Inside occupations:

comparing the task descriptions of 160 occupations across eight EU member states

Paper prepared for ISOL

November 16, 2011

Kea Tijdens*, Esther De Ruijter**, Judith De Ruijter** [1]

* Erasmus University Rotterdam/ Department of Sociology & University of Amsterdam/Amsterdam Institute for Advanced Labour Studies (AIAS), email: k.g.tijdens@uva.nl

** AO Consult, Tilburg, Netherlands, email: e.deruijter@aoconsult.nl, j.deruijter@aoconsult.nl

Keywords: occupations; work tasks; measurement; job content; expert rating; jobholder rating

Abstract

The Task-Based Approach uses occupational titles as a key variable for including characteristics beyond an individual's general human capital, but hardly discusses the quality of data on occupational titles. This paper does so from a cross-country comparative perspective. Do similar occupational titles refer to the same work activities, as assumed often but not empirically tested? For 160 occupations, the similarity of ten task descriptions per occupation across eight countries is investigated. Using a web-survey, experts and jobholders rated the frequency and importance of each task (N = 4,197 raters). Four hypotheses are investigated: occupations are similar across countries, occupations are similar within countries, countries are similar within regard to cross-occupation interrater agreements

---

and experts and jobholders are similar in their ratings? Half of the occupations reveal no agreement across raters, one third shows a weak/moderate agreement and one in ten shows a strong agreement. Comparing four countries, within-country similarity is larger than across countries. Comparing countries, Spain demonstrates the highest agreement, Poland and Netherlands the least, and Germany lies in between. Jobholders demonstrate higher agreement than experts. Europe-wide testing of work activities in occupations seems a viable goal.

1 Introduction

In recent years, the Task-Based Approach has gained attention in complementing human capital explanations for wages and labour market mobility. This approach aims at adding occupation-specific characteristics to an individual's general human capital. For this purpose the occupational variable is usually clustered into categories, for example manual versus white-collar jobs, high- versus low-skilled, agricultural versus manufacturing jobs, computer use versus non-use, and alike (see Brynin, 2002, for an overview). A further categorization of occupations is undertaken by Autor and Handel (2009), developing multi-item, additive scales for assessing the relative merits of job-level measures. In explaining MNE's offshoring practices, Becker et al (2009) distinguish non-routine and more interactive tasks. For long the skill dimensions of occupations have been studied. Workers' educational attainments are a poor proxy of their jobs' skill levels, challenging for other measures of the requirements of the job occupied. Although some occupational classifications include references to the required skill levels, as will be discussed hereafter, there is a need for including various dimensions of skill (e.g. Ingram and Neumann, 2006; McGuinness and Sloane, 2011). In summary, occupational titles are a key variable for including characteristics beyond an individual's general human capital. This challenges the quality of data on occupational titles. This paper does so from a cross-country comparative perspective.

Does an Italian plumber engage in the same activities as a plumber from France, Poland or the United Kingdom? This paper aims to investigate the similarity of occupations across EU member states. This is a relevant topic for both academic and practical reasons. The academic relevance refers to the social and behavioural sciences. Here, occupation is a key variable in a wide strand of studies, for example, those involving job analyses, school-to-work transitions, gender segregation, skill mismatches, competency modelling, professionalization and the like. In these fields, many cross-country comparative studies are based on the implicit assumption that similar occupational titles refer to the same work activities, but an empirical basis for this assumption is lacking.

The practical relevance refers to European policies. Over the past decade the European Union has issued a number of policies with a focus on occupations, for example, the New Skills New Jobs initiative, the policy to harmonise European Vocational Requirements, and the EURES portal for

matching vacancies and job seekers across borders, amongst others (European Commission, 2008a, 2008b). In the years to come, the European Skills, Competences and Occupations taxonomy, ESCO, will be developed in an effort to create a multilingual framework of occupations, skills, competences and qualifications (ESCO, 2010). Yet so far, no plan has been presented to investigate whether occupational titles have the same meaning across countries beyond semantic similarities or are being categorised similarly in an occupational classification. This highlights the need for methodological exploration and the development of an empirical underpinning for assessment of the work activities of occupations in Europe.

Jobs consist of similar sets of tasks performed within the context of an organisation. Beyond organisations, the wording changes from jobs to occupations. Occupations consist of a cluster of similar job titles, thus of similar main tasks. According to Tomaskovic-Devey (1995), the concept of occupation is especially relevant in comparative research, since studying jobs limits generalisations to the work organisation context and hampers national or international comparison. The worldwide International Standard Classification of Occupations (ISCO) is the most widely used for occupational comparisons. The recent ISCO-08 distinguishes 9 major groups at the highest level of aggregation, stepwise breaking these groups down into 433 occupational units at the classification's lowest 4-digit level of aggregation. For the purpose of self-identification in a closed survey question 'What is your occupation', these 4-digit units are not sufficiently detailed. Survey respondents may not consistently fit their job titles into the aggregated 4-digit units, which may result in an aggregation bias. As part of the EU-FP7 funded EUROCCUPATIONS project, the 4-digit ISCO-08 occupational units have been further detailed into a 5-digit level of aggregation, resulting in a list of 1,440 occupational titles. This list has been used to select the 160 occupations, used in the comparison of the task descriptions of across 8 EU member states in this paper.

Jobs and occupations are both multi-dimensional concepts with hundreds of indicators, ranging from attributes such as decision latitude, stress levels and skill requirements to compensation, responsibilities, accountability and the like. For many decades these attributes have been developed and tested in job analyses, job evaluation and personnel psychology studies (Dierdorff and Morgeson, 2009). In a meta-analysis of job analysis interrater reliability studies, Voskuijl and Van Sliedrecht (2002) found that of all attributes the task descriptions were rated least reliable (.29). In a recent study, Dierdorff and Morgesson (2009) confirm that interrater reliability decreases when ratings are made on job-related tasks compared to generic tasks. Keese (2009) finds that in surveys of enterprises' skill requirements, employers often tend to report the tasks that need to be performed in a job rather than focusing on skills or qualifications. Although difficult, occupational task descriptions are an essential starting point for ascertaining the required skills and competences of occupations. In the United States, the O*NET® Database of Occupations includes hundreds of descriptors, including the empirically tested descriptions of tasks performed in the occupations (Tippins and Hilton, 2010).

In the past decade, skill requirements have been investigated for a range of occupations across the European Union, for example, the bricklayers study by Brockmann et al. (2010). Some studies have also described the tasks performed in occupations within countries, but hardly any study has undertaken a cross-country empirical investigation of that nature. The few which did so focused on partial labour markets and were based on anecdotal rather than systematic research, for example, the study of unskilled occupations by Valenduc et al. (2008). This paper therefore focuses on the tasks performed in occupations, breaking new ground by investigating for the first time the similarity of the tasks performed in particular occupations within and across European countries. It aims to take a step toward a systematic empirical assessment of the tasks performed in occupations, covering the European labour market.

This paper has two objectives. First, it aims to analyse the similarity of occupations within and across countries using interrater agreements statistics. For this purpose, a methodology for drafting and subsequently measuring the main tasks performed in occupations in eight European countries has been developed and applied in a web-survey. Second, it aims to compare experts and jobholders with regard to their ratings. Section 2 details the empirical investigations of the work activities in occupations, including measurement, sampling and ratings issues. Section 3 outlines the methodology used in this study, including the selection of occupations, the drafting of task statements and the testing of similarity. Section 4 presents the findings of the study. Section 5 draws the main conclusions.

2 Occupations and the related work activities

2.1 Occupational information systems

Quite a number of industrialised countries maintain national occupational information systems or databases. Of these, the O*NET® Center has most frequently published on methodological issues. In Europe, national occupational databases are maintained in Austria (AMS-Qualifikationsbarometer), in Germany (Kompetenzenkatalog BerufeNet), in France (ROME), in Sweden (Taxonomy database), in Poland, in Belgium/Flanders (VDAB), in Finland (Ministry of Employment and the Economy) and a number of other countries. In Asia, Korea has its KEIS employment center. Most of these databases are associated with public employment services and student or career counseling. Typically, these databases include occupational titles, short and long task descriptions, skill requirements, and attributes such as wages or entry restrictions for disabled persons. The skill requirements are usually generic categories, but in ROME, for example, occupational titles are matched with registered educational categories. Apart from the O*NET® publications, not much information in English is

publicly available regarding the methodologies used to initiate and maintain occupational information systems.

Worldwide, labour force surveys (LFS) mostly ask the respondent's occupation using an open response format survey question, and then undertake substantial recoding efforts to classify the reported job titles into the national occupational classifications (Tijdens, 2010). LFS data allows researchers to rank occupations for variables such as skill levels, gender composition, supervisory duties, and the like (for example Dumont 2008). Labour force surveys are however not suitable for surveying work activities and skill requirements for occupational information systems, because this would require either a set of follow-up questions per occupation or a large set of survey questions for generic tasks and requirements. Empirical testing of job descriptions and skill requirements is typically performed only for selected workplaces or occupational groups (Tippins and Hilton, 2010). Sampling limitations also point to the practice to select workplaces or occupational groups. Sampling jobholders from a sampling frame of occupational registers is restricted to a limited set of occupations, for example, nurses or lawyers. Sampling the entire labour market runs the risk of insufficient observations for small occupations, for example, hatmakers.

The O*NET® Center uses two methods for investigating occupations (Rivkin et al., 2001; Tippins and Hilton, 2010). The establishment method is based on a probability sample of establishments and a sample of employees in these establishments. The occupational expert method is used for occupations that are more difficult to locate within establishments. Experts can be job analysts, individuals with knowledge of the occupational group, or members of occupational associations. Sometimes, jobholders are also referred to as experts. Surveying experts aims to gather the required information in an efficient manner, though it may not always be clear how these experts have acquired their occupational knowledge (O*NET®, 2002). The occupational expert method is used when the establishment method is not feasible, thus when the number of jobholders is small, when occupations are new or emerging, or when jobholders are in remote locations or scattered over many workplaces.

In conclusion, there is not one sampling strategy for collecting occupational information. This is due to the absence of a sampling frame of all jobholders in all occupations and the need to collect sufficient observations for occupations with few jobholders, requiring extremely large sample sizes when sampling the labour force. The most common method targets only selected occupations, surveying experts, supervisors, or jobholders.

2.2   What do workers do?

An organisation's business process model determines which work activities its workforce has to perform and which not. Work activities consist of tasks, clustered into jobs. The task set determines what workers do in their jobs. This so-called work-process oriented approach identifies work

activities that are central to a particular job role and then identifies the personal attributes required to achieve appropriate outcomes (Winterton et al, 2005). Tasks are not assigned randomly to jobs. Though extensively addressed in socio-technical job design, the mechanisms underlying the grouping of tasks into jobs have hardly been empirically investigated. Most likely, a task set is limited to one skill specialisation and it does not vary greatly with regard to the skill levels required for each task. Most likely, smaller organisations cluster broader sets of tasks into one job title, part-time jobs are assigned smaller sets of tasks, and hierarchical structured organisations design a job's task set such that a chain of related job titles will constitute a career path. In their study of the purchasing profession, Mulder, Wesselink, and Bruijstens (2005) describe that when starting a job profile development project, it is essential to acknowledge the stratification of related job titles.

Many organisations use job analysts to describe the jobs' task sets. For several reasons, such as new technologies, specific workplace practices regarding the division of work, the need for workers to temporarily replace their colleagues, or a worker's own initiative, jobholders' daily tasks may differ in whole or in part from the description of the task set. In a study of 1,432 secretaries in the Netherlands, 80% reported knowing their job description, and of these 43% reported performing the tasks according to the description, 49% did more tasks, 4% did other tasks, and 3% did fewer tasks (Tijdens, 2005).

Job titles can be aggregated to occupational titles to facilitate communication beyond the organisation. The processes underlying the grouping of tasks into jobs will be similar to the grouping of job titles into occupations, whereby occupational units at higher aggregation levels will consist of an increasingly heterogeneous bundle of job titles and an even more heterogeneous bundle of tasks. Hence, to test the similarity of occupations with regard to the task sets, occupational titles at disaggregated levels should be used.

In recent decades, a shift from traditional task analyses to competency modelling can be observed. It is argued that competencies are critical to organisational survival, productivity and continual improvement, and therefore, it is important to analyse and describe these competencies (for example Rodriguez et al., 2002). Note that competencies are closely related to skill levels and comprise a combination of skills, knowledge and attitudes that are necessary for particular occupations (Le Deist and Winterton, 2005; Winterton et al., 2005). A concern regarding the measurement of competencies is their validity as measurable constructs (for example Lievens et al., 2004; Markus et al., 2005). Most competency modelling approaches are directed to identifying relevant competencies without analysing the occupation's task set. Information on tasks decreases the complexity of competency judgements by experts and bolsters reliability of competency measurement (for example Dierdorff and Wilson, 2003; Lievens et al., 2004; Voskuijl and Van Sliedregt, 2002). Thus, for occupational information systems aiming to include competencies, it is important to do so by departing from information on the occupation's task set.

2.3 Measuring tasks in occupations

O*NET® has two approaches for measuring work activities in occupations. A generic approach aims to measure the so-called Generalized Work Activities (GWA) and a job-specific approach aims to measure work activities in occupations for which the GWA are not sufficiently detailed (Tippins and Hilton, 2010). GWAs are measured independent of the specific occupational title and therefore GWA tests include a large list of items. GWA tests take respondents 60-90 minutes, leading to substantial non-response rates in the O*NET® surveys. According to Tippins and Hilton (2010), GWAs are unavoidably vague, implying that they should not be considered tasks, but rather the underlying behavioural dimensions of work activities. In addition, the authors state that GWAs are culturally biased, in that they intrinsically favour certain job types, which may result in a biased ranking of jobs, for example, a gender bias or a bias towards managerial jobs.

In the job-specific approach, highly specific tasks represent the most detailed descriptors commonly used in work analysis (Dierdorff and Wilson, 2003). They are not applicable across different jobs and occupations, for example, accountants and paramedics will have different sets of tasks associated with their occupations. In this approach, the work activities in predefined occupations are investigated beyond the organization-specific job evaluation methods, where job descriptions may consist of several hundred tasks. In the O*NET® methodology, the list is reduced by including only tasks that a vast majority of jobholders (80% was the goal) could be expected to perform (Dierdorff and Morgeson, 2009). In recent years, a rapidly growing number of task descriptions for occupations are being posted on the Internet, either as a brief description or as a list of tasks. The websites stem from occupational information systems, industry-level organisations, professional organisations, the ILO, statistical agencies, public employment services, and the like. The availability of occupational descriptions has facilitated, for example, the drafting of task sets for the 433 occupational titles at 4-digits in ISCO-08 (Hunter, 2009). Yet, few empirical tests have investigated to what extent the jobholders in these occupations undertake the described task sets in different countries. In other words, are aggregated occupational titles reflecting homogeneous task sets within and across countries?

Many occupational classifications refer to the skill levels of occupations, which are also not empirically tested across countries. Testing might resolve concerns such as those of Dumont (2006), who concluded from wage regressions for four EU countries that the ISCO-88 skill levels were not very reliable and varied across countries. An empirical underpinning of the required skill levels for all 4-digit ISCO-08 occupations may require European-wide analyses, similar to the one by Mouw and Kalleberg (2010) on the role of occupations in explaining the increase in wage inequality in the USA. A recent strand of economic literature uses a so-called task-based view, aiming to classify occupations according to their core tasks and then considering the set of formal and informal skills required to

carry out these tasks. The authors predominantly use attributes such as IT-use and the incidence of routine tasks to analyse wage growth and outsourcing of jobs (Autor et al., 2003; Dumont, 2008).

In summary, the generic versus job-specific approaches touch upon major methodological issues. The generic work descriptions facilitate non-occupation-specific sampling and can therefore be asked in any survey. Yet, the list of descriptions is so long that surveying is time consuming for respondents and the descriptors may be so generic that they do not meet the needs of the major user groups of the occupational information system. In contrast, job-specific descriptions result in much shorter surveys but require occupation-specific sampling or occupation-based screening questions with follow-up questions, which is only possible in computer-assisted face-to-face or telephone surveys, or in web-surveys.

2.4   Rating occupations

In GWA and job-specific surveys, respondents are asked how frequently tasks are performed in a particular job and how important these tasks are for performing the job, mostly to be rated on five-point Likert scales. These scales are used to compute interrater agreement across and within occupations. Variance in ratings can stem from unsystematic sources (random error) which is attributable to the person making the rating, and from systematic sources which reflect legitimate differences in the task sets of jobholders (Tsacoumis and Van Iddekinge, 2006). The authors argue that high levels of interrater agreement among job analysts may not necessarily reflect high validity, because it could result from excessive simplification of stimulus materials.

In past decades, discussions focused on the best choice of raters, whether job analysts, jobholders, supervisors, HR officers, or members of professional associations. Jobholders are the source for most O*NET® data, providing information in the domains of job knowledge, GWAs, work context, education and training, work styles, and tasks (Tippins and Hilton, 2010). Rating by jobholders is associated with some major problems, which are extensively discussed by Morgeson and Campion (2000). These problems are also acknowledged in job analysis research (Lievens et al., 2004). Jobholders tend to assess the required skill level of their occupation to be higher than it actually is (for example Van der Klink and Boon, 2003). For example they might indicate supervision of others, while in fact they only instruct new colleagues. Jobholders are also strongly influenced by the corporate context in which they are employed, whereas occupational descriptions assume that they overcome these corporate boundaries. There is some evidence indicating that compared to job analysts, jobholders have more difficulty in making the judgments needed in work analysis (Dierdorff and Wilson, 2003). Other research indicates that jobholders can provide psychometrically sound ratings of more concrete characteristics of occupations, such as tasks, but that their ratings of more abstract concepts, such as abilities or job-related personality traits, are less reliable (Dierdorff and

Morgeson, 2009). Yet in a study on new and emerging occupations, O*NET® (2009) concludes that jobholders are most suited for collecting data on these types of occupations. In conclusion, it is not yet decided whether experts and jobholders provide reliable response, when surveying tasks.

3 Hypotheses, method and data collection

3.1 Research objective

This paper aims to investigate the similarity of 160 occupations across eight European countries: the United Kingdom, Germany, France, Spain, Poland, Belgium, the Netherlands and Italy. For this purpose, for each occupation a detailed task set has been drafted, as will be explained in the next section. Experts and jobholders, further called raters, have been asked to rate the task sets with regard to the frequency and importance of each task in the set. The following hypotheses have been investigated:

1) Occupations are similar across the eight countries, thus the interrater agreement within occupations is high.

2) Occupations are similar within each country, thus the interrater agreement within occupations is high within each country.

3) Countries are similar within regard to across-occupation interrater agreements.

4) Occupations are rated similarly by experts and jobholders, thus the within-occupation interrater agreement for each occupation does not differ across experts and jobholders.

For this study, the EUROCCUPATIONS project conducted a multilingual web-survey. This and the subsequent steps in the research are described in the next sections.

3.2 Selecting occupations and drafting tasks statements

The first part of the EUROCCUPATIONS project addressed the validity of self-identification of occupations when survey respondents are using web-surveys. Most survey respondents tend to report their job title in great detail, as they know it from their employment contract, a job evaluation scheme, or common usage in the workplace. For an open-ended survey question about occupation, web-surveys are disadvantageous because of the absence of interviewers who can control for unidentifiable or aggregated responses (Tijdens, 2011). Web-surveys can also apply a search tree for respondents' self-identification of their occupation. For this purpose, the project defined occupations with greater precision than the 433 ISCO-08 occupational units at 4-digits by adding further digits, aiming to cover all major occupations in the labour forces of the eight countries. The project's initial list totaled 1,440

occupations (Jacobs, 2007). The project employed the following definition to identify the level of aggregation of occupations:

"An occupation is a bundle of job titles, clustered in such a way that survey respondents in a valid way will recognize it as 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" (Tijdens, 2010, p.16).

The second part of the EUROCCUPATIONS project aimed at investigating the similarity of occupations. For this purpose, 160 occupations were selected from the list of 1,440 occupations. Four criteria were applied for this selection, namely (1) variation in skill level and in ISCO major groups, (2) variation in gender composition (male-dominated, mixed, and female-dominated occupations), (3) preference for occupations with large numbers of jobholders, and (4) preference for 'blurred' occupations with wide demarcation lines, for example, managers, process operators, waiters (De Ruijter et al, 2007).

Departing from the O*NET® (2002) guidelines for the description of tasks, EUROCCUPATIONS developed standardised procedures and formats for the drafting of approximately 10 unique task descriptions for each of the 160 occupations (De Ruijter and De Ruijter, 2007). Whereas O*NET® uses predominantly generic task descriptions, EUROCCUPATIONS aimed for occupation-specific task descriptions, because the discourse in Europe focuses on the comparability of occupational titles across countries and on the matching of job seekers with specific skills to specific occupations. Each national EUROCCUPATIONS team drafted a set of task descriptions for some 20-30 occupations, resulting in task descriptions for 160 occupations. This research was based on desk and Internet research using information from at least six different sources such as O*NET®, ROME, BIBB, VDAB, (inter)national competency profiles, and the like. Based on a comparison of the task descriptions from the different sources, the teams drafted on average 10 tasks (from 8 to 12) per occupation, following the standardised procedures regarding the level of aggregation and the use of language. All task descriptions were drafted in English. They were checked for conformity to the procedure, Anglo-Saxon bias (specifically inclusion of non-English sources), and for international validity and consistency. All team members commented on all task descriptions and revisions followed, if needed. The final task descriptions were then translated into the languages of the participating countries. These task descriptions, the dataset, the codebook, and the project deliverables are freely available (Tijdens et al., 2011, see www.EUROCCUPATIONS.org).

3.3 Detailing the data collection

For the testing of the 160 occupations, the project invited occupational experts to rate the task descriptions with regard to frequency and importance of tasks and to indicate the skill requirements

using an expert web-survey. Experts were defined as individuals who had knowledge about occupations, for example, people from professional organisations, interest groups, vocational training institutions, representatives of employers' or employees' organisations, policy officers in industry-level organisations, occupational advisors, HR professionals or supervisors in companies. The project teams undertook great efforts to recruit experts through their own networks and through expert organisations. In total, 2,468 experts from eight countries were recruited.

A web-survey was obviously the most suitable method to collect data on 10 tasks for 160 occupations in eight countries. Computer-assisted face-to-face interviews were too expensive, telephone interviews were not suitable because respondents could not instantly comprehend the task descriptions, and postal and other paper-based survey modes could not cope with 8 * 160 survey versions. In 2007 the web-survey was developed, tested and translated. It was online from late 2007 until early 2009. In total, the 2,468 experts completed 2,950 questionnaires, whereby 14 experts completed 10 or more questionnaires, 126 experts completed between 2 and 9 questionnaires, and the remaining experts completed 1 questionnaire (Table 1). This was more than the target of 2,560 completed expert questionnaires (160 occupations * 2 raters * 8 countries), but for a number of occupations the target of at least ten raters per occupation was not reached.

In order to gather sufficient occupational data, and also to facilitate comparison of jobholder and expert ratings, jobholders were recruited to complete the web-survey. The survey questions concerning the tasks were identical for both jobholders and experts. For the recruitment of jobholders, a weekly changing advertisement seeking workers in particular occupations was posted on the frequently visited WageIndicator websites on work and wages in each of the eight countries (see www.WageIndicator.org), inviting web-visitors working in these occupations to complete the questionnaire and offering a prize incentive. Between February 2008 and January 2009, a total of 1,247 jobholders completed a questionnaire (Table 1).

- -

INSERT TABLE 1 ABOUT HERE

- -

For 151 of the 160 occupations, the dataset holds at least two observations per occupation for at least six of the eight countries. For these 151 occupations, in total 4,129 questionnaires have been completed and the number of raters ranges from 12 for 'Telecommunication equipment installer or repairer' to 136 for 'Electrical engineer'. These 4,129 questionnaires have been used in the analyses in

section 4. For the comparison of countries, the analyses could only be performed for Spain, Germany, Netherlands and Poland, due to an insufficient number of raters in the other countries.

3.4   Testing the similarity of occupations

In the questionnaire, the frequency of each task description was measured using a 5-point Likert scale ranging from never to daily, whereas the importance of each task was measured using a 3-point scale ranging from not at all important to very important. To investigate the amount of similarity between occupations across and within countries and across raters, a proximity or distance measure was used. The interrater agreement is a proxy for similarity and measures "the extent to which the different judges tend to assign exactly the same rating to each object" (Tinsley and Weiss, 2000, p. 98). Here, for the interrater agreement within occupations the single-item rWG has been used (James et al., 1984, 1993). When multiple judges rate a single target on a single variable using an interval scale of measurement, the interrater agreement may be assessed using the rWG index, which defines agreement in terms of the proportional reduction in error variance, $rWG = 1 - S^2_X s^2_E$. Here, E is the variance expected when there is a complete lack of agreement among the judges (LeBreton and Senter, 2008). This is the variance obtained from a theoretical null distribution representing a complete lack of agreement among judges. A new variable was computed, indicating the variance observed within each occupation across the set of raters. Next, we compared the observed variances to the variance expected if judges had responded randomly. A rWG =.70 suggests that 30% of the variance in the ratings is still error variance.

In a first step, for each task the variance across raters was calculated for the frequency and the importance of tasks. In a second step, for each occupation the variance in the frequency across the task descriptions per occupation was computed as in equation (1). In equation (1) we divided by 2 because the expected variance for equal distribution is equal to 2, given a scale ranging from 0 to 4.

(1) rWG_freq = 1 - (obs_var/2)

For the task importance the rWG was computed as in equation (2). In equation (2) we divided by 1 because the expected variance for equal distribution is equal to 1, given a scale ranging from 0 to 2.

(2) rWG _impor = 1 - (obs_var/1)

Then, the mean rWG per rated occupation was calculated by adding the rWG's for task frequency and task importance and dividing them by the number of tasks in the occupation.

4 Results

4.1 Occupations are similar across countries

Our first hypothesis assumes a high interrater agreement within occupations, irrespective of the country under study. The first column in Table 2 shows the inter-rater agreements (rWG) summarized in six categories, revealing that the similarity of task frequency within the 151 occupations varies greatly. It ranges from -1.52 for 'Carer for the disabled' to .92 for 'Wood processing plant operator'. For half of the 151 occupations it reveals a lack of agreement or even no agreement at all concerning the frequency of tasks. For more than one third of the 151 occupations a weak or moderate agreement can be noticed. Only one in ten occupations shows strong agreement. The overall conclusion is that the frequency of tasks in occupations is rarely similar across raters.

Regarding the task importance, the findings reveal greater consistency across the raters. Half of the 151 occupations reveal a strong or very strong agreement. One third shows a weak or moderate agreement, whereas the remaining occupations reveal a complete lack of agreement. Detailed analyses (not in the Table) point to a weak relationship: the lower the occupation's skill level, the higher the interrater agreement scores.

- - - - - - - - - - - - - - - - - - - -

INSERT TABLE 2 ABOUT HERE

- - - - - - - - - - - - - - - - - - - -

4.2 Occupations are similar within countries

The limited degree of similarity found in the previous section might be due to differences across countries, raising the question of whether similarity improves when investigated per country. Our second hypothesis therefore assumes that within countries the task frequencies and the task importance in occupations are similar across raters. This second hypothesis could only be studied for four countries, as the number of raters was too low in the other four countries. Table 3 shows that in Spain more than 80% of the rated occupations fall in the category 'moderate agreement' or higher. For Germany, the comparable percentage is 58%, for the Netherlands 43% and for Poland 48%. This supports the assumption that within countries, the similarity of ratings of the frequency of tasks in occupations is greater than across countries.

- - - - - - - - - - - - - - - - - - - -

INSERT TABLE 3 ABOUT HERE

- - - - - - - - - - - - - - - - - - - -

4.3   Countries are similar across occupations

Do countries differ with regard to their occupational practices? Do some countries have stricter divisions of labour practices than other countries? Our third research objective addresses country differences. The data allow for computing the country-specific extent of agreement rates across all occupations. Note that although a minimum of 10 raters per occupation is mostly required, in our cross-country cross-occupation analyses this requirement could not always be met. Occupations that had been rated by no or very few raters have been excluded from the analyses, as the number of occupations in the second column of Table 4 shows.

Table 4 reveals the results of the analyses. It shows that the mean rWG for Spain is highest of all four countries, indicating a strong agreement across occupations, whereas a lack of agreement can be noticed for both Poland and Netherlands. Germany is in between with a weak agreement. An identical exercise has been conducted for the importance of the tasks. The second panel of Table 4 reveals a moderate agreement across all raters and all occupations and a strong agreement for Poland, Spain and United Kingdom.

These findings are not surprising. The Spanish labour market is characterized by a large share of small organizations and small organizations are assumed to have the most blurred and wide occupational demarcation lines. Hence, the division of work is low, implying that many of the tasks in a given occupation are performed frequently. The high rWG indicates that understanding of this phenomenon is shared across the country.

- - - - - - - - - - - - - - - - - - - -

INSERT TABLE 4 ABOUT HERE

- - - - - - - - - - - - - - - - - - - -

4.4   Experts and jobholders rate similarly

Our fourth hypothesis assumes that experts and jobholders do not differ with regard to their ratings. For this purpose, the interrater agreements on task frequency per occupation have been compared between the experts and jobholders, assuming no country differences. Table 5 reveals the results of this analysis, whereby occupations with less than 10 raters have been excluded from the analyses. It

should be noted that the experts have rated 146 occupations with at least 10 raters per occupation, whereas this is 40 occupations for the jobholders. As explained in section 3, this is due to the fact that jobholders were recruited specifically for occupations with too few expert raters. The table shows that for the experts one third of the occupations reveal at least a moderate agreement, whereas this is almost half of the occupations for the jobholders.

- -

INSERT TABLE 5 ABOUT HERE

- -

To investigate the differences across experts and jobholders, Table 6 reveals that jobholders on average reveal a higher interrater agreement compared to experts for both the task frequencies and the task importance. This finding should be interpreted with caution, because differences between experts and jobholders might be partly due to the differences in the number of raters in the two groups.

- -

INSERT TABLE 6 ABOUT HERE

- -

5 Conclusion

This paper started with the question of whether an Italian plumber performs the same activities as a plumber from Germany, Poland or France, which is a relevant question for both scientific and practical reasons. This paper aimed to investigate the similarity of 160 occupations across eight EU member states. The study builds on the work of the O*NET® Center in the USA and its occupational information system. O*NET® distinguishes two approaches for analyzing work activities, namely the generic versus job-specific descriptions of work activities. The generic work activities refer to underlying behavioral constructs in all occupations, thus requiring ratings for a wide range of items from all respondents. The job-specific descriptions of work activities depart from unique tasks per occupation, and thus requiring ratings from respondents with knowledge about the occupation. The generic approach allows for labour force-wide sampling, whereas the job-specific approach assumes occupation-specific sampling or occupation-based screening questions with follow-up questions,

which is only possible in computer-assisted face-to-face or telephone surveys, or in web-surveys. This paper adapts the job-specific approach.

Four hypotheses have been investigated, questioning whether occupations are similar across the eight countries, if occupations are similar within a country, if countries are similar within regard to cross-occupation interrater agreements and if occupations are rated similarly by experts and jobholders. For this study, the EUROCCUPATIONS project conducted a multilingual web-survey. A methodology has been designed for drafting and subsequently measuring the main tasks conducted in 160 occupations in eight European countries, using a web-survey. Occupational experts and jobholders have been invited to complete the survey (N = 4,197 raters for 151 occupations with sufficient respondents). The data has been analysed with regard to the similarity of occupations within and across countries, using interrater agreements statistics.

The findings show that across eight EU countries half of the occupations reveal no agreement, one third reveal a weak or moderate agreement and one in ten reveals a strong agreement. Interestingly, there is more agreement among raters concerning the importance of tasks than the frequency of tasks. Comparing four countries, in Spain more than 80% of the occupations reveal moderate agreement (Germany 58%, the Netherlands 43%, Poland 48%), supporting the assumption that within countries the similarity of occupations is larger than across countries. Disregarding the occupations, Spain reveals the highest agreement, Poland and Netherlands lack agreement, and Germany is in between. Comparing experts and jobholders, the latter reveal higher agreement.

Our study reveals that empirical testing of task descriptions could be undertaken for a wide range of occupations across Europe. It poses the idea of building a Europe-wide tasks library for measuring work activities in occupations. Based on the EUROCCUPATIONS pilot study, this seems a viable goal. Finally, the study supports the idea of a large scale, Europe-wide investigation of the skill requirements and certificates needed to perform occupations, using the same methodology.

6   References

Autor, D., Levy, F. and Murnane, R. (2003), 'The skill content of recent technological change: An empirical investigation', *Quarterly Journal of Economics*, 118, 1279-1333.

Autor, D.H. and Handel, M.J. (2009), Putting Tasks to the Test: Human Capital, Job Tasks and Wages, NBER Working Paper No. 15116, National Bureau of Economic Research, Cambridge, MA.

Becker, S.O., Ekholm, K., Muendler, M.A. (2009*), Offshoring and the onshore composition of tasks and skills*. Stirling Economics Discussion Paper 2009-18.

Brockmann, M., Clarke, L. and Winch, C. (eds) (2010), *Bricklaying is more than Flemish bond. Bricklaying qualifications in Europe* (Brussels: European Institute for Construction Labour Research).

Brynin, M. (2002), 'Overqualification in Employment', *Work, Employment and Society*, 16(4), 637-654.

De Ruijter J and De Ruijter E (2007) Procedures task descriptions key occupations. Rotterdam: Erasmus University Rotterdam, EurOccupations deliverable D05.

De Ruijter J, De Ruijter E and Jacobs J (2007) Key occupations and cluster division. Rotterdam: Erasmus University Rotterdam, EurOccupations deliverable D01c.

Delamare Le Deist, F. and Winterton, J. (2005), 'What is competence?', *Human Resource Development International*, 8, 27-46.

Dierdorff, E. and Morgeson, F. (2009), 'Effects of Descriptor Specificity And Observability on Incumbent Work Analysis Ratings', *Personnel Psychology*, 62, 601–628.

Dierdorff, E. and Wilson, W. (2003), 'A meta-analysis of job analysis reliability', *Journal of Applied Psychology*, 88, 635-646.

Dumont, M. (2006), 'The reliability–or lack thereof–of data on skills', *Economics Letters*, 93, 348–353.

Dumont, M. (2008), *Wages and employment by level of education and occupation in Belgium* (Brussels: Federal Planning Bureau, working paper 22-08).

ESCO (2010), *Overview of ESCO - a European classification of Skills, Competences and Occupations* (Brussels: European Commission, DG Employment / DG Education and Culture).

European Commission (2008a), *New skills for new jobs. Anticipating and matching labour market needs* (Brussels: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – COM (2008) 868 final).

European Commission (2008b), *Recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning* (Brussels, Commission of the European Communities).

Hunter, D (2009), *ISCO-08 Draft definitions* (Geneva: ILO).

Ingram, B.F. and Neumann, G.R. (2006) 'The Returns to Skill', Labour Economics, February 13(1), 35-59.

Jacobs J (2007) Structure and translations extended list. Rotterdam: Erasmus University Rotterdam, EurOccupations deliverable D01b.

James, R., Demaree R. and Wolf G. (1984), 'Estimating within-group interrater reliability with and without response bias', *Journal of Applied Psychology*, 69, 85–98.

James, R., Demaree R. and Wolf G. (1993), 'rwg: An assessment of within-group interrater agreement', *Journal of Applied Psychology*, 78, 306–309.

Keese, M. (2009), *Overview of OECD Work on Skills*. Paper presented at the OECD/Cedefop International Seminar on Measuring Generic Work Skills Using the Job Requirements Approach: Thessaloniki.

LeBreton, J. and Senter, J. (2008), 'Answers to 20 Questions About Interrater Reliability and Interrater Agreement', *Organizational Research Methods*, 11, 815-852

Lievens, F., Sanchez, J. and De Corte, W. (2004), 'Easing the inferential lap in competency modeling: The effects of task-related information and subject matter expertise', *Personnel Psychology*, 57, 881-904.

Markus, L., Cooper-Thomas, H., and Allpress, K. (2005), 'Confounded by competencies? An evaluation of the evolution and use of competency models', *New Zealand Journal of Psychology*, 34, 117-126.

McGuinness, S., P.J. Sloane (2011) Labour Market Mismatch Among UK Graduates: An Analysis Using REFLEX Data, *Economics of Education Review,* 30(1): 130-145.

Morgeson, F. and Campion, M. (2000), 'Accuracy in job analysis: Toward an inference-based model', *Journal of Organizational Behavior*, 21, 819-827.

Mouw, T. and Kalleberg, A. (2010), 'Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s', *American Sociological Review*, 75, 402-431.

Mulder, M., Wesselink, R. and Bruijstens, H.C.J. (2005), 'Job profile research for the purchasing profession', *International Journal of Training and Development*, 9, 185-204.

O*NET® (2002), *Appendix D - The Development of the Occupational Information (O*NET®) Analyst Database* (Raleigh, O*NET® Center).

O*NET® (2009), *New and Emerging Occupations of the 21st Century: Updating the O*NET®®-SOC Taxonomy. Summary and Implementation* (Washington DC: paper prepared for U. S. Department of Labor, Employment and Training Administration, Office of Workforce Investment, Division of Workforce System Support).

Rivkin, D., Lewis, P., Cox, S. and Koritko, L. (2001), *Pilot Test Results: Testing Subject Matter Expert Methodology for Collecting Occupational Information for O*NET®* (Washington DC: Paper prepared for O*NET®).

Rodriguez, D., Patel, R., Bright, A., Gregory, D. and Gowing, M. (2002), 'Developing competency models to promote integrated human resource practices', *Human Resource Management*, 41, 309-324.

Tijdens K (2005) Onderzoeksrapport Een functie met inhoud 2004. Een enquete naar de taakinhoud van secretaressen 2004, 2000, 1994. Amsterdam: University of Amsterdam, AIAS Working Paper 05-33.

Tijdens K (2011) Dropout rates during completion of an occupation search tree in web-surveys. Submitted for publication.

Tijdens K, De Ruijter E and De Ruijter J (2011) Codebook and explanatory note on the XYZ dataset about the job content of 150 occupations. Amsterdam: University of Amsterdam, AIAS Working Paper 107.

Tijdens, K.G. (2010) *Measuring occupations in web-surveys, the WISCO database of occupations.* Amsterdam: University of Amsterdam, AIAS Working Paper 86.

Tinsley, H. and Weiss, D. (2000), 'Interrater reliability and agreement', in H. E. A. Tinsley and S. D. Brown (eds), *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, (San Diego, CA: Academic Press), pp. 95-124.

Tippins N. and Hilton M. (eds) (2010), *A Database for a Changing Economy: Review of the Occupational Information Network (O*NET®) Panel to Review the Occupational Information Network (O*NET®)*, (Washington DC: The National Academies Press).

Tomaskovic-Devey D. (1995), 'Sex composition and gendered earnings inequality. A comparison of job and occupational models' in J.A. Jacobs (ed), *Sex inequality at work*, (Thousand Oaks: Sage Publications), pp. 23-56.

Tsacoumis S. and Van Iddekinge C. (2006), *A comparison of incumbent and analyst ratings of O*NET® skills.* (Alexandria, VA: HUMRO paper FR05-66. Available: http://www.onetcenter.org/reports/SkillsComp.html).

Valenduc, G., Vendramin, P., Krings, B.-J., Nierling, L. (2008), *How restructuring is changing occupations. Case study evidence from knowledge-intensive, manufacturing and service occupations.* Leuven: Katholieke Universiteit Leuven.

Van der Klink M. and Boon J. (2003), 'Competencies: the triumph of a fuzzy concept', *International Journal of Human Resources Development and Management,* 3, 125-137.

Voskuijl O. and Van Sliedregt T. (2002), 'Determinants of interrater reliability of job analysis: A meta analysis', *European Journal of Psychological Assessment*, 18, 52-62.

Winterton J., Delamare Le Deist F. and Stringfellow, E. (2005), *Typology of knowledge, skills and competences: Clarification of the concept and prototype* (Thessaloniki: Cedefop).

Table 1 Number of questionnaires from experts and jobholders, and number of experts, breakdown per country

	Belgium	France	Germany	Italy	Netherl.	Poland	Spain	UK	Total
Expert q.	204	345	456	26	690	614	404	211	2950
# of experts	127	158	451	22	553	584	373	200	2468
Jobholder q.	183	26	113	139	348	136	202	100	1247
Total	387	371	569	165	1038	750	606	311	4197

Source: EUROCCUPATIONS database

Table 2 Distribution of 160 occupations over seven categories for interrater agreement (rWG) for frequency of tasks and importance of tasks, including the mean number of raters per category.

	Frequency of tasks			Importance of tasks		
	Freq.	%	# raters	Freq.	%	# raters
No agreement (rWG < 0)	39	24.4	32.8	9	5.6	30.6
Lack of agreement (rWG. 00-.30)	38	23.8	26.1	13	8.1	27.7
Weak agreement (rWG. 30-.50)	24	15.0	26.8	21	13.1	33.9
Moderate agreement (rWG. 50-.70)	33	20.6	23.0	35	21.9	25.7
Strong agreement (rWG .70-.90)	16	10.0	27.6	71	44.4	24.3
Very strong agreement (rWG .90-1.00)	1	0.6	14.0	2	1.3	5.5
Missing (<2 raters)	9	5.6	7.3	9	5.6	24.0
Total	160	100.0	26.2	160	100.0	26.2

Source: EUROCCUPATIONS database.

Table 3 Distribution of 160 occupations over seven categories for interrater agreement (rWG) for frequency of tasks and average number of raters per category, break down by country

	Spain			Germany			Netherlands			Poland		
rWG	Freq.	%	# raters	Freq.	%	# raters	Freq.	%	# raters	Freq.	%	# raters
< 0	7	4.4	3.7	27	16.9	4.7	36	22.5	10.8	33	20.6	5.5
.00-.30	8	5.0	4.3	13	8.1	4.0	17	10.6	4.9	15	9.4	6.3
.30-.50	11	6.9	3.8	12	7.5	5.2	22	13.8	7.4	20	12.5	5.4
.50-.70	27	16.9	5.1	16	10.0	5.1	23	14.4	6.6	21	13.1	5.0
.70-.90	44	27.5	3.9	20	12.5	5.5	22	13.8	8.0	26	16.3	4.9
.90-1.00	51	31.9	3.4	34	21.3	2.8	11	6.9	3.5	18	11.3	4.2
missing	12	7.5	1.8	38	23.8	1.1	29	18.1	1.2	27	16.9	2.0
Total	160	100%	3.8	160	100%	3.6	160	100%	6.5	160	100%	4.7

Source: *EUROCCUPATIONS database*

Note: rWG < 0 no agreement; rWG .0-.29 Lack of agreement; rWG .30-.49 Weak agreement; rWG .50-.69 Moderate agreement; rWG .70-.89 Strong agreement; rWG .90-1.00 Very strong agreement; missing indicates <2 raters or missing values on the ratings.

Table 4 Mean country-level interrater agreement (rWG) for frequency of tasks and for importance of tasks per occupation

	N	Minimum	Maximum	Mean rWG	Std. Dev.
mean rWG_frequency of task					
Netherlands	131	-1.86	1	0.24	0.64
Poland	133	-2.02	1	0.28	0.67
Germany	122	-1.68	1	0.40	0.62
Spain	148	-1.21	1	0.70	0.36
mean rWG_importance of task					
Netherlands	115	-0.75	1	0.66	0.40
Poland	103	-0.42	1	0.74	0.27
Germany	97	-1.00	1	0.72	0.40
Spain	138	-0.64	1	0.80	0.23

Source: EUROCCUPATIONS database

Note: N refers to the number of occupations included in the analyses.

Table 5 Distribution of 160 occupations over seven categories for interrater agreement (rWG) for frequency of tasks and mean number of raters per category, break down by country

rWG	Experts and jobholders		Experts		Jobholders	
	Freq.	%	Freq.	%	Freq.	%
< 0	39	24.4	39	24.4	5	3.1
.00-.30	38	23.8	34	21.3	9	5.6
.30-.50	24	15.0	27	16.9	9	5.6
.50-.70	33	20.6	29	18.1	10	6.3
.70-.90	16	10.0	15	9.4	7	4.4
.90-1.00	1	0.6	2	1.3		
< 10 raters	9	5.6	14	8.8	120	75.0
Total	160	100.0	160	100.0	160	100.0

Source: *EUROCCUPATIONS database*

Note: rWG < 0 no agreement; rWG .0-.29 Lack of agreement; rWG .30-.49 Weak agreement; rWG .50-.69 Moderate agreement; rWG .70-.89 Strong agreement; rWG .90-1.00 Very strong agreement; missing indicates <2 raters or missing values on the ratings.

Table 6 Mean rater-category interrater agreement (rWG) for frequency of tasks and for importance of tasks per occupation

	N	Minimum	Maximum	Mean rWG	Std. Dev
meanrwg_frequency of task					
Experts + jobholders	159	-1.53	1.00	0.22	0.45
Experts	151	-1.70	1.00	0.21	0.51
Jobholders	95	-0.97	1.00	0.43	0.38
meanrwg_importance of task					
Experts + jobholders	159	-0.48	0.98	0.59	0.29
Experts	151	-0.67	0.98	0.55	0.33
Jobholders	78	0.23	1.00	0.82	0.13

Source: EUROCCUPATIONS database

Note: N refers to the number of occupations included in the analyses.