MEASURING OCCUPATIONS IN WORLDWIDE WEB-SURVEYS

KEA TIJDENS
University of Amsterdam/AIAS, Netherlands

Abstract: This paper summarizes the design principles underlying the WISCO Database of Occupations for the measurement of occupations in multi-country web-surveys by means of self-identification. It is discussed why the Database has been designed, and its source list, search tree and translations. The Database holds almost 1,600 occupational titles. Using a wide variety of sources, the list of occupational titles has been compiled as part of the FP6 funded EurOccupations project. It is explained how the Database deals with issues such as skill levels, corporate hierarchies, job ladders, managerial and supervisory occupations, craft versus manufacturing occupations, composite occupations and the methods used for translations. Finally, the paper sketches briefly how the database is used in the worldwide WageIndicator web-survey. For more information see www.euroccupations.org or www.wageindicator.org

Key words: occupations, measurement, search tree, web surveys, database.

1. Introduction

Occupation is a key variable in socio-economic research. In paper-based, telephone or face-to-face surveys, it is mostly asked in an open response format. In contrast, web-surveys offer a unique possibility for a closed response format, using a search tree. As part of the EU-FP6 funded EurOccupations project (2006-2009, nº 028987, www.EurOccupations.org), a free downloadable database of occupations, designed for eight EU-member states. The project aimed to provide a tool for self-identification in surveys, measuring occupations with a greater precision than the 4-digit units of the International Standard Classification of Occupations ISCO by adding further digits. It did not aim for revising ISCO or any other occupational classification. In co-operation with the continuous, worldwide WageIndicator web-survey, www.wageindictor.org, usability tests have been performed and occupational titles for countries have been added. Thanks to these two projects, the **W**orld database of **ISCO** occupations WISCO could be drafted, including:

A source list of 1,594 distinct occupational titles in English
Country-specific translations of these occupational titles
A 3-tier search tree that allows respondents to navigate through the database

In many ways, the social sciences may profit from the WISCO Database of Occupations. First, when used in multi-country web-surveys, it will increase comparability of the occupation variable across countries. Second, when used in web-surveys with large sample sizes, the detailed occupational titles allow for analyses of sub-samples previously not possible. Third, the database can be used in computer-assisted face-to-face surveys, when the interviewer turns the screen to the respondent.

2. The procedure of drafting and testing the WISCO Database of Occupations

Defining occupations

The primary aim of WISCO Database of Occupations is its use for valid self-identification of occupation in web-surveys. Given that respondents prefer to indicate their job titles rather than aggregated categories, the source list of occupational titles had to be close to the wording used in job titles, thus requiring a long list of occupational titles. However, these occupations have to be distinct from each other, because synonyms or overlapping occupational titles may confuse respondents. Yet, the longer the list, the higher the average respondents' reading-time and the higher the likelihood of dropout during survey completion. The source list has to optimise between the demand to include as many distinct occupational titles as possible to facilitate valid self-identification and the demand to be as brief as possible to reduce reading time.

The length of the source list is further determined by the search tree, consisting of a 2- or 3-tier tick list, detailing broad categories in the 1st tier to detailed items in the 2nd and 3rd tier. To prevent visitors from scrolling, a standard search tree on a computer screen can cope with some 20 items in the 1st tier and up to 20*20=400 items in the 2nd tier and 20*20*20=8,000 items in the 3rd tier. As 400 occupational titles definitely are too few, the search tree had to consist of 3-tiers with a limit of 8,000 occupational titles. A third argument for an efficient source list is the number of jobholders. Occupations with few jobholders are preferably not included, whereas occupations with large numbers of jobholders are preferably broken down in two or more occupational titles. As a rule of thumb, we used a 0.01% limit of the labour force. Large occupations have been been broken down into several distinct occupational titles.

The WISCO Database of Occupations employs the following definition: "*an occupation is* a bundle of job titles, clustered in such a way that survey respondents*, that respondents in a valid way will recognize as at their job title; an occupation identifies a set of tasks distinct from another occupation; an occupation should have at least a not-negligible*

number of jobholders and it should not have an extremely large share in the labour force".

A stepwise procedure to draft the source list

Since 2001, when the WageIndicator web-survey started in the Netherlands, an occupation search tree has been developed gradually, using the NOC of Statistics Netherlands. In 2004, when the web-survey expanded to surrounding countries, new search trees and source lists were applied. Some countries preferred different search trees, other countries had occupation lists from their statistical offices, but these were not translated in English. So, a search tree and a source list to generate cross-country comparable data was lacking.

EurOccupations started in May 2006 and aimed at drafting an occupations database for the eight – at that time - largest EU member states, Belgium, France, Germany, Great Britain, Italy, the Netherlands, Poland, and Spain. Its source list should fully correspond with ISCO-08. In September 2006 draft 3 of ISCO-08 was published (ILO 2006). Its 447 occupational units at 4-digit level were taken as the point of departure for the source list of occupations. The occupations additionally specified in this draft were added to the list, if considered to be distinct from occupations already in the list. Occupations, listed in the detailed Alphabetical index of occupational titles for ISCO-88(COM), that could be expected to have large numbers of jobholders were added to the source list. Frequently reported occupations in the 2004-06 German and Netherlands in the additional open response question in the WageIndicator web-survey were added to the source list, particularly of importance for new and emerging occupations and for occupations that had to be detailed. Finally, national occupational classifications from UK, Belgium, USA and Canada were reviewed. Occupations that were assumed to have large numbers of jobholders were added to the source list. Early 2007, the source list held 1,433 occupational titles. This list was translated by professional translators and carefully checked by the national labour market experts of the EurOccupations and WageIndicator partners. The comments of translators and experts mostly related to occupational demarcation lines, e.g. two occupational titles in the source list were not considered distinct in the country at stake. In these cases, either one occupation was removed from the source list or one occupation was not included in the national list.

In Spring 2008, ILO published the final version of its ISCO-2008 classification with 433 occupational titles at 4-digit level (ILO 2007). Compared to the 2006 draft, the number of occupations was reduced and some occupations were assigned a different skill level. The WISCO source list was accordingly adapted and again critically reviewed with regard to internal consistency and suitability within the search tree. Finally, the source list was updated taken into account the results of the EurOccupations similarity test of the 150 occupations, selected from the initial EurOccupations source list. The final WISCO Database of Occupations source list counted 1,594 occupations (Tijdens and Jacobs 2009a). It is posted at the website www.euroccupations.org.

In Summer and Autumn 2009, the revised WISCO Database of Occupations was implemented for all countries in the WageIndicator web survey. In addition to the eight EurOccupations countries, the WISCO Database of Occupations currently holds translations for Brazil, Bulgaria, China, Czech Republic, Denmark, Greece, Finland, Hungary, Indonesia, Russia, Slovakia, Sweden, South Korea, and Turkey. National labour market experts checked translations for countries for which the language was available, e.g. Angola, Argentina, Azerbaijan, Botswana, Chile, Colombia, India, Malawi, Mexico, Mozambique, Paraguay, South Africa, United States, Zambia, and Zimbabwe. For 2010 translations for Hindi, Norwegian, Romanian and the three Baltic states are foreseen. Countries are allowed to add or remove occupations in their national lists, because the 0.01% limit of the labour force will vary across countries. If new occupations are added, the experts are asked to identify the related ISCO-4 digit occupation.

ISCO distinguishes four skill levels. It has no way to handle country-specific deviations of the skill levels. This challenged WISCO Database of Occupations to phrase occupational titles such that respondents will assess their appropriate skill level. EurOccupations has undertaken a systematic empirical investigation regarding the skill levels of 150 occupations in eight EU member states. The vast majority of these occupations did not vary largely with regard to job content, but to skill levels (see Deliverables D07-D22, 2009). In the years to come, an empirical measurement of required skill levels of occupations worldwide will be possible with the WageIndicator web-survey data, using the jobholder's attained educational level, the years of experience and the self-assessed match between attained education level and job level.

Usability test

By mid 2007 the search tree and the translations of the initial source list were implemented in the WageIndicator web-survey for three usability tests (Tijdens and Jacobs 2009b). Until 2008/Q4, in total 171,443 respondents from 33 countries responded. Test 1 was an inspection of the distributions over the ten 1-digit ISCO groups for seven EurOccupations countries. The results seem reasonable, and no weird outliers could be detected. Test 2 related to respondents' feed-back. In this period, the WageIndicator web-sites received over five thousand emails from visitors but less than 20 complained about the search tree. Their comments referred to not listed occupations and to the paths in the search tree and have been taken into account. For test 3, the distributions over the 3-digit ISCO-88 occupational groups have been compared between the web-survey data and

the 2001 ELFS for Poland, Spain and UK. This comparison is hampered because the years of survey are not similar, both variables are based on cross-over tables, notably from the NOC's into ISCO-88 (ELFS) and from ISCO-08_draft 3 into ISCO-88 (WageIndicator), and the sampling and survey modes are different, notably a random sampled face-to-face survey versus a volunteer web-survey. Taken into account these dissimilarities, the results are satisfactory, though better for the UK than for Poland, with Spain in between. The most striking differences reveal that the *Computing professionals* are largely overrepresented in WageIndicator and that the *Shop, stall and market salespersons and demonstrators* are underrepresented. The former might be explained from this groups' self-selection into the web-survey and from growing employment since 2001. The major lesson learned was that the shop assistant occupations the source list needed more detail.

3. The search tree and readability issues

When offering a choice-set with a large number of occupations, respondents must be able to navigate through the list. In web-surveys, three techniques facilitate choices from long item lists. First, an alphabetically sorted drop-down can be used, but this technique is limited to a few hundred items. Second, a search tree with two or three tiers can be used if the list exceeds a few hundred items, and this technique has been used for WISCO. Third, an open format question with text recognition can be used, but this assumes a database with occupational titles and their synonyms. The latter are currently not available in WISCO, but may be so in the near future.

In the WISCO Database of Occupations search tree, the 1st tier uses a mixture of broad occupational groups and industry groups, e.g. *Agriculture, nature, animals, environment* or *Care, children, welfare, social work*. Job-vacancy sites use similar job families. Since 2004, this 1st tier has been in use in the WageIndicator web-survey and it has proven to be a good entry. Note that this 1st tier is different from the 1-digit ISCO-08 major groups, as the ISCO hierarchy is designed for taxonomy purposes and not to facilitate self-identification. All 2nd tier items are nested in the 1st tier, whereby one 2nd tier item can be nested into two or more 1st tier items, if needed. The 3rd tier includes all source list occupations, again allowing multiple nesting. The search paths are designed such that they are most obvious for occupations with large numbers of jobholders or with predominantly low-skilled jobholders.

Search trees assume alphabetical sorting within each tier, because in most languages reading from a to z adapts to the way individuals tend to select an item from a long list. This is disadvantageous because respondents might be tempted to tick an item at the top. Yet, randomly sorted lists are more disadvantageous because they are counter-intuitive and therefore increase reading time. Translators are instructed to profit from alphabetical sorting by clustering related occupations to ease self-identification, e.g. *Forestry advisor*, *Forestry helper*, *Forestry manager*, *Forestry planter*, *Forestry technician*. This order is preferred over e.g. *Manager forestry*.

In case of an open response format unidentifiable occupations have to be classified in the appropriate residual occupational category, called "*not elsewhere classified*". ISCO-08 has 27 residual units. In order to include all 4 digit ISCO-08 unit groups, these residual occupations had to be included in the source list though in case of self-identification the concept of residual occupations doesn't make sense. Respondents will not read the entire choice-set and then conclude that their occupation is not present. Two strategies have been followed to solve this problem. First, for reasons of readability, all 27 residual occupations have been rephrased as "*Occupational unit X, all other*". Second, these residual occupations are always sorted at the bottom of the appropriate 3rd tier, assuming that respondents have red all occupational titles in that particular 3rd tier list before deciding to tick the residual occupation.

Self-identification assumes reading skills and therefore poses high demands on readability. The wording of an occupational title should be brief, easy to understand, and unambiguous. Reading time should be brief to reduce the risk of survey completion break-off. So, the singular has been preferred over the plural. *Bee-keeper* has been preferred over *Apiarist*. Synonyms have been avoided. Feminine occupational titles have been restricted to a minimum. The readability issues do not hold for all countries. In some countries both male and female occupational titles have to be used. For Germany, the *DTP operator* is translated into *DTP Operator/in*. In this case, the disadvantage of increasing respondents' average reading time is considered smaller than the advantage of adhering to the national understanding of gender equality. Additionally, the source list must be easy to understand and unambiguous for translators. So, *Chef cook* is preferred over *Chef,* and *Helper* is preferred over *Labourer*. Translations by national labour market experts are preferred over translations by professional translators, and nationally used occupational titles are preferred over literal translations.

The concepts of careering, job ladders and job-enlargement blur the demarcation lines across occupations, whereas clarity is critical for valid self-identification. This turned out most problematic for the assistant occupations. Is the assistant plumber part of a job ladder to become a plumber and thus one occupation, or not? This will vary worldwide and therefore the word assistant has been avoided as much as possible. Occupations including junior and senior job titles are not used. For respondents who prefer to report the hierarchy within their occupation, the WageIndicator web-survey has a follow-up survey question where these categories can be ticked.

4. The source list and occupational hierarchies

Introduction

Worldwide, an increasing share of the labour force is employed in large and medium-sized organisations. These organisations usually have a well-developed division of work, shaping hierarchical demarcation lines between occupations. For two reasons, this issue of corporate hierarchy had to be solved. First, ISCO has assigned different skill levels to different positions within the hierarchy. Second, respondents prefer to report their position within the hierarchy. Valid self-identification assumes that occupational titles are clear with respect to the corporate hierarchy. Using our knowledge about corporate structures, all occupations had to fit into a stylized, six-layer corporate hierarchy (Table 1). It is discussed hereafter.

Table 1: The stylized, six-layer corporate hierarchy applied in the WISCO Database of Occupations

Hierarchy	Description	ISCO major group
OCC+4	CEO, board members and area managers of large firms or organisations (50 or more employees)	Major group 1
OCC+3	Managers of institutions, centres, branches and alike Company director, chief executive 10-50 employees	Major group 1
OCC+2	Departmental managers, using a stylized setting of 14 departments	Major group 2-3
OCC+1	First line supervisors	In same group as occupation
OCC	Occupation	-
OCC-1	Helpers	Major group 9

CEO's, directors, managers (OCC+4 and OCC+3)

The ambiguity in the occupational title of manager has sometimes been problematic for occupational classifications (Elias and Birch, 1994). In some languages, no clear differences exist between the hierarchical manager and the person who is responsible for a product or a service within the organisation, mostly also called manager, e.g. *Account manager*. The search tree plays a major role for valid self-identification of managers. One 1st tier item is used: *Management, direction*. This leads to four 2nd tier items, notably *Department manager*; *Highest management level in organisation with more than 500 employees*; *Highest management level in organisation with 50-500 employees*; *Highest management level in organisation with less than 50 employees*. In the 3rd tier a choice of the manager occupations is presented, covering all occupations in OCC+4 and OCC+3, for example the *Livestock farm manager*. All managers of institutions, centres, branches and alike (OCC+3) are also located in the thematic search paths, for example *Agriculture, nature, animals, environment* > *Livestock* > *Livestock farm manager*. At this 3rd tier list, also the *Livestock farm worker* and the *Livestock farmer* is listed, so that respondents can make a valid choice with regard to their occupation.

Heads of departments or branches (OCC+2)

ISCO does not identify department/group/team/division managers (OCC+2), but the WISCO Database of Occupations does. In the WageIndicator web-survey these are commonly reported occupations. For this purpose, a stylized horizontal corporate structure of 14 departments has been designed, such as *Administration*; *Technical support, quality control*; *Logistics, purchasing*; and *Marketing*; and occupational titles have been added for the managers of these departments (see Tijdens 2009 for more information).

First line supervisors (OCC+1)

The US occupational classification employs occupational titles for first line supervisors. ISCO-08 does not, though this issue has been heavily debated. In the WageIndicator web-survey respondents prefer to report being a first-line supervisor, because the open format follow-up question is frequently used for this purpose. Therefore, the source list has distinct occupational titles for first-line supervisors (OCC+1), e.g. *First line supervisor personal care workers, First line supervisor protective service workers*, or *First line supervisor retail sales workers*. For the drafting of the source list, it is assumed that not all occupations have a related first-line supervisor, but that the incidence of first-line supervisors is likely in those unskilled and semi-skilled occupations, where the work is performed mostly in groups. The source list includes in total 47 first-line supervisor occupations. First line supervisors are assigned the same ISCO code as the occupation that they supervise.

Helpers (OCC-1)

According to ISCO, helper occupations are mostly classified as unskilled occupations, and they are classified distinct from the related occupation. The source list includes almost twenty distinct occupational titles for helpers, e.g. *Assembling helper, Building construction helper, Cattle station helper*, or *Road construction helper*. Helpers (OCC-1) are assigned an ISCO skill level different from the related occupations.

5. The source list: miscellaneous

No additional information needed on industry, firm size or employment status

For the recoding of open format occupation data, industry, firm size and employment status are used mostly as auxiliary variables. When using the WISCO Database of Occupations no additional survey questions are required for recoding. Consequently, some occupational titles include a reference to industry, e.g. *Dairy-products process controller* and *Chemical products process controller*. Other occupational titles refer to firm size, notably *Company director, chief executive 10-50 employees*, *Company director, chief executive 50-500 employees*, and *Company director, chief executive >500 employees*.

As for employment status, the source list does not differentiate occupations in this respect. Worldwide countries vary largely with respect to the proportion self-employed in occupations, implying that the source list would almost be double the size when including a reference to employment status in the occupational title. In case the source list is used for constructing a Socio-Economic Classification, the survey needs to include an additional question on employment status. Alternatively, percentages of self-employed per occupation can be provided for a large number of countries using data of the WageIndicator web-survey, which has a question on employment status.

Avoiding gender bias

In any occupation list, gender bias should be avoided. Analyses on Labour Force Survey data for 8 countries indicate that on a 3-digit level male-dominated occupations indeed appear to be specified on a more disaggregate level than female-dominated occupations in the 2007 initial source list (De Ruijter, De Ruijter and Jacobs 2009). The issue was solved by breaking down large female-dominated occupational units into detailed occupational titles. For example, the nurse occupation was broken down into 11 distinct occupational titles.

Coping with composite occupations

Small organisations tend to employ workers in composite jobs. Respondents may therefore want to classify themselves in more than one occupation. Web-surveys using search trees have two solutions to this problem. The first solution includes an instruction to the survey question "What kind of work do you do?" that in case of composite jobs the occupation should be ticked with the highest skill level or otherwise the one on which most time is spend. Respondent may also be offered a follow-up open response format question to specify their composite occupational title. The second solution is allowing respondents to tick more than one occupation. Unfortunately, due to technical constraints the WageIndicator web-survey does not facilitate a second choice, but may do so in the years to come.

Handicraft workers and machine-operators

In the cause of the 20th century, small-scale workshops have been replaced by factories and craft occupations by machine-operators due to industrialisation and technological innovations. Countries vary with respect to the degree that these processes have taken place. Nevertheless, even in highly industrialised countries traditional craft occupations exist, supplying handicraft goods for commercial markets. In the source list the machine operator and the handicraft workers are assigned distinct occupational titles, e.g. the *Handicraft weaver, knitter, embroiderer* and the *Weaving machine operator*, or the *Handicraft leather worker* and the *Shoemaking machine operator*. For food manufacturing, the word handicraft worker is not applicable. For bakers and butchers, in most countries the occupational titles will refer primarily to retail trade and a different phrasing is used for comparable occupations in manufacturing.

Obsolete occupations

Occupational structures are dynamic because in all countries the occupational composition of the labour force varies over time. The 0.01% threshold for including an occupation is static. This raises the issue how to detect occupations that become obsolete and how to detect whether this is a country-specific or a worldwide phenomenon? As the supply of machinery for factories, building sites, IT services and offices is increasingly a globalised business, it might not leave much room for country-variation with regard to obsolete occupations. Unfortunately, this observation is not underpinned with empirical results. Worldwide, very few research results are available with regard to occupational dynamics, most likely due to a lack of appropriate data. In the source list, occupations are not included that are judged obsolete given the technological state-of-the-art in manufacturing and services, such as *Riveter* or *Lead burner*.

New and emerging occupations

In growing businesses it is likely to observe a growing division of labour. The travel agency business used to have the occupation *Travel agent*, but as this business grew, the number of distinct occupations grew too and the source list includes five occupational titles now. A major challenge for any occupational classification is how to become aware of the new and emerging occupations (The National Center for O*NET Development 2006). In ISCO-08 a list of IT occupations was added (Hunter 2006). To trace new and emerging occupations, the WageIndicator web-survey will continue to use an open response format question for respondents, after having ticked an

occupational title from the search tree. This offers a possibility for a worldwide monitoring of new and emerging occupations, and if needed, allow for adding these to the WISCO Database of Occupations.

6. Conclusion

This paper summarizes the principles underlying the design of the WISCO Database of Occupations, whose primary aim is the measurement of occupations by means of self-identification in multi-country web-surveys. The database has three components, notably a source list of 1,594 distinct occupational titles in English, country-specific translations of these occupational titles, and a 3-tier search tree allowing respondents to navigate through the database. All occupational titles in the source list are coded according to ILO's international classification of occupations, ISCO-08. The database has been designed as part of the FP6 funded EurOccupations project for eight countries. It has been tested and used in the worldwide WageIndicator web-survey, and extended to more countries. The 2007 draft of the database was successfully used in the web-survey of 33 countries, and the 2009 final version is implemented for 48 countries.

The EurOccupations part of the database is freely available from the project website. The WageIndicator Foundation will continue to use the database for its continuous, worldwide web-survey. This web-survey also allows for detecting new and emerging occupations because of its follow-up open format survey question asking if the respondents want to add additional information about their occupation. Data-collectors and academics from all over the world are invited to use and/or to contribute to the WISCO Database of Occupations. Suggestions for improvements of occupational titles in a particular language/country are welcomed and so are reflections on the search tree. Additions for new countries/languages are particularly welcomed. All messages can be sent to the author.

The ultimate objective of the WISCO Database of Occupations is to collect worldwide survey data on detailed, comparable occupations. If this data-collection develops satisfactory, it might be used for a research agenda focussing on three objectives. First, to investigate empirically the required skill levels of occupations, aiming to understand cross-country similarities and differences. Second, for investigating empirically the occupational dynamics over time, aiming to understand the mechanisms that explain the growth and decline of occupations. Third, to investigate empirically the processes of the division of work within labour organizations, and particularly in corporate hierarchies, aiming to understand horizontal and vertical occupational demarcation lines.

7. Literature

De Ruijter, J., de Ruijter, E., Jacobs, J. (2009) *Gender bias in the EurOccupations database*, EurOccupations Deliverable D23. Internet: www.eurocupations.org

Elias, P. & Birch, M. (1994) 'Establishment of Community-Wide Occupational Statistics. *ISCO 88 (COM)', A Guide for Users* University of Warwick, Working paper Institute for Employment Research.

Hunter, D. (2006) *Occupations in Information and Communications Technology. Options for Updating the International Standard Classification of Occupations,* Geneva, ILO, Discussion Paper April 2006.

ILO (2006) Draft classification structure – ISCO-08. Including Annex 1 for updating ISCO-88 into ISCO-08, Geneva: International Labour Office.

ILO (2007) Resolution Concerning Updating the International Standard Classification of Occupations. Tripartite Meeting of Experts on Labour Statistics on Updating the International Standard Classification of Occupations (ISCO), 3 to 6 December 2007. Geneva, International Labour Office.

The National Center for O*NET Development (2006) New and Emerging (N&E) Occupations. Methodology Development Report, Prepared for U. S. Department of Labor, Employment and Training Administration, Washington, DC.

Tijdens, K.G. (2009) *Update of the underlying principles of the Database of Occupations of the EurOccupations project.* EurOccupations Deliverable D01a1. Internet: www.eurocupations.org

Tijdens, K. G. & Jacobs, J. (2009a) *The EurOccupations Database of Occupations.* EurOccupations Deliverable D01b1. Internet: www.eurocupations.org

Tijdens, K. G. & Jacobs, J. (2009b) *A test of the usefulness of the EUROCCUPATIONS database for self-administered computer-based questionnaires.* EurOccupations Deliverable D24. Internet: www.eurocupations.org