-agency reliability of the four-digit ISCO codes was only around 0.5, the consistency between the derived socioeconomic status and prestige scores was much higher than expected. For ISEI, the correlation between each pair of agencies was 0.90, which is higher than the correlations reported in the study by

Maaz et al. (2009). For ALLBUS, the correlation of SIOPS between Agency A and Agency B was 0.84, between Agency B and Agency C it was 0.82, and between Agency A and Agency C it was 0.85. For PIAAC, the correlation between Agency A and Agency B was 0.84. The high correlations of both SES and prestige scores confirm hypothesis H3.

#### 4.4. Differences in Answer Length and Reliability

In a final step, we analyzed the length of the answers given to the question about occupation in the two surveys. Because of the differences on how the questions on occupations were phrased in PIAAC and ALLBUS (see Subsection 3.1), we assume that PIAAC answers were more detailed, that is, longer. We measured answer length by counting the total number of words after combining all sub-questions. One drawback of this method is that we were not able to identify word repetitions. Thus, if respondents repeated words, these were counted each time.

Our analysis showed that the questions in PIAAC were answered, in fact, in more detail, leading to answers that were about twice as long as the answers in ALLBUS. The average number of words recorded by interviewers was 2.70 (*SD*: 2.22) for ALLBUS, with a minimum of one and a maximum length of 23 words. For PIAAC, the average number of words recorded was 5.52 (*SD*: 3.87), with a minimum of one and a maximum of 48 words. Furthermore, when systematically checking very common answers, such as “secretary” and “clerk”, it became obvious that in PIAAC, as compared to ALLBUS, most respondents had given more details than just a general job title. For example, PIAAC respondents more frequently added information such as “personal assistant to the CEO” to the general job title “secretary”, or “office clerk in accounting” to the job title “clerk”.

However, the crucial question is whether longer answers result in better – that is, more reliable – codings. To investigate this question, we calculated reliability coefficients based on length of answer (see Table 6). We found that the inter-agency reliability ( $\kappa$ ) did indeed vary with the length of the answer. However, contrary to what one might intuitively expect, reliabilities decreased as answers became longer, thus confirming hypothesis H4. This is in line with the finding by Conrad et al. (2016). A closer inspection of long answers (> 15 words) did not yield any specific clues as to the reasons why they appear to be more difficult to code. We suspect that fewer words might result in incomplete occupational

Table 6. Inter-agency reliability (Cohen’s kappa) by length of answer.

Answer length	ALLBUS				PIAAC	
	Answers in %	Agencies B & C	Agencies A & C	Agencies A & B	Answers in %	Agencies A & B
1 word	32.1	0.54	0.53	0.54	2.6	0.63
2 words	33.7	0.54	0.53	0.56	22.8	0.61
3 words	9.9	0.49	0.45	0.48	6.7	0.58
4 words	9.3	0.46	0.44	0.47	15.1	0.57
5 words	6.0	0.47	0.37	0.44	13.8	0.57
6 words	3.0	0.39	0.32	0.40	10.1	0.57
> 6 words	6.0	0.37	0.36	0.43	28.9	0.51

ALLBUS: N = 5,130; PIAAC: N = 4,159.

descriptions but that, for frequently used occupational titles, coders have internal coding rules for handling these (partially incomplete) answers. If the answers are longer, coders have more scope for decision-making, and simple rules can no longer be applied. Less consistent coding of longer answers may also be due to the fact that coders are reluctant to read longer texts and therefore base their coding on only parts of the answer.

Although this result is true for both ALLBUS and PIAAC we found, in general, that occupations in PIAAC were more reliably coded than in ALLBUS. For example, the reliability between Agency A and Agency B for one-word ALLBUS answers was 0.54, whereas it was 0.63 for one-word PIAAC answers (Table 6). Interviewer training and a more detailed question seem to have led to longer, more precise answers.

Finally, if longer answers are more difficult to code, it could be assumed that they are also more often uncodeable or not completely codeable. However, our findings do not support this assumption. Uncodeable or incompletely codeable answers were, on average, shorter than codeable answers: the means for completely codeable answers were 2.76, 2.71, and 2.71, while the means for not completely codeable answers were 2.26, 2.39, and 2.30 for Agencies A, B, and C respectively.

## 5. Conclusion

To better understand how the process of occupational coding affects outcome quality, we studied occupational data from two surveys, each coded by several agencies. We found that the share of uncodeable answers was comparatively small, ranging from 1.8% to 4.9%, depending on the coding agency and data source. However, a closer look revealed that what was deemed “uncodeable” varied between coding agencies. In a next step, we examined the coding results more closely and discovered characteristic, systematic differences in the way agencies interpreted ISCO and defined relationships between answers to be coded and ISCO categories. Notwithstanding the comparatively low agreement between occupational codings from the different agencies, the correlations for SES and occupational prestige scales were satisfactory (between 0.82 and 0.90). Finally, we observe a negative correlation between length of answers (number of words) and coding reliability both in ALLBUS and PIAAC. At the same time, coding reliability between the two agencies (A and B) that coded data from both surveys was higher for PIAAC than for ALLBUS. This difference remained when we controlled for length of answer. In line with Elias (1997, 13) we discuss our findings in relation to the following aspects: the *data to be coded*, the *coding rules and coding process*, and the *classification*.

With regard to the *data to be coded* the seemingly contradictory results concerning answer length are the most interesting. Several factors may have contributed to these findings. First, the intensive interviewer training in PIAAC may not only have increased answer length – with a general negative effect on reliability – but might also have led to “better” answers because interviewers were thoroughly trained on what kind of information was useful for occupational coding. Especially with respect to some frequently reported occupations that are difficult to code without additional information (e.g., educators, secretaries, or clerks) the interviewer training may have had, on balance, a positive effect on reliability. The higher percentage of four-digit codes for PIAAC data also points to the effect of interviewer training.

The second factor that may have contributed to the higher inter-agency reliability of PIAAC codings are differences in the way occupational information was collected. In particular, the first sub-question in PIAAC explicitly asking for the exact job title, may have elicited answers that were more suitable for ISCO coding. In addition, the detailed question about employment status that preceded the open-ended question about occupation in ALLBUS may have played a crucial role. This may have affected the answers given to the open-ended questions because respondents had just given some of this information in the preceding question on their employment status. Besides effects caused by the resulting differences in the core input material for occupational coding, it seems reasonable to assume an interaction effect of data provided and coding rules. The use of this ancillary information for occupation coding seems to have varied in extent or manner between the agencies involved. This could be an important factor that explains the lower inter-agency reliabilities observed for ALLBUS compared to PIAAC, where the question on employment status only differentiated between employed and self-employed.

Taking all these factors together, our results concerning the ALLBUS-PIAAC differences suggest that some of the inter-agency coding disagreements result from various differences in the survey instrument and the interviewer training, and the ensuing differences in the occupational data to be coded – partly in interaction with coding rules applied concerning this data. Therefore, identifying and understanding those characteristics of question wording and interviewer training that lead to the most suitable input material for occupational coding would be a promising field for further research.

Turning to the *coding rules and coding process*, we argue that the systematic deviations that we observed between codings from different agencies indicate that the rules and procedures laid down in the ISCO manual (ILO 2012) do not suffice to completely cover the coding process. This results in low inter-agency reliability and low validity of the results.

From a quality perspective, (inter-coder) reliability is important, but validity of codes is essential. Hak and Bernts (1996) pointed out that the validity of the coding process depends on the quality of coding instructions, whereas inter-coder reliability depends on the implementation of these instructions. This implies that insufficient reliability and validity of occupational codes have common sources. To elicit both reliable and valid information, interviewers should be trained to filter as little information as possible and to ask for an occupational title. Moreover, they should be familiarized with basic coding procedures (Cantor and Esposito 1992, 665).

To improve coding quality, agencies engaged in occupational coding create rules of their own that reflect their interpretation of ISCO, and these house rules lead to the large differences in coding decisions that we found. The shortcomings of the ISCO manual and the secondary, agency-specific coding rules are a serious threat to the validity of occupational codes. We assume that this threat is even larger in the cross-national context where these agency-specific rules most likely reflect particularities of national labor markets. To overcome this situation, we suggest that coding agencies state the rules they apply as clearly as possible and make them public. A publicly accessible body of rules would allow more systematic discussion and development by those engaged in occupational coding. For the time being, survey practitioners commissioning coding services should ask agencies about the coding rules they apply and, where necessary,

negotiate such rules. One may think that the problems resulting from differences in rules may be overcome by applying completely automatic coding routines. Indeed, such methods have been proposed, for example, based on statistical learning (for a discussion of such approaches, see Gweon et al. 2017). However, up to now these approaches are not able to code considerably more than half of the occupations, thus limiting their usability.

With respect to the *classification*, some problems related to coding ISCO originate from the national context in which it is applied, for example, “Meister” (master craftsman) in Germany. It can be expected that similar problems arise in other national contexts, as ISCO can be seen as a compromise between different national views and peculiarities (see Desrosières 1996, 19–21 for a concise description of the development of international comparative statistics). We have also seen that agencies make fundamentally different decisions when coding occupations. We were particularly surprised to see that there are considerable differences in the use of major groups, pointing to “boundary problems”. A further improvement of the ISCO manual, by including even more precise definitions and more detailed explanations, could certainly be helpful in dealing with problematic distinctions.

Given the multitude of jobs actually performed in the real world, rules and procedures laid down in a manual will never suffice to completely cover the coding process. Approaches supporting the coding of ISCO directly in the field may help to avoid some of the problems. The idea behind in-field coding is that respondents have better and more detailed knowledge of the type of work they carry out. Based on additional variables, such as employment status, public sector employment, or supervisory tasks, they could be asked for more details depending on their answer to the first question about occupation. Initial results of such approaches (Tijdens 2014, 2015; Schierholz et al. 2017) are promising. When discussing developments in the way occupational data is assessed and coded, it is also important to look into the effect that new approaches have on the survey process. In-field coding would introduce additional burden on respondents and interviewers, and increase survey costs. Nevertheless, improving results of occupational coding should be a central aim of efforts to improve the overall quality of surveys.

## 6. References

- Belloni, M., A. Brugiavini, E. Meschi, and K. Tijdens. 2016. “Measuring and Detecting Errors in Occupational Coding: An Analysis of SHARE Data.” *Journal of Official Statistics* 32(4): 917–945. Doi: <http://dx.doi.org/10.1515/JOS-2016-0049>.
- Bergmann, M.M. and D. Joye. 2005. “Comparing Social Stratification Schemata: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright.” *Cambridge Studies in Social Research* 10: 1–35. Available at: <https://www.sociology.cam.ac.uk/research/srg/cs10> (accessed January 2019).
- Billiet, J. and G. Loosveldt. 1988. “Improvement of the Quality of Responses to Factual Survey Question by Interviewer Training.” *Public Opinion Quarterly* 52: 190–211. Doi: <http://dx.doi.org/10.1086/269094>.
- Bundesagentur für Arbeit. 2011. *Klassifikation der Berufe 2010. Systematischer und alphabetischer Teil mit Erläuterungen*. Nürnberg: Bundesagentur für Arbeit.

- Bushnell, D. 1998. "An Evaluation of Computer-Assisted Occupation Coding." In *Proceedings of the International Conference New Methods for Survey Research*, August 21–22, 1998: 23–26. Chilworth Manor, Southampton, United Kingdom.
- Campanelli, P., K. Thomson, N. Moon, and T. Staples. 1997. "The Quality of Occupational Coding in the United Kingdom." In *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin, 437–453. New York: John Wiley & Sons, Inc.
- Cantor, D. and J.L. Esposito. 1992. "Evaluating Interviewer Style for Collecting Industry and Occupation Information." *Proceedings of the Section on Survey Methods, American Statistical Association*: 661–666.
- Conrad, F.G., M.P. Couper, and J.W. Sakshaug. 2016. "Classifying Open-Ended Reports: Factors Affecting the Reliability of Occupation Codes." *Journal of Official Statistics* 32(1): 75–92. Doi: <http://dx.doi.org/10.1515/JOS-2016-0003>.
- Desrosières, Alain. 1996. "Statistical Traditions: An Obstacle to International Comparisons?" In *Cross-National Research Methods in the Social Sciences*, edited by L. Hantrais and S. Mangen, 17–27. New York: Cassel.
- Elias, P. 1997. "Occupational Classification (ISCO-88): Concepts, Methods, Reliability, Validity and Cross-National Comparability." *OECD Labour Market and Social Policy Occasional Papers* 20. Doi: <http://dx.doi.org/10.1787/304441717388>.
- Freelon, D.G. 2010. "Recal: Intercoder Reliability Calculation as a Web Service." *International Journal of Internet Science* 5(1): 20–33. Available at: [http://www.ijis.net/ijis5\\_1/ijis5\\_1\\_freelon.pdf](http://www.ijis.net/ijis5_1/ijis5_1_freelon.pdf) (accessed January 2019).
- Ganzeboom, H.B.G. and D.J. Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research* 25: 201–239. Doi: <http://dx.doi.org/10.1006/ssre.1996.0010>.
- Ganzeboom, H.B.G. and D.J. Treiman. 2003. "Three Internationally Standardised Measures for Comparative Research on Occupational Status." In *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*, edited by J.H.P. Hoffmeyer-Zlotnik and C. Wolf, 159–193. New York: Kluwer Academic/Plenum Publishers.
- Ganzeboom, H.B.G. and D.J. Treiman. 2012. "International Stratification and Mobility File: Conversion Tools." Amsterdam: Department of Social Research Methodology. Available at: <http://www.harryganzeboom.nl/ismf/index.htm>. Retrieved 2017/02/27.
- Geis, A. and J.H.P. Hoffmeyer-Zlotnik. 2000. "Stand der Berufsvercodung." *ZUMA-Nachrichten* 47: 103–128. Available at: [https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma\\_nachrichten/zn\\_47.pdf](https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_47.pdf) (accessed January 2019).
- Gweon, H., M. Schonlau, L. Kaczmirek, M. Blohm, and S. Steiner. 2017. "Three Methods for Occupation Coding Based on Statistical Learning." *Journal of Official Statistics* 33(1): 101–122. Doi: <http://dx.doi.org/10.1515/jos-2017-0006>.
- Hak, T. and T. Bernts. 1996. "Coder Training: Theoretical Training or Practical Socialization?" *Qualitative Sociology* 19(2): 235–257. Doi: <http://dx.doi.org/10.1007/BF02393420>.

- Hoffmann, E., P. Elias, B. Embury, and R. Thomas. 1995. *What Kind of Work Do You Do? Data Collection and Processing Strategies When Measuring "Occupation" for Statistical Surveys and Administrative Records*. Geneva: ILO.
- Hoffmeyer-Zlotnik, J.H.P., D. Hess, and A. Geis. 2004. "Computerunterstützte Vercodung der International Standard Classification of Occupations (ISCO-88)." *ZUMA-Nachrichten* 55: 29–52. Available at: [https://www.ssoar.info/ssoar/bitstream/handle/document/20762/ssoar-zuma-2004-55-hoffmeyer-zlotnik\\_et\\_al-computerunterstutzte\\_vercodung\\_der\\_international\\_standard.pdf?sequence=1](https://www.ssoar.info/ssoar/bitstream/handle/document/20762/ssoar-zuma-2004-55-hoffmeyer-zlotnik_et_al-computerunterstutzte_vercodung_der_international_standard.pdf?sequence=1) (accessed January 2016).
- International Labour Office (ILO). 2012. *International Standard Classification of Occupations 2008 (ISCO-08): Structure, Group Definitions and Correspondence Tables*. Geneva: ILO.
- Maaz, K., U. Trautwein, C. Gresch, O. Lüdtke, and R. Watermann. 2009. "Intercoder-Reliabilität bei der Berufscodierung nach der ISCO-88 und Validität des sozioökonomischen Status." *ZfE* 12: 281–301. Doi: <http://dx.doi.org/10.1007/s11618-009-0068-0>.
- Schierholz, M., M. Gensicke, N. Tschersich, and F. Kreuter. 2017. "Occupation Coding During the Interview." *Journal of the Royal Statistical Society A* 181: 379–407. Doi: <http://dx.doi.org/10.1111/rssa.12297>.
- 't Mannetje, A. and H. Kromhout. 2003. "The Use of Occupation and Industry Classifications in General Population Studies." *International Journal of Epidemiology* 32: 419–428. Doi: <http://dx.doi.org/10.1093/ije/dyg080>.
- Tijdens, K. 2014. "Dropout Rates and Response Times of an Occupation Search Tree in a Web Survey." *Journal of Official Statistics* 30(1): 23–43. Doi: <http://dx.doi.org/10.2478/jos-2014-0002>.
- Tijdens, K. 2015. "Self-Identification of Occupation in Web Surveys: Requirements for Search Trees and Look-up Tables." *Survey Insights: Methods from the Field*. Doi: <http://dx.doi.org/10.13094/SMIF-2015-00008>.

Received August 2017

Revised February 2018

Accepted April 2018