ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Extracting Dimensions of Job Quality from Online Employee Reviews

Cath Sleeman

ESCoE Discussion Paper 2024-01

January 2024

ISSN 2515-4664

DISCUSSION PAPER

Extracting Dimensions of Job Quality from Online Employee Reviews
Cath Sleeman
ESCoE Discussion Paper No. 2024-01
January 2024

Abstract

This paper uses a novel dataset of online employee reviews, from Indeed's UK website, to enhance our understanding of job quality. Keywords within the reviews were extracted and then clustered to form a taxonomy of job quality. It is the first UK taxonomy that is derived from employee reviews and the first to be updatable in real time. Analysis of the taxonomy shows that the emphasis placed on different dimensions of work quality has shifted over time, with a marked increase in references to culture, atmosphere and the broader workplace environment. The paper also reveals differences between occupations; workers in some occupations value pay and rewards, while others value the workplace atmosphere and environment. The dataset and methods presented here could form a useful early warning system to detect changes in the dimensions of work quality and to measure the impact of new working practices. More broadly, the paper shows the potential of a novel dataset to improve our understanding of job quality.

Keywords: job quality, wellbeing, web data, machine learning, taxonomy

JEL classification: J28, I3, C45, C38

Cath Sleeman, Nesta
cath.sleeman@nesta.org.uk

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

Extracting dimensions of job quality from online employee reviews

Dr Cath Sleeman, Nesta

Abstract

This paper uses a novel dataset of online employee reviews, from Indeed's UK website, to enhance our understanding of job quality. Keywords within the reviews were extracted and then clustered to form a taxonomy of job quality. It is the first UK taxonomy that is derived from employee reviews and the first to be updatable in real time. Analysis of the taxonomy shows that the emphasis placed on different dimensions of work quality has shifted over time, with a marked increase in references to culture, atmosphere and the broader workplace environment. The paper also reveals differences between occupations; workers in some occupations value pay and rewards, while others value the workplace atmosphere and environment. The dataset and methods presented here could form a useful early warning system to detect changes in the dimensions of work quality and to measure the impact of new working practices. More broadly, the paper shows the potential of a novel dataset to improve our understanding of job quality.

Keywords: job quality, wellbeing, web data, machine learning, taxonomy

JEL Classifications: J28, I3, C45, C38

Acknowledgements: This research has been funded by the Economic and Social Research Council (ESRC), grant ES/S012729/2, as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE).

1 Introduction

This paper shows how a particularly novel dataset might help to identify and monitor the dimensions of job quality in the UK. The UK has no official taxonomy that captures the determinants of job quality, and there is no system to monitor changes in these over time.

The lack of monitoring around the dimensions of job quality is concerning for a number of reasons. Without monitoring, it is much harder to target policies that would curb damaging work practices and, more broadly, improve worker well-being. From an economic viewpoint, this lack of monitoring may even be acting as a drag on productivity. Using UK data and previous studies, Bosworth and Warhurst (2021) found "a clear and compelling link between good work and improved worker productivity".[1] An OECD review by Arends, Prinz and Abma (2017) also found strong evidence for a "negative relationship between job stress or job strain and individual at-work productivity and for a positive relationship between job rewards and productivity" (page 4).

The novel dataset used in this paper is 1.5 million reviews left by employees on Indeed's UK website (hereafter *Indeed*). In each review, a current or former employee discusses their experience of working at a specific organisation. The reviews are public and are principally intended to help job seekers gain insight into different workplaces.

The main text fields in these reviews have never been used to study job quality, either in the UK or abroad. With this unique dataset, the paper presents a method for extracting the dimensions of job quality directly from the reviews. Workers are free to write about any aspect of their work experience in as much detail as they wish. By contrast, in a survey-based approach (such as the CIPD's Good Work Index) workers can only provide their views on the questions posed. The dataset of reviews can potentially surface more granular dimensions of job quality. On the flipside, surveys can ensure a representative sample of workers, which is not the case for employee reviews. The full list of limitations, in regards to employee reviews, is explored in subsection 3.6 and the representativeness of reviewers (across occupations) is measured in subsection 5.1.

In regards to benefits, the method presented in this paper provides an inexpensive mechanism for the timely detection of changes in the dimensions of job quality. Employee reviews could act as an early detection system for the impact of new work practices. It would also expose the occupations which are most impacted by these practices.

[1] The quote is from this blog which accompanies the paper.

There are several definitions of job quality and a number of closely related concepts (such as ‘decent work’ and ‘good work’).[2] The method presented in this paper does not rely on adopting a particular definition of work quality, because the dimensions of job quality are surfaced directly from the opinions of employees. However, the method does make a significant assumption: namely that the topics mentioned by workers in their public reviews are the same topics that affected their well-being. One point of support for this assumption is that *Indeed* actively encourages workers to give their opinions about the parts of a job that they found hardest and those that they found most enjoyable. A second assumption, that can’t be met, is that the reviews are representative of *all* workers. This is because not all workers are equally inclined to leave a review and the content of their reviews may be influenced by being public and by seeing the reviews of colleagues. Due to these biases, the taxonomy will likely be more representative of some sectors than others. This assumption is explored further in subsection 5.1.

2 Literature review

The literature review, below, highlights the three unique contributions of the paper. This is the first study to examine the *main review field* within *Indeed* reviews. Other papers have focussed either on the much shorter *pro* and *con* fields or only on the *numerical ratings* left by reviewers. The main review field provides a much richer picture of employees’ work experiences. This is also the first study (among those that use online reviews) to construct a *multi-level* taxonomy of work quality. Other studies have extracted a set of topics, at a single level of granularity. A multi-level approach allows the identification of smaller, but still important, aspects of work quality that could be missed in a coarser clustering. Finally, none of the papers discussed below have considered variations in the dimensions of job quality, either overtime or between occupations. These variations are explored in subsections 5.4 and 5.5.

2.1 Studies using *Indeed* reviews

Only two other studies have used *Indeed* reviews. Both are focussed on US-based organisations and neither use the main text of the review. Sainju (2020) used the reviews to identify the determinants of job satisfaction. The paper used reviews of Fortune 50 companies and utilised the *pro* and *con* fields only. These fields contain a string of words (e.g. ‘free food, good hours’) and are usually much shorter than the main review field. The paper found that in most cases, there are less than 10 tokens in these fields. Howe and Bisel (2020) also used *Indeed* reviews, but focussed on the relationship between the sentiment of the review and how helpful others found that review (visitors can rate a review as ‘helpful’ or ‘unhelpful’). They focussed on US military and US corporate organisations.

[2] Warhurst et. al. (2017) provides an exhaustive summary.

2.2 Studies using Glassdoor reviews

A number of papers have used reviews from Glassdoor, which is an alternative site on which workers can leave reviews. The studies mentioned below are those that concentrate on job quality or use the text of reviews.

Symitsi et. al. (2020) examined 349,550 reviews from Glassdoor, left by workers at 40,915 UK firms. The structure of Glassdoor reviews enabled the authors to separate positive and negative comments. They used topic modelling to identify 12 positive and 12 negative topics. The paper examined the information content of these dimensions to predict variables such as employee turnover and firm profitability.

Dabirian et. al. (2016) examined 38,000 reviews of the highest and lowest ranked employers on Glassdoor. The authors used IMB Watson (an artificial intelligence tool) to extract seven 'branding value propositions' that current, former, and potential employees care about when they collectively evaluate employers.

Ramos et. al. (2020) examined 15,000 reviews of 15 major IT companies based in the US. They used a Support Vector Machine model to identify the 10 frequent nouns (and other features) that were most salient to the overall score granted by the reviewer.

Finally, Jung and Suh (2019) used an alternative website, called Jobplanet.co.kr, which is the largest online company review site used by employees in South Korea. They gathered 232,400 reviews from 2014 to 2017. By using topic modelling (specifically Latent Dirichlet Allocation) they obtained 30 factors of job satisfaction.

In comparison to these studies, this paper considers a much larger number of reviews and a larger number of firms. It is also the first to use these reviews to create a *multi-level* taxonomy of job quality, and examine how the dimensions of work quality vary across occupations and over time. The advantage of a multi-level taxonomy is that it allows smaller, but still important, aspects of work quality to be identified (which may be missed in a single-level taxonomy).

2.3 The measurement of work quality in the UK

In 2016, the Government commissioned Matthew Taylor to lead a review examining the impact of modern business practices on the world of work. The Taylor Review of Modern Working Practices was published in 2017 (Taylor, 2017), and recommended that the Government place quality of work on an equal footing with quantity of work when measuring success. Moreover, it specifically recommended that "The Government should identify a set of metrics against which it will measure success in improving work, reporting annually on the quality of work on offer in the UK" (Taylor, 2017, page 103).

Later that year the Government's Industrial Strategy (HM Government, 2017, page 118) accepted Taylor's recommendation to identify a set of measures against which to assess job quality and success. It also laid out five aspects that they believed to be foundational: overall worker satisfaction; good pay; participation and progression; well-being, safety and security; and voice and autonomy. This was followed, in December 2018, by the Government's Good Work Plan (HM Government, 2018), which revealed that the Government had asked the Industrial Strategy Council to consider the most appropriate way in which to measure quality of work. However the Council was abolished in 2021, following the release of the Government's Build Back Better plan (HM Government, 2021). Although the plan emphasised the need to create 'high-quality work' (page 22), there were no references to the measurement of work quality.

While there is still no official measure of work quality, others have made progress. Feltstead et. al. (2019) developed a job quality quiz (based on questions in the Skills and Employment Survey) which covers 10 domains (and has a particular focus on work intensity and opportunities for learning). Another example is the Chartered Institute of Personnel and Development (CIPD) who publish an annual Good Work Index (Norris-Green and Wheatley, 2022), based on a survey of over 6,000 workers across the UK. The survey began in 2018 and it reports on seven dimensions of work quality, shown in Figure 1. These dimensions were based on a review by Warhurst et. al. (2017), who identified six aspects of work quality. The CIPD then split one of these aspects into two ('intrinsic characteristics of work' became 'job design and the nature of work' and 'social support and cohesion') (page 3 of Sarkar and Gifford, 2018).

The seven dimensions of job quality, by the CIPD, were then enriched by the Measuring Job Quality Working Group (Irvine et. al, 2018). This independent group was set up following the Taylor Review to consider the challenges of measuring job quality. Their final report contained 18 metrics to capture the CIPD's seven dimensions of Work Quality, and these are shown in Figure 1 below. Subsection 5.2.3 compares these seven dimensions (and 18 metrics) to the topics extracted from the online employee reviews.

At the end of 2022, the ONS published a standalone analysis on eight of the 18 metrics, using data from the Annual Population Survey (Baeck and Uzzell, 2022). This was an update to, and expansion of, a previous piece published in 2019. Finally, it is important to mention the many (and ongoing) *international efforts* to measure job quality, including [the ILO's Decent Work Indicators](), [the OECD's job quality framework]() and [Eurofound's European Working Conditions Surveys (EWCS)]().

Figure 1: The CIPD's seven dimensions of work quality and the 18 metrics recommended by the Measuring Job Quality Working Group. Source: Irvine et. al. (2018, page 5).

3 Data

The dataset for this study was provided by *Indeed* to Nesta, under a Data Sharing Agreement. It consists of just over 1.5 million reviews (all publicly available) that were left by current and former employees on *Indeed's* UK site between March 2012 and April 2019 inclusive. This section provides further details on the reviews, as well as descriptive statistics and a discussion of the dataset's strengths and limitations.

3.1 The process of leaving a review

Anyone is permitted to leave a review on *Indeed's* website. The reviewer does not need to prove that they are, or were, an employee of the organisation that they are reviewing. *Indeed* reserves the right not to post a review if it violates their standards[3]. These standards include, but are not limited to, prohibiting reviews on the basis of obscene language, spam content, contact information, critical comments on other reviews, allegations of illegal activity and defamatory content. Businesses cannot remove reviews left by employees. They can post public replies to the reviews, though these replies were not part of the dataset for this study. Anyone can flag a review for removal but they must explain why it breaks *Indeed's* terms of service[4].

3.2 The structure of a review

Figure 2 shows an example of an employee review which is contained in the dataset. The review begins with a title, chosen by the reviewer, and their overall rating (on a scale of 1 to 5 stars, inclusive). This is followed by a row of text giving the job role of the reviewer, whether they are a former or current employee, the location of their job and the date of their review. Beneath those fields is the text of the review (henceforth referred to as the *main review field*), which is the focus of this study. This block of text is sometimes followed by short lists of pros and cons (though not all reviewers complete these fields, as in the example shown). A reviewer can also give ratings for categories such as work/life balance and management. Below the review, visitors to *Indeed's* site can indicate if they found that the review was helpful. They can also share or report the review.

[3] Available at https://uk.indeed.com/survey/review?attributionid=review-this-company-btn-reviews-tab&cid=63b9122fbd20f7f9.
[4] Available at https://indeed.force.com/employerSupport1/s/article/205297680?language=en_US.

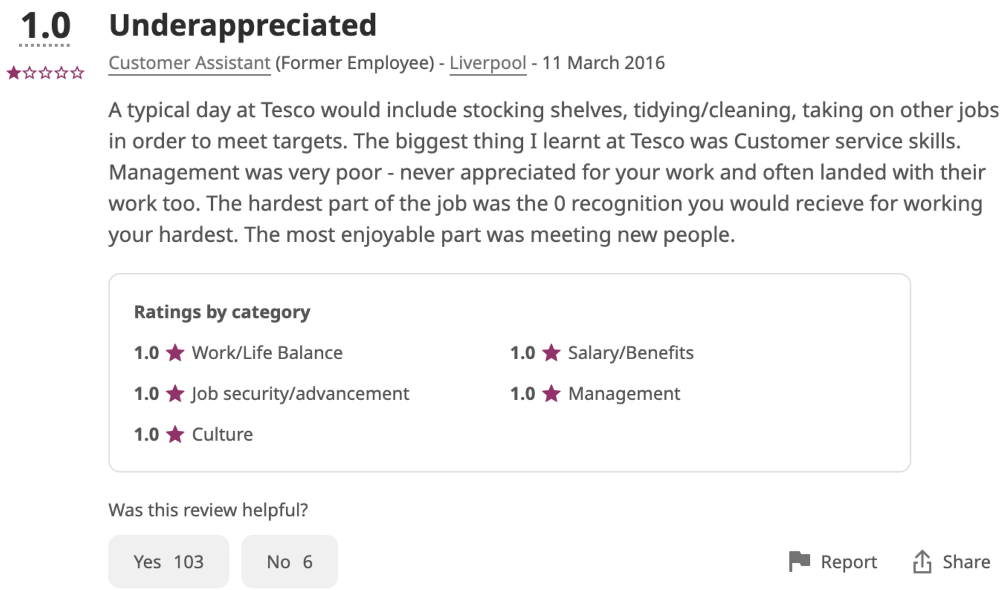

Figure 2: An example of a review (available at https://uk.indeed.com/cmp/Tesco-40b6123a/reviews/underappreciated?id=76a9a9b8c0c84163).

3.3. The main review field

The main review field has a minimum of 150 characters. Reviewers are free to write whatever they like, subject to the guidelines mentioned above. *Indeed* also cautions reviewers against providing any confidential company information or personally identifiable information. To assist in their writing, reviewers are presented with a set of prompts (see the bullet points on the right-hand side of Figure 3 below). The impact of these prompts on the content of the reviews is explored in subsection 5.4.

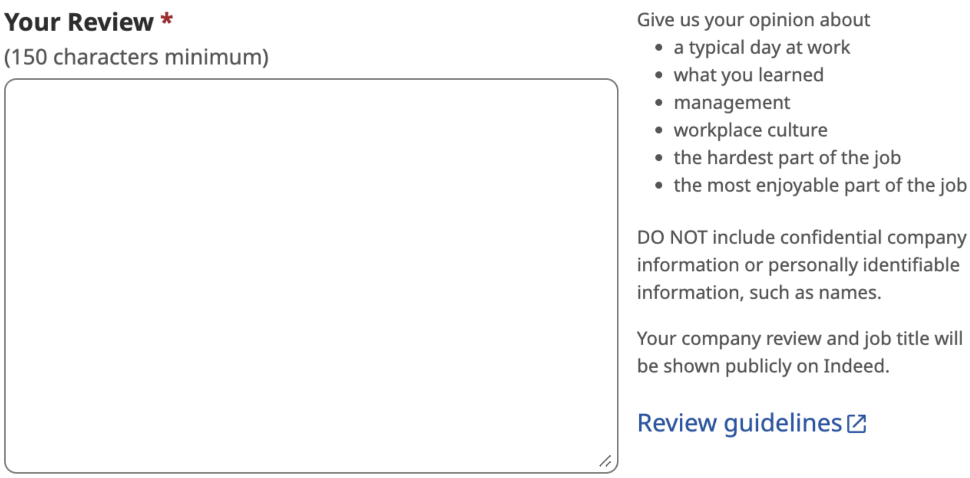

Figure 3: Part of the webpage presented to workers when they leave a review (in August 2022).

Most reviewers touch on three aspects in the main review field. They typically begin by describing the tasks and duties in their role, such as stocking shelves and tidying/cleaning from the example in Figure 2. The reviewer will then go on to describe certain features of their job which impacted their well-being. In the review shown above, these features were management, level of recognition, and the ability to meet new people. It is these features that the paper aims to extract from the reviews and capture within a Quality of Work taxonomy. The final aspect in most reviews is the worker's feelings towards those aforementioned features, such as 'enjoyment' or feeling 'never appreciated', as in the review shown in Figure 2.

3.4 Descriptive statistics

Figure 4 below shows the number of reviews that were left each month. While the average number was around 17,500 per month, this figure varied substantially and passed 35,000 on two occasions. This fluctuation was likely due to advertising campaigns that encouraged workers to leave reviews. The variation is unlikely to impact the dimensions of work quality mentioned within reviews, and subsection 5.4.1 shows that the breakdown of occupations by reviewers did not change substantially over the sample period.

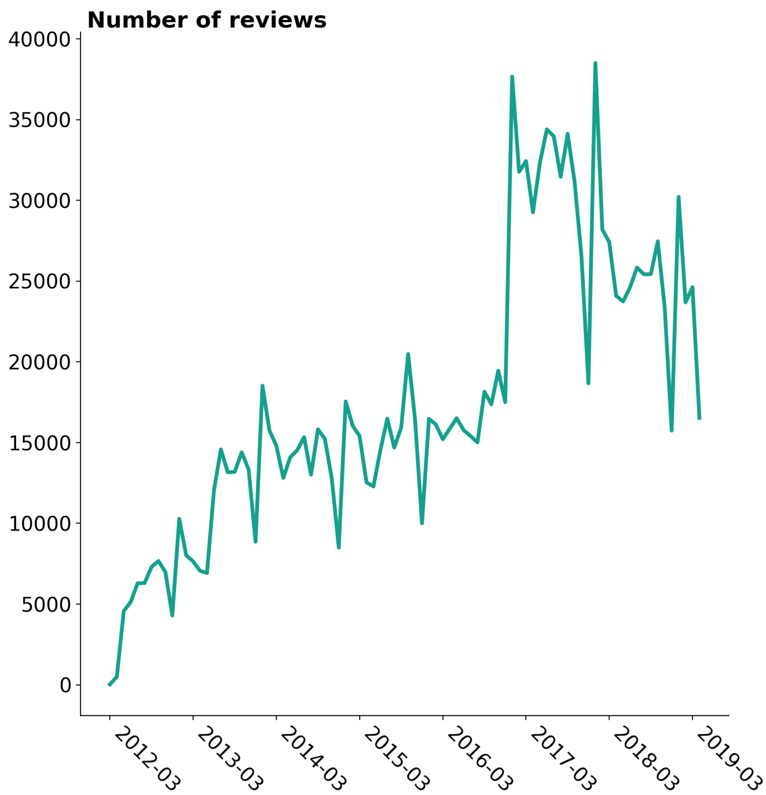

Figure 4: The number of reviews left per month.

Turning to the organisations that were reviewed, there are approximately 743,000 unique organisations in the dataset. The number of reviews per business is highly skewed. Just under 85% of organisations were reviewed once, and 97% had five reviews or less. At the other end of the distribution, the five companies with the highest number of reviews were Marks and Spencer (4,400), ASDA (5,600), J Sainsbury Plc (7,200), Tesco (11,000) and McDonald's (12,000).

This skewed distribution likely reflects the fact that the most-reviewed companies are all large employers in sectors that typically have a high turnover of employees. Both their size and rate of employee turnover generates a large pool of potential reviewers.

Another factor that impacts the shape of this distribution is a tendency by reviewers to use variations of their employer's name. For example, in addition to the 11,000 reviews for 'Tesco', there were 13 reviews for 'tesco distribution centre', 4 reviews for 'tesco's', 2 reviews for 'tesco extra dudley' and many other variations. In total, an additional 1,000 unique organisations in the dataset contained the word 'tesco'. The reviews for these organisations should have been subsumed within the organisation called 'Tesco'.

Finally, the features of the main review field are described. In a random sample of 100,000 reviews, there was an average of 3 sentences in each review, with a median of 2 sentences. In regards to tokens (which are similar to words), there was an average of 54 tokens per review, with a median of 37.

The text of reviews were also compared to detect duplicates. One review was deemed to be a duplicate of another if it contained the exact same text as the original review in the main review field. In a larger random sample (50% of the complete dataset) there were 4,200 reviews that were a duplicate of at least one other review in the sample. This is less than 0.6% of the sample. The most common case of duplication occurred when a reviewer simply copied the prompts suggested by *Indeed* (shown Figure 3) and pasted these into the main review field. In these instances, the reviewer may have wanted to rate the company without leaving a review.

3.5 Strengths of employee reviews

There are two benefits of using employee reviews to study job quality. First, the blank box in which workers write their reviews gives them complete freedom to speak about any aspect of their work life. This freedom enables the detection of new dimensions of job quality as they emerge. By contrast, in a survey-based approach to measuring job quality, the surveyor must presuppose the important aspects of job quality and workers are then constrained to give their thoughts on just these aspects. It is important to mention, of course, that in some instances this systematic approach can be a strength of surveys (and a weakness of reviews), namely where workers are unaware of job aspects that are posing risks to their mental or physical health (Felstead et. al., 2019).

A second benefit is that employee reviews have the potential to provide very timely insights on job quality. Although this particular dataset ends in 2019, new reviews are continually being left on the *Indeed* website. Moreover, these reviews can be collected at no cost. In contrast, surveys are typically expensive to run and there is inevitably a delay between the survey being carried out and the results of the survey being published.

3.6 Limitations of employee reviews

The limitations of employee reviewers come from their potential for bias and noise.

3.6.1 Bias

The first source of bias originates from the workers who choose to leave a review. This group may not be representative of the wider workforce, and may have different concerns in relation to job quality. The extent of this bias depends on a number of factors, including awareness of the *Indeed* website, as well as the perceived risks and benefits of leaving a review. For example, self-employed workers may be less inclined to leave a review. Subsection 5.1 examines which occupation groups are under or over-represented in the dataset.

A second source of bias relates to the content of reviews and the dimensions of work quality that reviewers choose to mention. The public nature of reviews, and the presence of previous reviews, may both influence the topics touched upon by a reviewer. Unfortunately it is not possible to gauge the extent of this bias.

3.6.2 Noise

There are two respects in which this dataset is ‘noisy’. First, the main review field mixes descriptions of tasks with features of the job that affected the reviewer’s well-being. A reviewer may mention ‘back pain’ in relation to the conditions that they treated (for example, as a physio), or the phrase may refer to pain that they experienced as a result of lifting boxes. This mixture of tasks and aspects of job quality can be difficult to disentangle. For this reason, the paper avoids exploring health conditions or other dimensions of work quality that might also be mentioned as tasks within a job. Instead, subsection 5.5 concentrates on ‘pay’ and ‘atmosphere’, as these two features of work quality are rarely referred to as duties within a job. This limitation may also make it difficult to discern the degree of 'decision latitude' that workers have in their jobs, which is an important component of job stress (Karasek, 1979).

Second, many of the reviews are written in a very informal style and this complicates the task of extracting the dimensions of work quality. Misspellings and mistakes in grammar are numerous. For example, 'manager' is misspelt as 'manger', 'maneger' and 'mananger'. The frequency of spelling mistakes rules out the use of pre-trained language models, while grammatical mistakes make it more difficult for position-of-speech taggers to accurately identify the nouns within the reviews. The Methods section, below, describes the measures taken to limit the impact of this noise.

4 Methods

There are two methodological components of this study. The first is an algorithm that assigns job titles (left by reviewers) to standard occupation codes (SOC). The purpose of this algorithm was to assess the occupational groups that are under- and over- represented in the reviews. The second, and main, component is a method that creates a taxonomy of work-related terms, using the main text of the reviews.

4.1 Assigning job titles to occupational groups

To assess the under- and over- representation of occupations in the dataset, the reviewers' job titles were assigned to occupational groups. The proportion of reviews in each occupational group was then compared to the proportion of all workers in the same group. The latter proportion was based on employment figures from the Annual Population Surveys between 2012/13 and 2018/19.[5] Over this period, the highest proportion of workers (9.2%) were employed in elementary administration and service occupations, followed by 8.2% in administrative occupations.

The occupational groups used were the set of 2-digit 2010 SOC codes.[6] While the 3-digit and 4-digit levels would have provided more granular insights on representation, many job titles left by reviewers were too broad (e.g. 'engineer'). These titles could not have been assigned to the narrow 3-digit and 4-digit occupations (which distinguish between 'civil engineer' and 'electrical engineer').

The algorithm, described below, takes a reviewer's job title and assigns it to one of the 25 2-digit occupational groups. A similar algorithm, called CASCOT, was developed by the Warwick Institute

[5] The survey results are based on the year starting in April and ending in March.
[6] Employment data for that period is not available for the more recent 2020 SOC groups.

for Employment research. The tool appears to have good accuracy but unfortunately the methodology is not open and the tool is not free for applying the algorithm to thousands of job titles.

4.1.1 Method

The algorithm, described below, relies heavily on the ONS's SOC 2010 Coding Index. This Index is a list of over 36,000 job titles. For each title, the ONS provides the most appropriate 4-digit SOC code.

The algorithm contains five steps, which are shown in Figure 5. These steps reflect the need to balance speed with accuracy.[7]

Step Zero: Cleaning. Before applying the algorithm, the job titles from the reviews and the job titles from the coding index are both cleaned. This process includes converting the text to lowercase and removing most punctuation. Both sets of titles are also standardised: US spellings are converted to UK spellings (e.g. 'organizer' is changed to 'organiser'), common abbreviations are manually identified and expanded (e.g. 'kp' is expanded to 'kitchen porter') and certain words and phrases are either combined or separated (e.g. 'storekeeper' is changed to 'store keeper').

A number of reviewers left their job title blank or wrote variations of 'NA', 'prefer not to say' or 'worker'. A list of these phrases was assembled to ensure that they were not assigned to occupational groups. In total, 0.6% of the sample did not contain a job title. Additionally, some reviewers described themselves as volunteers or students. The experiences of these groups appeared to be quite different to those of workers. For that reason, their reviews were excluded from the rest of the analysis; they comprised 2.1% of the sample.

Step One: Exact match. The algorithm searches for the reviewer's exact job title in the ONS's Coding Index. If a match is found, the corresponding 2-digit SOC code is returned.[8]

Step Two: Two-word match. If an exact match cannot be found, the algorithm searches for two words from the reviewer's job title that are contained within a title from the Coding Index. Stop words are excluded, as are words that do not indicate a reviewer's occupation (such as 'part', referring to 'part time'). The two words need not be adjacent to each other, in either the reviewer's job title or the

[7] Approaches that considered the semantic meaning of the word were disregarded for their lack of speed.

[8] If multiple matches are found, the job title that excludes qualifiers is chosen. Qualifiers are extra fields within the Coding Index which indicate that the job title should only be assigned to that SOC code when the job is located in a particular industry or meets an additional qualifier. For example, an 'Administrator' is assigned to SOC code 4159 in the SOC Coding Index, while an 'Administrator' with the qualifier of 'hospital service' (i.e. they are a Hospital Service Administrator) is assigned to SOC code 3561. If all the titles contain qualifiers, then the most common 2-digit SOC code is assigned.

Coding Index. In instances where multiple matches are found, the algorithm chooses the most common 2-digit SOC code from the matches.

Step Three: One digit gap. If a two-word match is not found, the algorithm searches for a job title in the Coding Index that differs by only one letter from the reviewer's title. This small step is intended to catch minor spelling mistakes, such as a reviewer who describes their job as 'shop asistant'. Allowing two or more characters to differ led to inaccuracies and was prohibitively slow.

Step Four: TF-IDF. In this step, the algorithm identifies terms within the reviewer's job title that are closely associated with a particular SOC code. This association is measured by constructing a term frequency-inverse document frequency (TF-IDF) matrix from the Coding Index. The 'documents' are the 25 2-digit SOC codes and the 'terms' in each document are the concatenated job titles which have been assigned to these codes. Very common words (that appear in over 85% of documents) are excluded, as are stop words and non-informative terms. Words within the reviewer's job title are then vectorized and the cosine similarity between the title and each 2-digit SOC code are calculated. If the maximum similarity for a SOC code exceeds a threshold of 0.1, then that SOC code is chosen.

Step Five: Single word match: The final step looks for single words within the reviewer's job title that exactly match to a single-word job title from the Coding Index.[9] For each word, these matches are stored and the most popular SOC code is chosen.

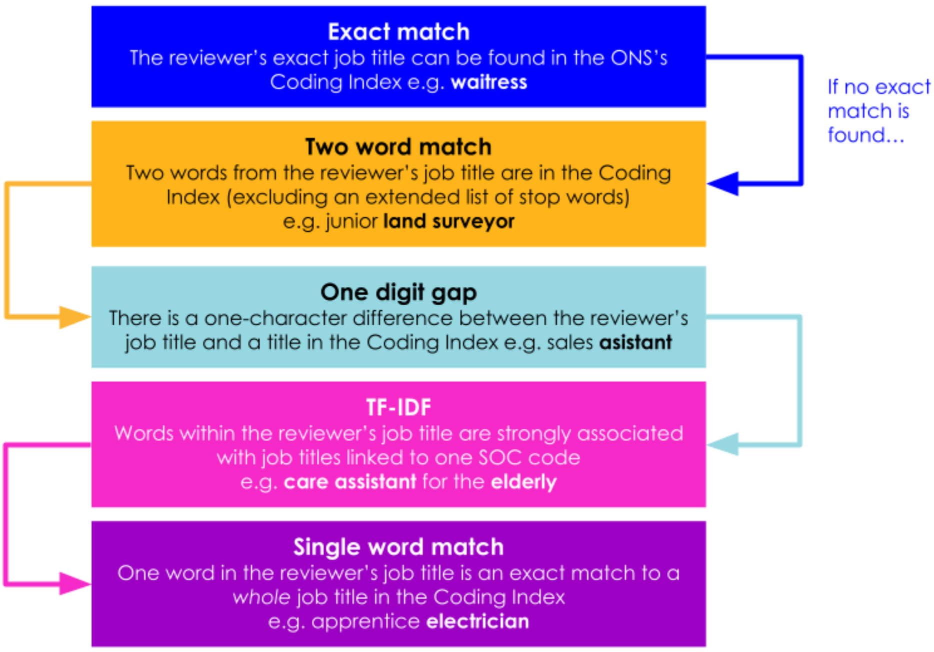

Figure 5: The five steps within the algorithm that assigns job titles to occupational groups.

[9] The word 'assistant' is ignored if the job title contains other words and if it is the first word within the title (e.g. assistant accountant). In this case the word does not help to identify the most appropriate SOC code.

4.1.2 Comparison with CASCOT

Assigning job titles to occupational groups is necessarily subjective, and so it was not possible to evaluate the algorithm against a ground truth. Instead, two comparisons were carried out. In the first, the 100 *most common* job titles, left by reviewers, were identified and the 2-digit SOC codes predicted by the algorithm above were compared with those predicted by the CASCOT algorithm[10]. In the second comparison, this exercise was repeated using a *random* sample of 100 job titles. In both comparisons there were reasonable levels of agreement. Amongst the 100 most popular titles, CASCOT and this paper's algorithm agreed on 78% of occasions (at the 2-digit level).[11] In the random sample, the agreement was 69%. The disagreements occurred mainly where this paper's algorithm did not output a SOC code for the job title, while CASCOT did (for titles such as 'intern' and 'work experience').

4.2 Building a taxonomy of work-related topics

The second, and main, component of the methodology required constructing a taxonomy of work-related topics from the main review field. The taxonomy contains multiple levels to capture both broad and narrow topics, and each level consists of branches that refer to different work-related topics.

4.2.1 Extracting common nouns and measuring similarities

The first steps to construct the taxonomy are extracting a set of nouns that will reside within the taxonomy and then creating a network which captures the semantic similarity between these terms. This process is described below in Figure 6.

[10] https://cascotweb.warwick.ac.uk/#/classification/soc2010.
[11] Instances where both algorithms failed to predict a SOC code (e.g. for 'student') were counted as agreement.

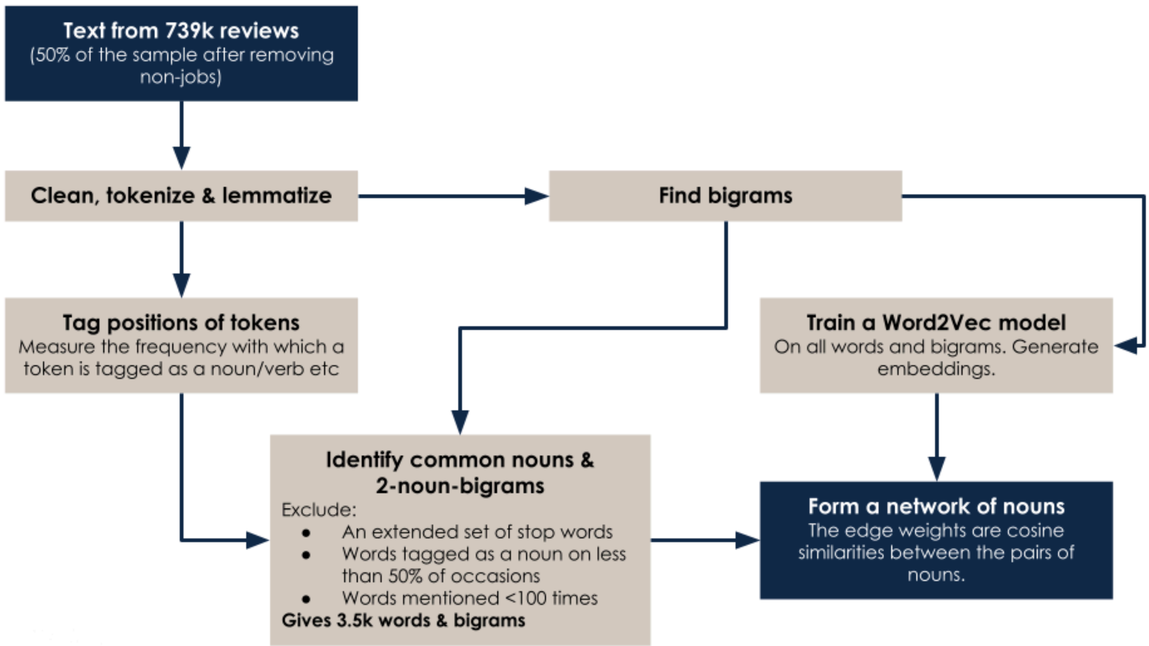

Figure 6: Process to create a network of nouns from the reviews.

The starting point was a set of 739,000 reviews, which represents a random 50% sample of the dataset (after excluding reviews from students, volunteers, interns and those undertaking work experience). The text in the reviews was reduced to lower case and most punctuation was removed. The reviews were then tokenized - splitting each review into sentences and then into tokens. Each token was lemmatized, taking inflected forms of words back to their base form. Bigrams[12] were identified, such as ‘workplace culture’ and ‘team leader’.[13]

A Word2Vec model was then trained on the cleaned tokens and bigrams. This model is a shallow neural network and uses a continuous-bag-of-words approach to predict a target word based on the surrounding words. The hidden weights within the model allow each word to be represented as a numerical vector, which in turn, enables the measurement of semantic similarity between any two words.

The decision to estimate new embeddings, rather than using those from a pre-trained model, was driven by the domain-specific nature of the corpus and its unique features. The reviews contain many terms where their meaning within a work context is different to their most common meaning in a general corpus of text. Examples of these words include ‘culture’ and ‘promotion’. The Word2Vec model can capture these context-specific meanings and is particularly good at detecting the meaning

[12] A bigram is a sequence of two adjacent tokens, which are usually words.
[13] Only those two-word combinations that occurred at least 100 times, across the sample of reviews, were considered for bigram status.

of commonly misspelt words and abbreviations. As an example, the figure below shows the terms that the model detects as being most similar to 'manager'. It correctly detects the misspelling of manager ('manger') and successfully identifies 'gm' (general manager) and 'md' (managing director).

No.	Words and bigrams that are most similar to 'manager'
1	manger
2	supervisor
3	regional manager
4	gm
5	team leader
6	management
7	manageress
8	managing director
9	director
10	md

Figure 7: The ten words that are semantically most similar to 'manager' in the trained Word2Vec model.

The most common nouns (words and bigrams) were then extracted. Only words that were identified by the position-of-speech tagger as being nouns on at least 50% of occasions were kept.[14] Separately, nouns that were mentioned less than 100 times were dropped. Stopwords were also excluded, and the standard list of stopwords was extended to include a further 51 terms. The extended list included terms that had been incorrectly identified as nouns (e.g. 'shes'), terms in which two words had been concatenated (e.g. 'workersthe') and terms that lacked meaning (e.g. 'thing'). This process left a total of just over 3,500 nouns on which to build the taxonomy. A network was constructed from these nouns in which the edge weight between any two nouns captured the cosine similarity of their embeddings.

4.2.2 Hierarchical clustering

The hierarchical clustering was performed using a Python package written by Karlis Kanders called *cluster-utils*, which performs an adapted version of a method outlined by Poulin and Theberge (2018). The two authors proposed a new method of consensus clustering which produces stable results and

[14] For bigrams, each word had to pass this threshold.

significantly reduces the resolution limit problem (which is the minimum size of a community that can be found by clustering).

The *cluster-utils* package, which implements a version of the authors' approach, consists of two steps, which are outlined below in Figure 8. In the first step, the Leiden clustering algorithm is run repeatedly on the network of words and the results of the clusterings are stored. These clusterings are used to create a new network. The edge weight between any two words in this new network is based upon how frequently the two words were clustered together in the first step. In the second step, the Leiden algorithm is run repeatedly on this new (more stable) network. An adjusted mutual information metric was calculated for each clustering and the one that agreed most with all the other clusterings was chosen as the final consensus partition. In this instance, the Leiden algorithm was run 100 times in the first step and 50 times in the second. The results were not sensitive to these two values.

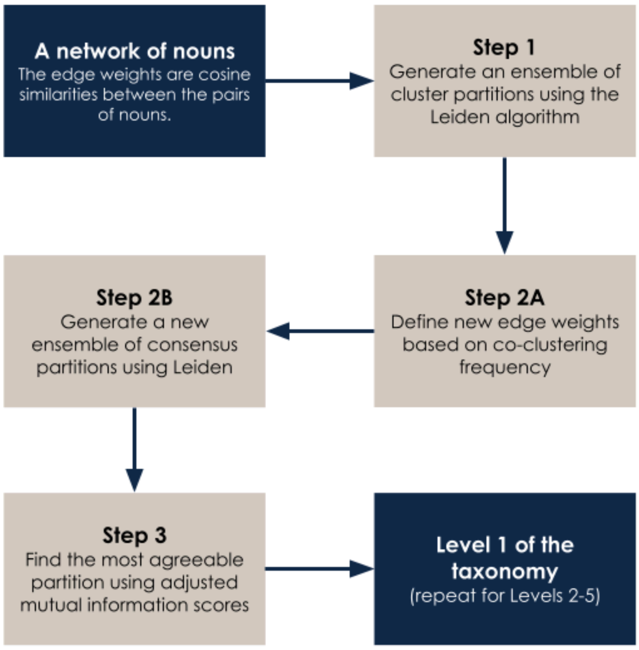

Figure 8: Steps to form a work taxonomy from a group of common nouns extracted from reviews.

The two-step method, described above, only generated the top (or first) level of the taxonomy. In order to give depth to the taxonomy, this clustering procedure was repeated separately for the nouns in each first-level cluster. This gave a second level, containing more granular groups of nouns. This process was repeated three more times to create a final taxonomy of nouns that contained five levels.

In the final step, each cluster in the taxonomy, at each level, was named. For those clusters containing three or fewer nouns, their names were formed by simply joining the nouns together (e.g. 'salary, compensation, remuneration'). For larger clusters, the mean similarity between each noun (and every other noun in the cluster) was calculated. The three nouns with the highest mean similarity were joined together to form the cluster's name. Many other approaches were tried but this method was found to produce the most sensible names.

5 Results

As mentioned above, all the results are based on a random sample, representing 50% of the reviews (or approximately 755,000). This random sample contains approximately 231,000 unique job titles.

5.1 The representation of occupations

This subsection uses the occupational assignment algorithm (described in subsection 4.1 above) to explore how occupations are under- or over- represented in the reviews. As a reminder, the algorithm uses five steps to find a SOC code. The chart below shows the specific step that successfully found an appropriate SOC code. Almost half of the 755,000 job titles (which includes many duplicates) were assigned to an occupation group by finding an exact match within the ONS's SOC Coding Index. Most others were assigned using either the two-word match or the TF-IDF steps. Only 10.5% of job titles could not be assigned to an occupation. A frequent reason for an unassigned job title is that a reviewer's title did not contain enough information.

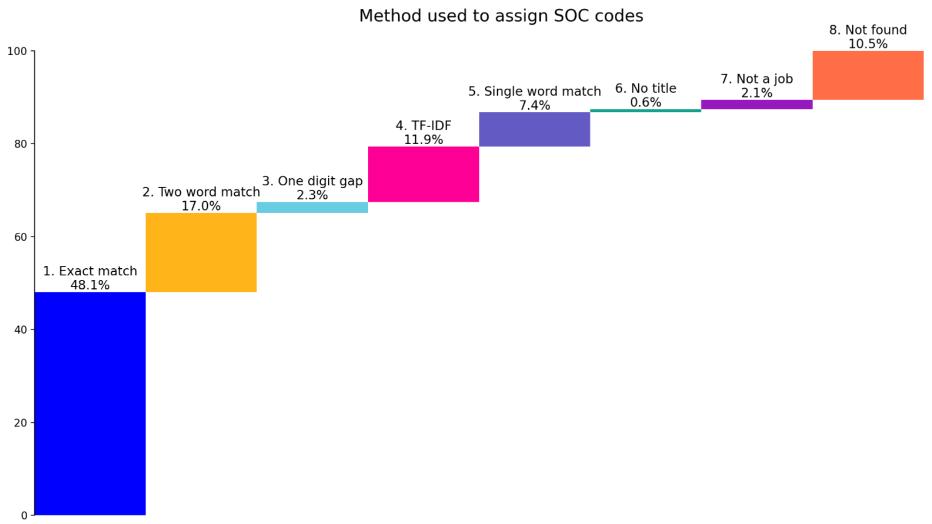

Figure 9: The method used to assign all job titles to SOC codes (including duplicates).

The distribution of occupational codes amongst the reviews was then compared to official employment data (from the Annual Population Surveys between 2012/13 and 2018/19 inclusive). The dates were chosen to mimic the range of the sample. The figure below shows the representation quotient for each 2-digit SOC code - this is the proportion of reviews divided by the proportion of UK employees. Values above one (in blue) indicate that the proportion of reviews in the *Indeed* dataset exceeds the same proportion in the wider workforce. Reviews from workers in Customer Service Occupations are vastly overrepresented; their proportion of reviews is more than 2.5 times their proportion of employment. Several elementary occupations and sales occupations are also overrepresented.

The representation quotient will reflect two aspects: worker's likelihood of leaving a review and the turnover within each occupation.[15] Both aspects may be contributing to the overrepresentation of reviews from Customer Service Occupations.

Conversely, reviews from Health Professionals and several skilled trades are underrepresented. For each of these groups, their proportion of reviews is less than half their proportion of employment. This may be because many Health Professionals are employed by a single organisation (the NHS), and so there are few alternative employers to compare and contrast. For those working in skilled trades, many are self-employed and so may have less incentive to use the platform or they may have recorded their job title as 'self-employed' which could not be assigned to an occupation group.

[15] Their chance of leaving a review will depend, in turn, on factors such as their use of *Indeed's* job search platform.

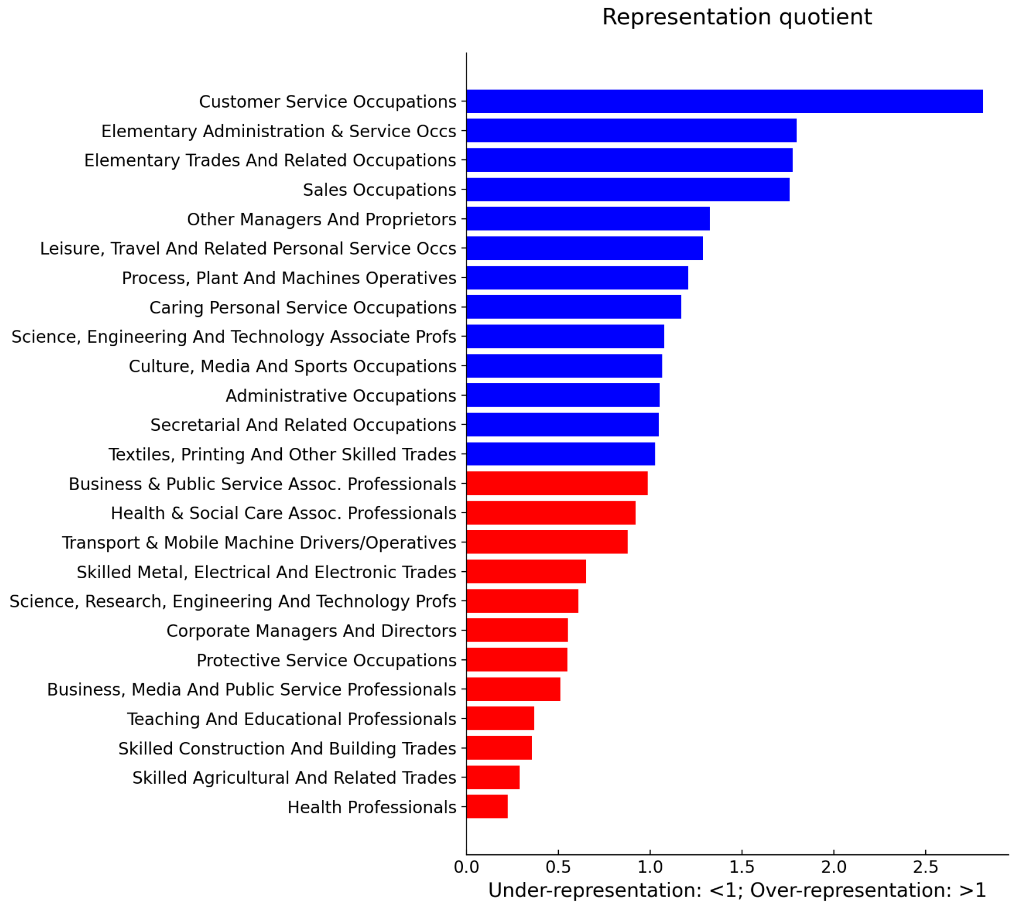

Figure 10: The under- and over- representation of occupations amongst the reviews.

In light of these results, the reviews could be resampled (with weights) to improve representation. However, given the extent of underrepresentation, this process would require repeatedly sampling the same reviews or drastically cutting the sample size. For that reason, the sample was not reweighted and instead it is acknowledged that the resulting Quality of Work taxonomy (shown in the next section) is more representative of those occupations in blue (in Figure 10).

5.2 A UK work taxonomy

5.2.1 Whole taxonomy

The figure below shows the first two levels of the whole taxonomy. The first level of the whole taxonomy contains four branches. The top branch (which is called ‘emotion, incompetence, consequence’) captures words that relate to the dimensions of workers’ well-being. It is this branch that is of key interest in the paper, and it is subsequently referred to as the ‘Quality of Work taxonomy'. This branch contains just over 850 words and mentions of these words across the reviews account for around 50% of the mentions of all nouns that are contained in the whole taxonomy.

The remaining three branches of the whole taxonomy capture other aspects of work. The second branch contains words that describe job titles and workplaces. The third and fourth branches contain tasks within a job and items used to carry out those tasks. Specifically, the third branch refers to items used in skilled trades and elementary occupations, while the fourth branch refers to items for desk-based jobs.

Although the first branch is entirely related to dimensions of job quality, there are a handful of other clusters, capturing dimensions of well-being, that are situated in other branches of the first level. For example, the third branch ('tube, tank, bucket') contains a cluster relating to shift patterns and another relating to the weather. The 'remote' location of these branches is likely due to the fact that they are not universal dimensions of work quality and only affect workers in some occupations.

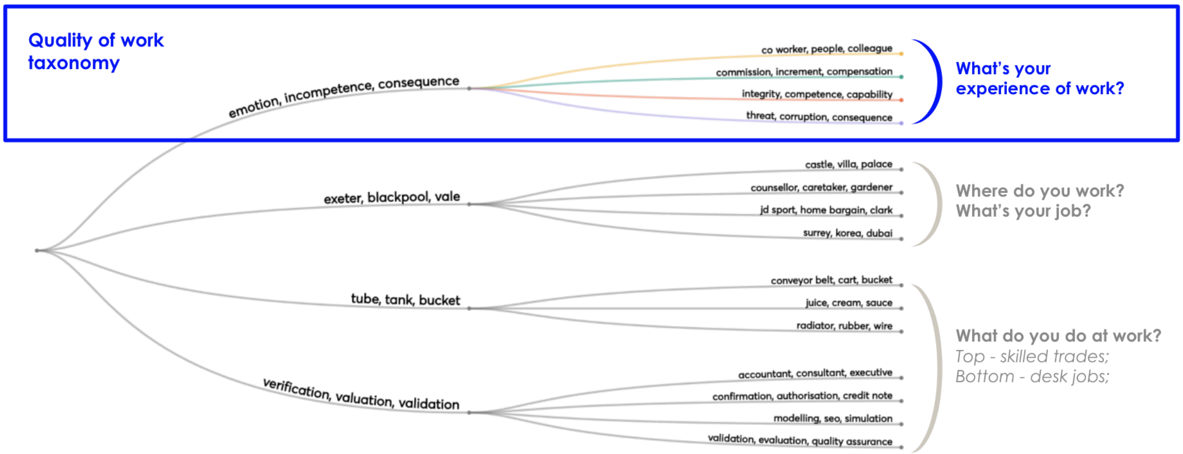

Figure 11: The first two levels of the whole work taxonomy.

5.2.2 Quality of Work taxonomy

The figure below shows the first three levels of the Quality of Work Taxonomy (which, as mentioned, come from the first branch of the whole taxonomy). The names of some clusters do contain spelling errors because these (misspelt) terms were frequently used within reviews.

The first branch (called 'co-worker, people, colleague') contains words that describe who reviewers interact with at work. It includes many words used to describe colleagues (such as co-workers, staff, workmate) as well as references to positive relationships (such as bond, connection and laugh).

The second branch ('commission, increment, compensation') captures the terms and benefits of work. It includes nouns related to pay, holiday and opportunities for progression.

The third branch ('integrity, competence, capability') contains many of the positive dimensions of worker well-being. These range from reflections on the behaviours of others (including professionalism and transparency), the attributes of the broader organisation (such as their philosophy and vision) and the impact on workers (including a sense of pride and accomplishment).

The fourth and final branch ('threat, corruption and consequence') encapsulates negative experiences at work. It includes a number of negative emotions (anger, anxiety and uncertainty), as well as the events that may have triggered these emotions (shortages, cutbacks and racism).

While the broad groupings are sensible, they would require further refinement by domain experts before being adopted. This data-driven taxonomy could act as a starting point for an official taxonomy.

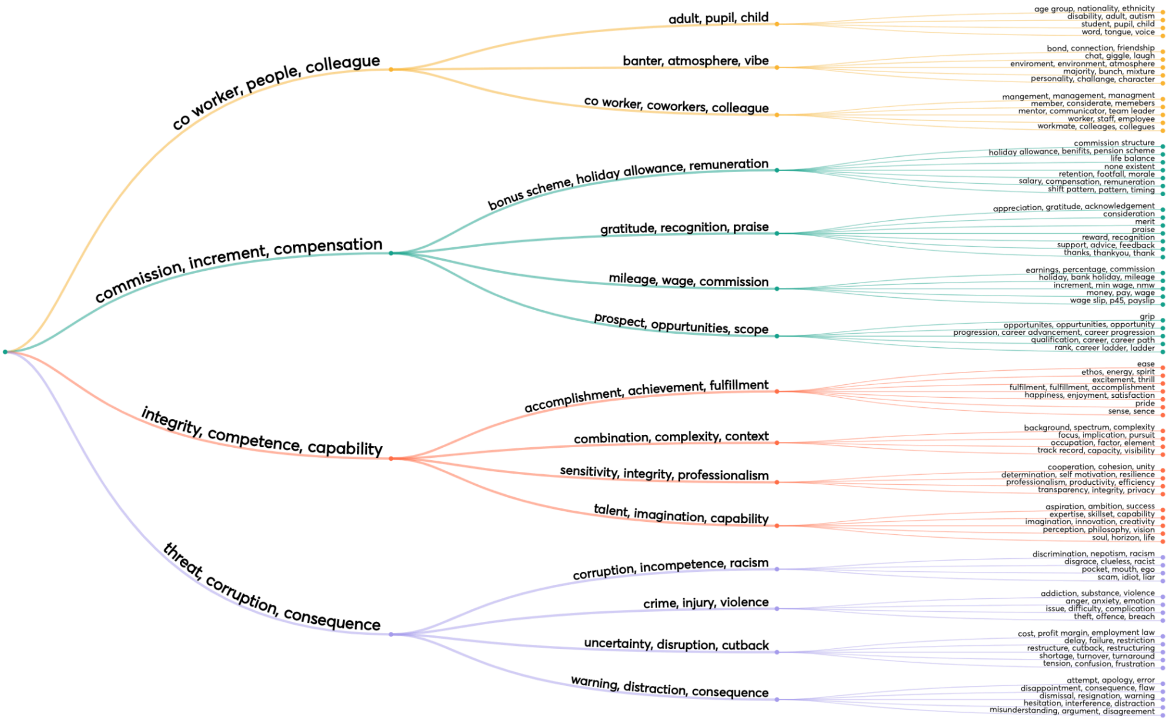

Figure 12: The Quality of Work taxonomy.

5.2.3 Comparison with the CIPD's dimensions of work quality

The Quality of Work taxonomy, shown above, does not map neatly to the seven dimensions of work quality that are used by the CIPD. However, many of the words within each dimension can be found within the taxonomy. For example, the second branch of the taxonomy (which is called 'commission, increment, compensation') aligns well to the CIPD's dimension of 'pay and benefits'. That second branch splits into four further branches and the lowest of these (which is called 'prospect, opportunities and scope') is similar to the CIPD's 'job design and nature of work' dimension - they both cover opportunities for progression. A final example is that the third branch (called 'integrity, competence and capability') broadly aligns with the CIPD's dimension of 'social support and cohesion'.

In addition to encompassing the seven dimensions, the Quality of Work taxonomy includes several other dimensions. Crucially, these dimensions are not captured by the 18 metrics of job quality that have been proposed by the Measuring Job Quality's Working Group and are described below:

Non-financial recognition: Within the branch of the taxonomy that captures work terms and benefits ('commission, increment, compensation'), there is an entire cluster of words that refer to non-financial rewards. This cluster is called 'gratitude, recognition, praise' and it includes terms such as 'thank you', 'appreciation' and 'acknowledgement'. The presence of these terms suggests that non-financial rewards may be an important dimension of job quality that is not currently captured within the 18 metrics.

Atmosphere, culture and environment: In the *Indeed* reviews, workers frequently speak about an organisation's culture, atmosphere and environment. The closest dimension within the CIPD's seven dimensions is 'Social Support and Cohesion'. The proposed metrics for this dimension concentrate on an individual's relationships with their colleagues and immediate boss. An additional metric could be added to get a sense of the individual's satisfaction with the broader atmosphere created by senior management.

Integrity, transparency and vision: Reviewers also frequently give their reactions to the way in which their organisation is run. This includes thoughts on the visibility and transparency of senior management, as well as the organisation's philosophy, ethos and vision. These do appear to impact workers' well-being, and so could be added as an extra dimension to CIPD's taxonomy.

Negative events: The reviews from workers include descriptions of many one-off events that were detrimental to worker well-being. The Working Group's set of proposed metrics do ask employees about physical injuries and periods of anxiety or depression. Ideally, the latter metric would be expanded to include *any* work-related episode of poor mental health. Further questions could be added to enquire about the possible causes of these events, with examples of discrimination, staff shortages, and the prospect of cutbacks.

5.3 Differences in the dimensions of work quality between managers and non-managers

This section explores differences in work-quality dimensions between those in managerial and non-managerial roles. To distinguish between reviewers in each role, the Word2Vec model was used to extract synonyms for 'manager', and these were narrowed down to a set of 19 words. Any reviewer whose job title included one of these words was judged to be in a managerial role. The words included the likes of 'leader' and 'executive', as well as acronyms such as 'tl' (for team leader).

Just under 15% of reviews were identified as belonging to managers. A simpler approach would have been to use job titles that had been assigned to 'SOC Code 1' which covers 'Managers, Directors and Senior Officials'. However this SOC code includes some non-managerial roles (such as Supplier) and omits some middle-management positions (such as a Sales Area Manager, which sits in SOC Code 3).

After separating the reviews of managers and non-managers, the frequency of word use by each group was calculated. Figures 13 and 14 show the words (within two branches of the taxonomy) that were significantly more likely to be mentioned by managers (on the right-hand side) and non-managers (on the left-hand side). The size of the circles reflect the combined mentions of the word by both groups. The analysis focuses on words from two (of the four) branches within the Quality of Work taxonomy: rewards (‘commission, increment, compensation’) and negative events (‘threat, corruption, consequence’). These two branches were found to contain the most interesting differences in word use between managers and non-managers. To test for significant differences in word use, a 2x2 contingency table was formed for each word. The table contained the total mentions of the word by all managers and by all non-managers, and the total mentions of all other words (in the quality of work taxonomy) by each of the two groups. A Chi-Squared test, using a 1% significance level, was then applied to test for significant differences in the relative use of the word between managers and non-managers.

The figures show that the two groups placed different emphasis on these two broad dimensions of work quality. In regards to the ‘terms and benefits’ branch (Figure 13), non-managers more frequently mentioned aspects relating to base pay and hours, while managers focussed relatively more on other benefits (e.g. bonus, benefit package), as well as career progression, opportunities for advancement, and work-life balance. Within the ‘negative events’ branch (in purple), non-managers spoke more frequently about interpersonal difficulties (including instances of discrimination, tension and favouritism), as well as the impact of these events on their health (anxiety, depression and stress). Managers spoke more about restructuring, shortages, downturns and complaints.

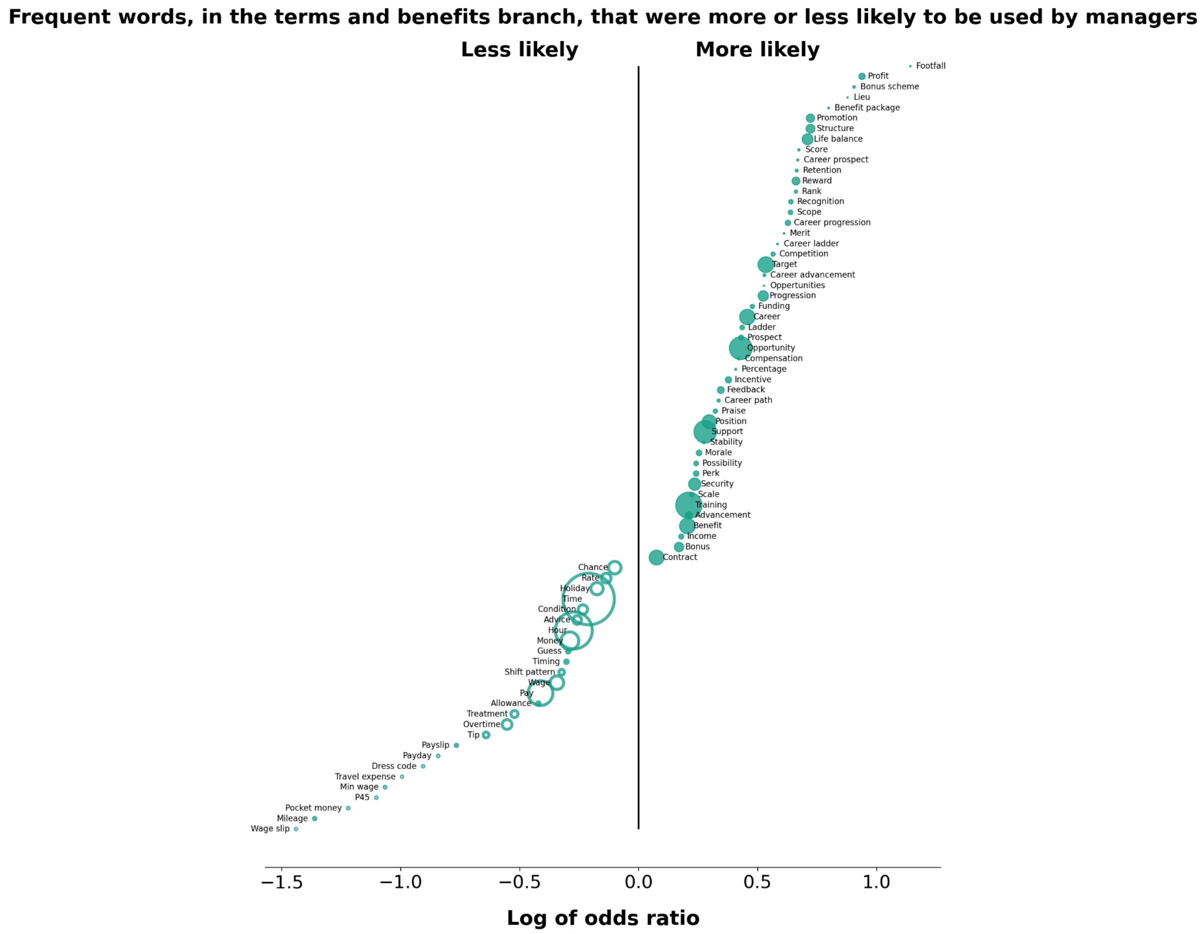

Figure 13: Words from the terms and benefits branch (in the Quality of Work taxonomy) that were significantly more likely to be used by managers (solid circles on the right-hand side) and non-managers (empty circles on left-hand side).

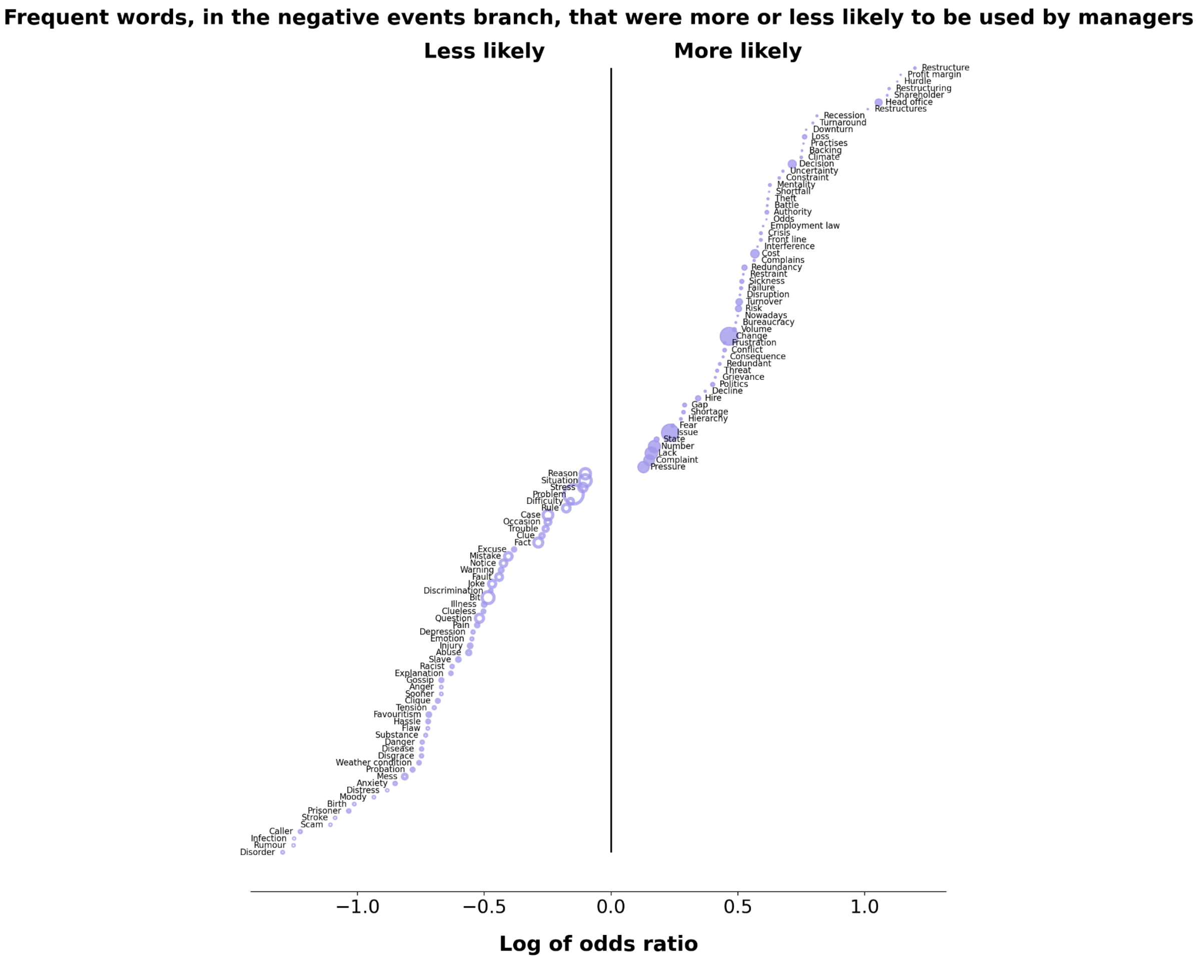

Figure 14: Words from the negative events branch (in the Quality of Work taxonomy) that were significantly more likely to be used by managers (solid circles on right-hand side) and non-managers (empty circles on left-hand side).

5.4 Changes in the dimensions of work quality over time

This section examines changes in the emphasis placed on the different aspects of work quality between 2012 to 2019.

5.4.1 The occupational mix over time

Figure 15 shows how the mix of reviewers' occupations changed over the sample period (based on their first-digit SOC codes). The mix remained more or less constant, which is critical to the interpretation of subsequent results. Specifically, it gives greater confidence that any changes in word-use reflect shifts in the dimensions of work quality, as opposed to changes in the mix of workers who left reviews. Of course, changes in the mix of reviewers cannot be completely ruled out as there may have been movements in the mix of jobs *within* each occupational group.

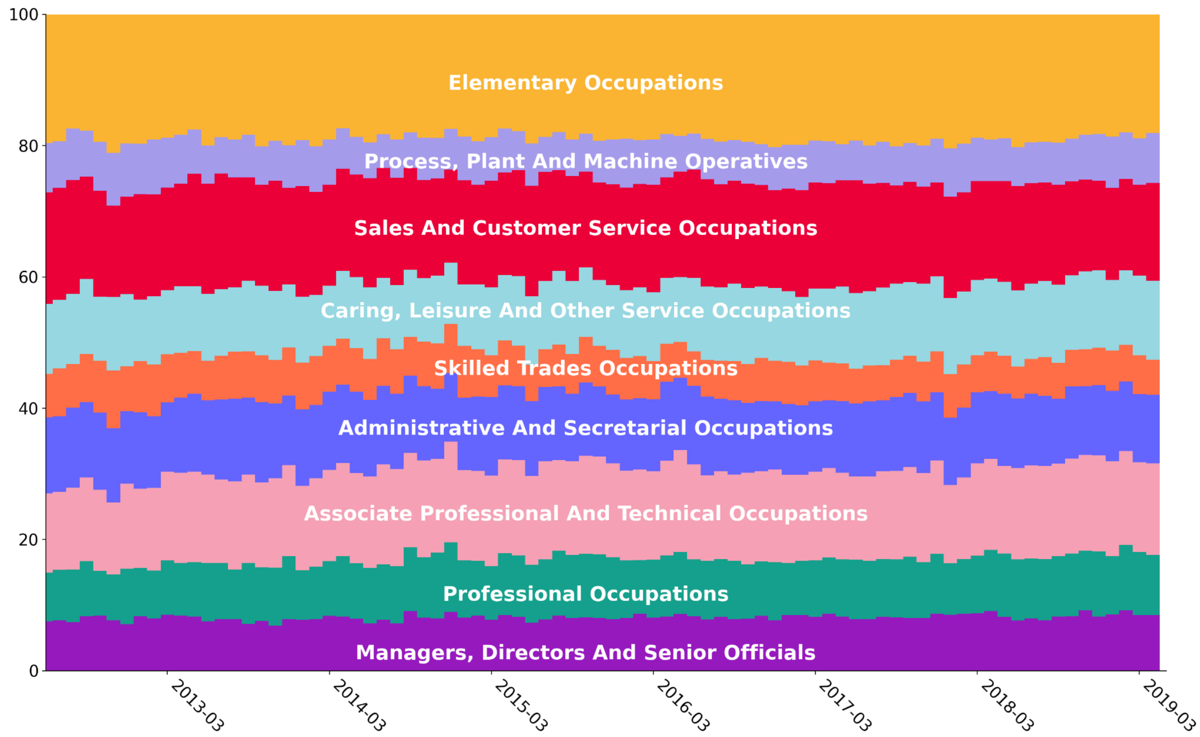

Figure 15: The breakdown of reviews by the broad occupational groups of reviewers (first-digit SOC codes).

5.4.2 Changes in word use over time

The evolution of word use over time was explored by splitting the 50% sample into two further sub-samples. The first sub-sample consists of reviews posted between 2012 and 2016 (inclusive). The second contains reviews posted between 2017 and 2019 (inclusive). Fewer reviews were posted in the early years, and so this split ensures that there are a roughly equal number of reviews in each of the two periods (360k and 380k respectively).

The analysis began by examining how the relative emphasis on the four different branches (in the first level) of the whole taxonomy had shifted between the two periods. This revealed that the share of mentions for words that belong to the Quality of Work branch grew slightly, by approximately 6 percentage points. In the first sub-sample, these words comprised 49% of all mentions of words within the whole taxonomy. In the second sub-sample this metric had risen to 55%.

The analysis below explores the changes within the Quality of Work taxonomy that were responsible for this growth. Figure 16 shows words within the Quality of Work taxonomy for which the total number of mentions changed significantly between the two periods (and for which the contribution to the overall growth exceeded 0.05 percentage points, in either direction). As before, significant changes (at the 1% level) were found by performing a Chi-squared test between the use of the word in the first and second sub-samples. The use of the word in a sub-sample was calculated by taking its mentions and scaling it by the total mentions of all words (from the Quality of Work taxonomy).

A number of changes are evident in the figure. First, there appears to have been a switch in the terminology that reviewers use to describe their co-workers. There were statistically significant declines in mentions of ‘co-worker’, ‘team’ and ‘colleague’. This is countered by significant rises in mentions of ‘staff’ and ‘employee’. For example, the term ‘co-worker’ was mentioned over 52.6k times in the first sub-sample, but only 3.5k times in the second sub-sample.

Second, there has been a shift away from speaking about the competences required in a job, with significant declines in mentions of ‘skill’, ‘ability’ and ‘knowledge’. Third, there appears to have been a rise in references to the terms and benefits associated with a job, with significant increases in mentions of ‘pay’, ‘salary’, ‘hour’ and ‘training’. Together, these two changes hint at a movement towards greater openness in the reviews, whereby reviewers give their thoughts on the job’s pay, rather than describing the skills required in the job.

Finally, reviewers also began to make more references to their wider workplace environment, using terms such as ‘atmosphere’, ‘culture’ and ‘environment’ more frequently. These shifts are explored in more detail below.

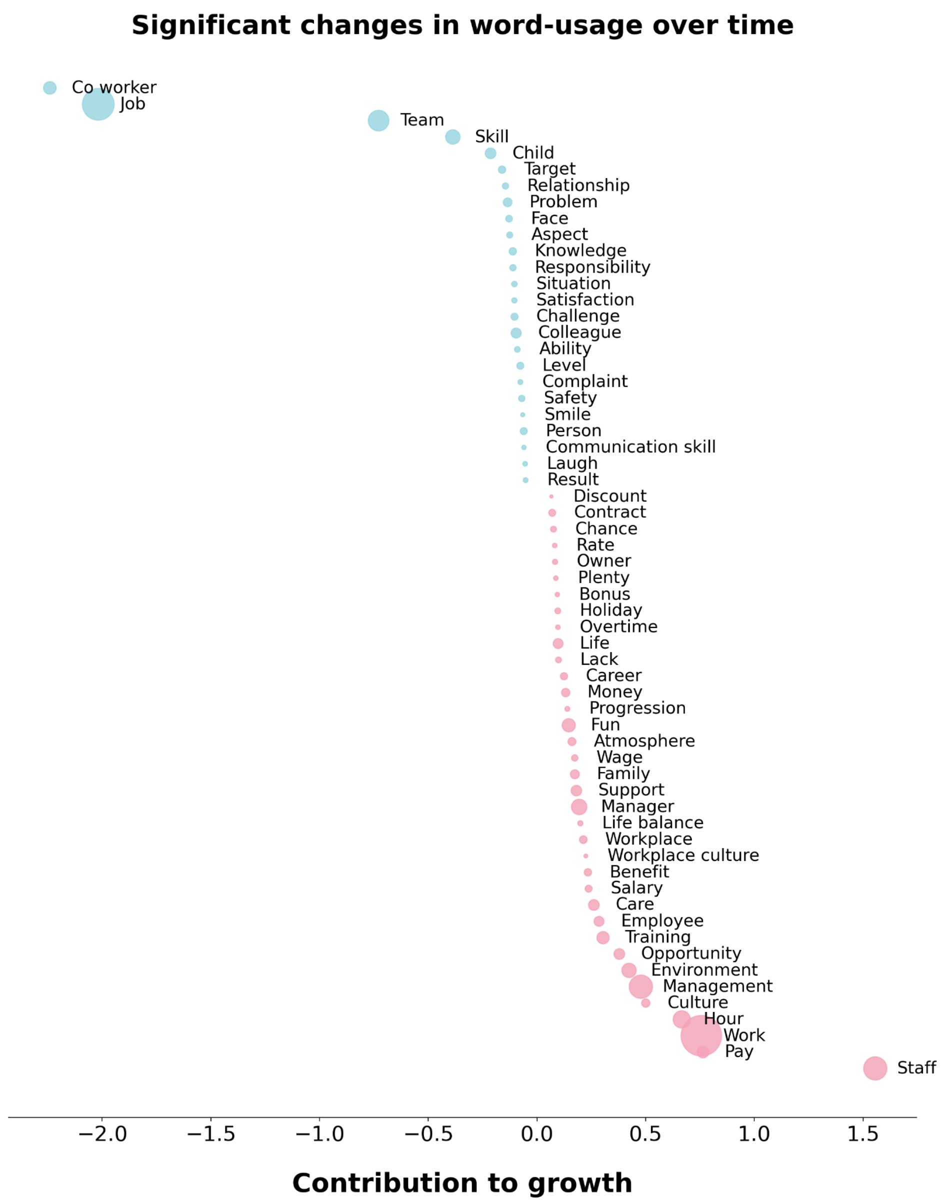

Figure 16: Words within the Quality of Work taxonomy whose number of mentions (as a proportion of all words within the taxonomy) changed significantly between the first and second halves of the sample. Only those words whose contribution exceeded 0.05 percentage points (either positively or negatively) are shown. The area of each circle reflects the total mentions of the corresponding word across both sub-samples.

There is a risk that the shifts described above may have been driven by changes in the prompts that are shown to reviewers. Amongst the current prompts, seen in Figure 3, half are non-specific (e.g. ‘the hardest part of the job’) but three refer to specific aspects of work (namely ‘learning’, ‘management’ and ‘workplace culture’) and these may have influenced the topics written about by reviewers. To complicate matters, one of these prompts has changed over time.

This concern was explored by calculating the frequency with which words were used on a month-by-month basis. A slow and steady change in the frequency of a word is unlikely to be the result of changes in the prompts by *Indeed*. If, instead, large jumps in word use (between two months) are observed, then a change in the prompts may have influenced the words used by reviewers.

Figure 17A shows, on a monthly basis, the percentage of reviews that contain the word at the top of the chart. Both 'culture' and 'workplace culture' exhibit sudden increases in 2017, which most likely reflects the introduction of the phrase 'workplace culture' into the prompts by *Indeed*. That said, the use of 'culture' was rising before the change, and there were steady gains in mentions of its synonyms - 'environment' and 'atmosphere'- over the sample period. Thus, despite the change in prompts, there is evidence that the workplace environment has become more important to workers over time.

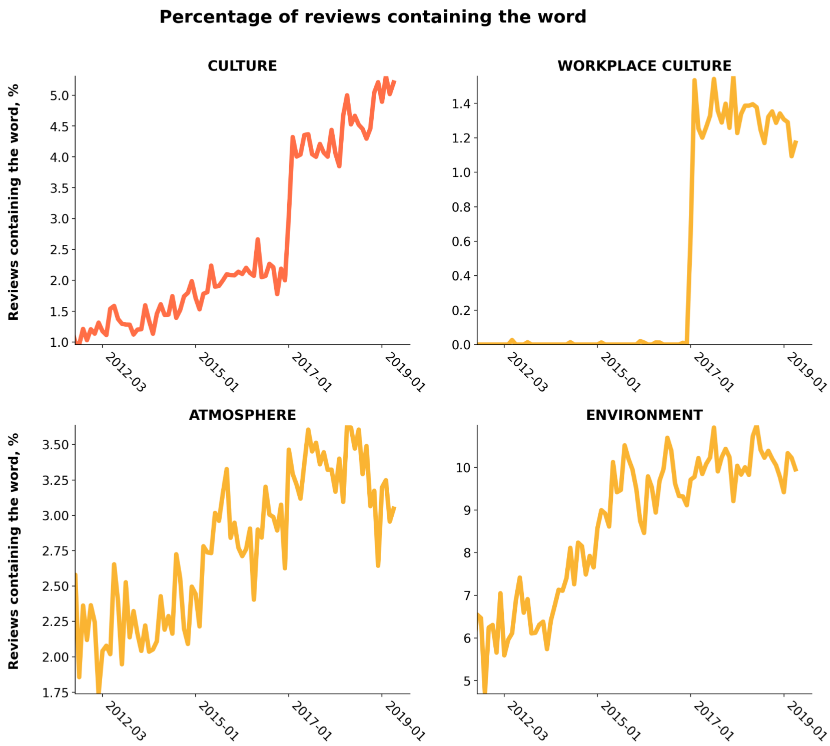

Figure 17A: Changes in word use over time. Caution: The scale on each vertical axis is unique. The colour of the line indicates the branch of the Quality of Work taxonomy to which the word belongs.

In regards to the increase in mentions around work terms and benefits, Figure 17B shows a steady rise over time. As mentioned, this result may stem from these components becoming more important to workers over the period. However, it is not possible to rule out the alternative hypothesis that workers simply became more willing or confident to talk about these potentially sensitive aspects over the same period.

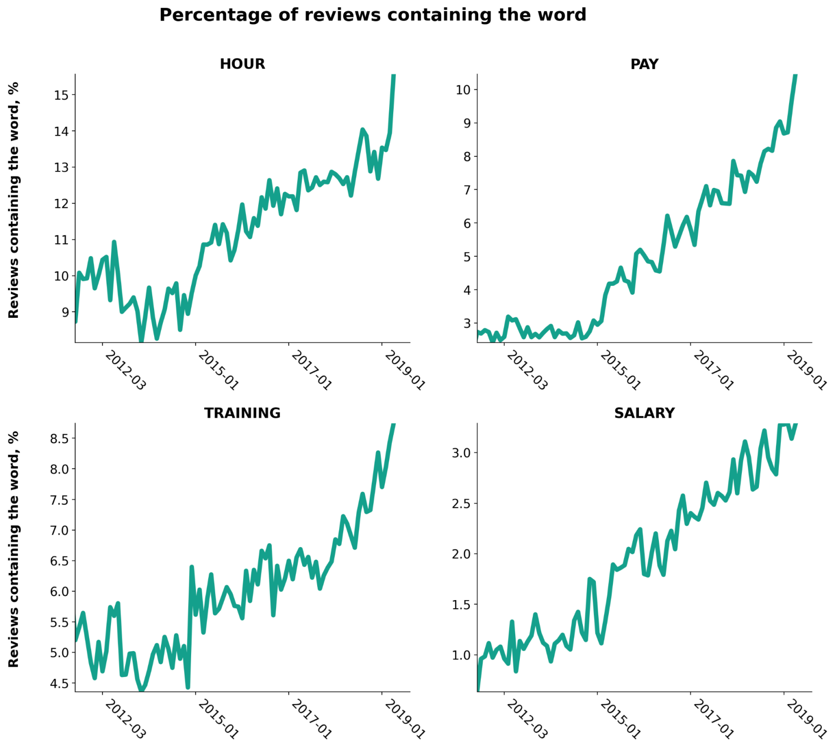

Figure 17B: Changes in word use over time. Caution: The scale on each vertical axis is unique. The colour of the line indicates the branch of the Quality of Work taxonomy to which the word belongs.

Regarding the change in the terms used to reference colleagues (observed above), this seems to be only partly driven by the change in one prompt. Figure 17C shows that use of the term 'co-worker' fell sharply at the start of 2017, and presumably reflects the removal of this word from the prompts. However, it was also declining before *Indeed* made this change. The term 'team' also declined steadily over the period, while 'staff' and 'employee' experienced gradual increases. Together, these movements suggest a genuine shift in the focus of reviewers, from immediate colleagues to a broader group of staff members.

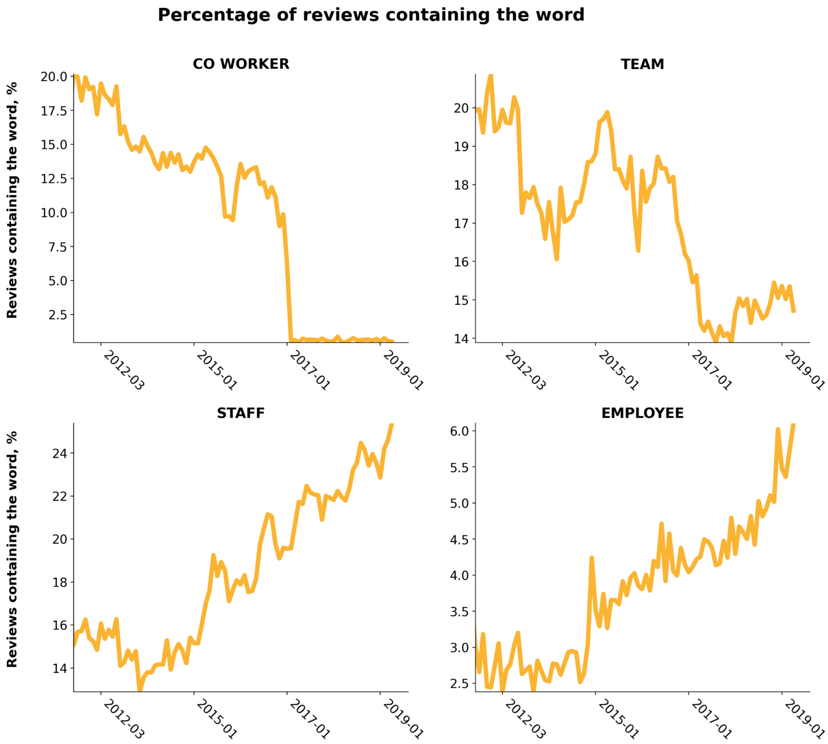

Figure 17C: Changes in word use over time. Caution: The scale on each vertical axis is unique. The colour of the line indicates the Quality of Work branch to which the word belongs.

Finally, in this section, attention is shifted to *all* words (within the whole taxonomy) that experienced significant changes in mentions between the two sub-samples. It is difficult to discern a theme amongst those words that became significantly *less* frequent between the two sub-samples. That said, there is a small group of terms that all relate to older technologies, such as ‘fax’, ‘newspaper’, and ‘telephone manner’. There is also, understandably, large declines in the use of ‘olympic’ which relates to the 2012 Olympics that was held in London.

Conversely, words that became much *more* frequent in reviews do share a common theme. They suggest much greater frankness in the reviews with significant increases in the use of terms such as ‘favouritism’, ‘nepotism’, ‘racism’ as well as the likes of ‘clown’, ‘donkey’ and ‘fool’. It is important to note that although these terms have become more frequent, their use overall remains relatively rare.

5.5 Differences in the dimensions of work quality between occupations

Both ‘pay’ and ‘atmosphere’ feature in the Quality of Work taxonomy. While the former is well defined, the latter is more ambiguous. This subsection explores how the importance of these two, quite different, components of work quality vary across occupations and job roles. Around 12% of all reviews mentioned either ‘atmosphere’ or ‘environment’ (with 95k total mentions), while 9% mentioned either ‘pay’, ‘salary’ or ‘wage’ (with 78k total mentions).

Figure 18 below shows the percentage of reviews in each broad occupation group (2-digit SOC code) that contained at least one term from each set of words.

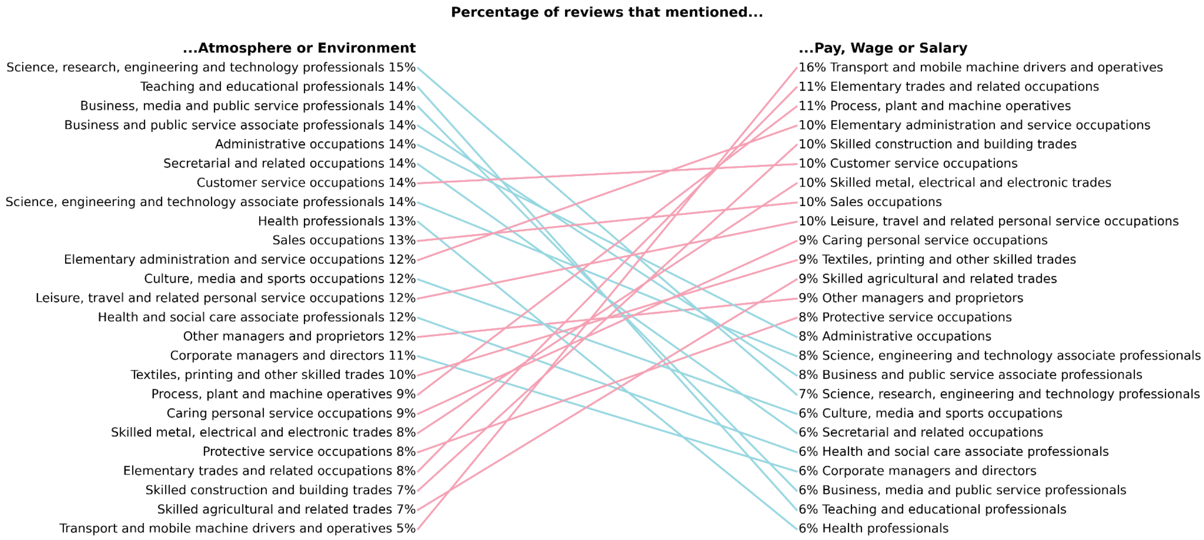

Figure 18: The percentage of reviews in each broad occupation group (2-digit SOC codes) that contained words relating to atmosphere (on the left-hand side) or pay (on the right-hand side). The ordering of the occupations reflects their ranking by these two metrics. These percentages are provided in brackets next to the name of each occupational group. The lines connect each occupational group. The colour of each line simply indicates its slope (positive or negative).

Figure 18 shows large variations in the frequency with which different occupational groups mentioned these two aspects of work quality. The percentage of reviews that referred to ‘atmosphere’ or ‘environment’ varied from 5% (for Transport and mobile machine drivers and operatives) to 15% (for Science, research, engineering and technology professionals). More broadly, workers in professional and administrative roles were most likely to mention these terms. In contrast, mentions of ‘pay’, ‘wage’, or ‘salary’ were more likely to be made by those in skilled trades and transport, with Transport and mobile machine drivers and operatives mentioning at least one of these terms in 16% of reviews. In contrast, Health professionals referred to these terms in just 6% of reviews. The figure also shows a clear negative correlation between the likelihood of mentioning either aspect of work quality.

To explore these findings further, focus is turned from (broad) occupational groups to (narrower) job titles. Attention is restricted to ‘common’ job titles, for which there were at least 100 reviews in the sample. The results are shown in Figure 19 below.

	…atmosphere or environment	**…pay, wage or salary**
Most likely to mention…	1. Research associate 2. Bar staff/waitress 3. Cafe worker 4. Bartender/waitress 5. Recruitment consultant 6. Office worker 7. Bar assistant 8. Marketing 9. Call centre advisor 10. Document controller	1. Domiciliary care worker 2. HGV class 2 driver 3. Class 1 driver 4. Former employee 5. Canvasser 6. Operative worker 7. Class 2 driver 8. HGV driver class 1 9. Line chef 10. HGV class 1 driver
Least likely to mention…	1. Driver/labourer 2. Ground worker 3. Paper round 4. Class 1 driver 5. HGV 2 driver 6. Courier driver 7. Lorry driver 8. Home care assistant 9. Maintenance 10. Domiciliary care worker	1. Nursery assistant (work experience) 2. Lunchtime supervisor 3. Animal care assistant 4. Childcare assistant 5. Teaching assistant (work experience) 6. Midday supervisor 7. Volunteer teaching assistant 8. Bar person 9. Classroom assistant 10. Operations director

Figure 19: The ten job titles of reviewers who were most (and least) likely to write reviews containing words relating to ‘atmosphere’ and ‘pay’. Only job titles for which there were at least 100 reviews were considered for inclusion.

The three roles that spoke most frequently about ‘atmosphere’ or ‘environment’ were Research associate, Bar staff/waitress and Cafe worker. For workers in these three roles, one or both of these words were mentioned in at least 23% of their reviews. On pay, the most frequent mentions came from drivers and security roles. HGV Class 2 Drivers mentioned the topic in almost 26% of their reviews. In contrast, those working in assistant-type roles (many with children) were least likely to speak about pay. Examples of these roles include childcare assistants, teaching assistants, nursery assistants, volunteer shop assistants, youth workers and research assistants. For each of these roles, pay, wage and salary were mentioned in less than 3% of reviews.

Example of a review by a childcare assistant: *"the best place to work. lovely staff and the most amazing work atmosphere. The children were always joyful and the other staff were so lovely and caring."*

Example of a review by an HGV driver: *"Not a bad company to work for not the greatest pay but easy work get left alone most of the time never forced to night out local family run company ..."*

Figure 20: Examples of reviews mentioning 'atmosphere' and 'pay'.

This subsection has shown that although 'pay' and 'atmosphere' are both dimensions of work quality, their relative importance varies across workers, and this variation correlates with occupation. This finding confirms the importance of using methods like stated and revealed preferences, within surveys of job quality, to capture the different emphases that employees may place on the various dimensions of work quality.

6 Conclusions

The UK has neither an official Quality of Work taxonomy nor a system for detecting changes in the salience of the dimensions that comprise job quality. This paper has shown how a novel dataset of employee reviews might help to both inform the construction of this taxonomy and aid the ongoing measurement of job quality.

The paper began by creating an algorithm that assigns job titles to occupations (using the SOC 2010 classification). This was necessary to measure the extent to which certain occupations were under- or over- represented in the reviews. A multi-level taxonomy of the frequent nouns mentioned within these reviews was created. A Word2Vec model captured the semantic relationships between terms and hierarchical clustering was used to group the nouns. One of the branches in the first level of the whole taxonomy was identified as containing the dimensions of job quality. This Quality of Work taxonomy was compared to the CIPD's own taxonomy. Differences between managers and non-managers, as well as changes over time, were examined. The paper concluded by contrasting the emphasis placed on pay and atmosphere by different occupations and job roles.

The paper contains four results. First, it identified aspects of work quality that are not currently within the CIPD's seven dimensions of work quality or the 18 metrics proposed by the Measuring Job Quality Working Group. These 'extra dimensions' include non-financial rewards, atmosphere, culture and environment, integrity, transparency and vision, and finally discrete negative events. Second, the paper showed that managers and non-managers emphasised different dimensions of work quality in their reviews. Managers placed greater emphasis on benefits and career progression, while non-managers focussed more on base pay and hours. In regards to the cluster of words that captured negative experiences, non-managers spoke more frequently about interpersonal difficulties, while managers referenced concerns around restructuring, shortages, downturns and complaints. The third finding is that the dimensions of work quality may be evolving overtime. There is evidence that the broader work environment (or culture) became more important to workers over the sample period. This finding is robust to a change in the set of prompts shown by *Indeed* to those leaving reviews. In addition, there were steady rises in references to work terms and benefits, including pay and hours. These components may have become more important or workers may have become more confident to speak about these aspects. Finally, the results revealed large differences across occupations in the emphasis placed on different dimensions of work quality. Workers in professional and administrative roles were most likely to mention ‘atmosphere’ or ‘environment’. In contrast, mentions of ‘pay’, ‘wage’, or ‘salary’ were more likely to be made by those in skilled trades and transport. There was also a clear negative correlation between the likelihood of mentioning these two aspects of work quality.

There are three directions in which the methods and results presented in this paper could be extended. First, the Quality of Work taxonomy could be used to inform the creation of an official Quality of Work taxonomy for the UK. The dimensions identified here would need to be confirmed by other approaches, such as carrying out a survey or gathering the views of experts. Particular attention should be paid to the views of workers in occupations that are not well represented in the reviews: Health Professionals and several types of skilled trades. Second, this dataset could act as an early warning system to quickly detect changes in the dimensions of work quality. Any newly identified dimensions could then be incorporated within official surveys to confirm the indications from this data source. A further study could examine how mentions of particular job-quality-dimensions vary with the economic cycle. For example, do reviewers begin to make more reference to ‘pay’ as wages stall? Third, there are two research projects that naturally follow on from this paper. The first would use *Indeed* reviews to explore the impact of COVID-19 on the dimensions of work quality. COVID-19 has caused long-lasting shifts in working behaviours, such as a dramatic rise in remote working. The second study would use a more up-to-date (or ‘live’) version of the dataset to track changes in job quality caused by the Cost of Living crisis. It would be particularly interesting to examine whether there has been an increase in references to ‘pay’ and ‘unions’. In conclusion, this novel dataset is a valuable source of information for discovering and exploring the dimensions of job quality.

References

Arends, I., C. Prinz and F. Abma (2017) "Job quality, health and at-work productivity", OECD Social, Employment and Migration Working Papers, No. 195, OECD Publishing, Paris. Link.

Baeck, A. and Uzzell, K. (2022) "Job quality in the UK – analysis of job quality indicators: 2021". ONS. Link.

Bosworth, D. and Warhurst, C. (2021) "Does good work have a positive effect on productivity? Developing the evidence base", Warwick Institute for Employment Research. Link.

Dabirian, A., J. Kietzmann and H. Diba (2016). "A great place to work!? Understanding crowdsourced employer branding", Business Horizons 10. Link.

Felstead, A., D. Gallie, F. Green and G. Henseke (2019). "Conceiving, designing and trailing a short-form measure of job quality: a proof-of-concept study", Industrial Relations Journal 50 (1). Link.

HM Government (2017) "Industrial Strategy: Building a Britain fit for a future", Industrial Strategy White Paper. Link.

HM Government (2018) "Good Work Plan", Industrial Strategy. Link.

HM Government (2021) "Build Back Better: our plan for growth". Link.

Howe, W. and Bisel, R. (2020) "Third-party online organizational reviews: Explaining review-and-rating patterns of the United States Military and large corporate organizations", Computers in Human Behaviors Reports 1 (1), 1-8. Link.

Irvine, G., D. White and M. Diffley (2018) "Measuring Good Work - The final report of the Measuring Job Quality Working Group" The Carnegie UK Trust and The RSA Future Work Centre. Link.

Jung, Y. and Yongmoo, S. (2019) "Mining the voice of employees: A text mining approach to identifying and analysing job satisfaction factors from online employee reviews", Decision Support Systems 123. Link.

Karasek, R. (1979) “Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign”, Administrative Science Quarterly, 24(2), 285–308. Link.

Norris-Green, M. and Wheatley, D. (2022) “CIPD Good Work Index 2022”, London: Chartered Institute of Personnel and Development. Link.

Poulin, V. and F. Theberge (2018) “Ensemble Clustering for Graphs”, International Conference on Complex Networks and their Applications, Springer, Cham. Link.

Ramos, R, S. Moro and P. Rita (2020) “What drives job satisfaction in IT companies?”, International Journal of Productivity and Performance Management 70 (2). Link.

Sainju, B. (2020) “Job Satisfaction and Employee Turnover Determinants in Fortune 50 Companies: Insights from Employee Reviews from Indeed.com”, Utah State University, All Graduate Theses and Dissertations, 7985. Link.

Sarkar, S. and J. Gifford (2018) “UK Working Lives Appendix 2: Methodology”, London: Chartered Institute of Personnel and Development. Link.

Symitsi, E., P. Stamolampros, G. Daskalakis and N. Korfiatis (2020) “The Informational Value of Employee Online Reviews”, European Journal of Operational Research (forthcoming), Link.

Taylor, M. (2017) “Good Work: The Taylor Review of Modern Working Practices”. Independent report. Link.

Warhurst, C., S. Wright and C. Lyonette (2017) “Understanding and measuring job quality Research report Part 1 – Thematic Literature Review”, Warwick Institute for Employment Research. Link.