Rethinking the measurement of occupational task content

Matthias Haslberger [iD]
University of Oxford, UK

Abstract

Which tasks workers perform in their jobs is critical for how technological change plays out in the labour market. This article critically reviews existing measures of occupational task content and makes the case for rethinking how this concept is operationalised. It identifies serious shortcomings relating to the theoretical content and the empirical implementation of existing measures. Based on survey data from European Union countries between 2000 and 2015, it then introduces novel measures of routine task intensity and task complexity at the International Standard Classification of Occupations 1988 two-digit level that address these shortcomings. The indices will contribute to a more theoretically informed understanding of technological change and benefit both labour economists and sociologists in investigating the nature of recent technological change.

JEL Codes: J23, J24

Introduction

Ever since the seminal article by Autor et al. (2003), routine-biased technological change (RBTC) has been the dominant explanation for employment and wage trends in the technological change literature. These authors developed a set of measures of occupational task content which have since, by virtue of academic primogeniture, enjoyed a near monopoly in the task literature. Only recently, scholars have begun subjecting them to

Corresponding author:
Matthias Haslberger, Nuffield College and Department of Social Policy and Intervention, University of Oxford, New Road, Oxford OX1 1NF, UK.
Email: matthias.haslberger@nuffield.ox.ac.uk

greater scrutiny. Invariably, they find conceptual and empirical problems with the standard indices (Fernández-Macías and Hurley, 2017; Handel, 2017; Sebastian and Biagi, 2018).

This article adds to this recent literature, arguing that the congruence of theoretical concepts and their empirical implementation needs to be improved and that suitable indices need to account for differences over time and between countries. The article then develops such improved measures using workers' self-assessments in the European Working Conditions Survey (EWCS). One illustrative improvement is the case of office clerks who in much of the existing literature are characterised as the most routine-intensive occupation. This assessment disregards the evolution of this occupational group since the 1970s as it has computerised, and clerks have taken over formerly managerial tasks. Using recent data and a more suitable set of variables, the approach proposed here places office clerks near the middle of the routine distribution, which is more in line with today's workplace requirements. The present article is therefore predominantly a methodological contribution which aims to enable a more theoretically informed understanding of technological change.

The article first provides an overview and critique of the existing operationalisations in the 'Overview and critique of existing operationalisations' section, followed by the theoretical case for an alternative approach. In 'Data and construction of the indices' I describe the data and the strategy for constructing the new measures of occupational task content. The 'Comparing measures of task content' section is devoted to quantifying the differences between the existing and new measures. The 'New opportunities for research' section proposes fruitful applications for the new indices, and the 'Conclusion' summarises the findings.

Overview and critique of existing operationalisations

What are routine tasks?

The RBTC approach explicitly asks 'what it is that people do with computers' (Autor et al., 2003: 1280) and thus takes a first step away from the black-box view of technological change that has been criticised by sociologists (Fernandez, 2001). In another seminal article in the task literature, a task is defined as a 'unit of work activity that produces output' (Acemoglu and Autor, 2011: 1118). Essentially, the RBTC argument predicts a reallocation of employment based on the task composition of occupations. The core of Autor et al.'s (2003) argument is worth quoting in full and stipulates

> (1) that computer capital substitutes for workers in carrying out a limited and well-defined set of cognitive and manual activities, those that can be accomplished by following explicit rules (what we term 'routine tasks'); and (2) that computer capital complements workers in carrying out problem-solving and complex communication activities ('nonroutine' tasks). (p. 1280)

Thus, production tasks are allocated to workers or capital based on comparative advantage in performing the respective tasks, where capital has an advantage in performing routine tasks and workers have an advantage when it comes to non-routine tasks. Elsewhere in the article, the authors argue that routine tasks 'require [the] methodical

repetition of an unwavering procedure' (Autor et al., 2003: 1283), thus introducing the element of repetitiveness. Other definitions require that routine tasks be 'expressible in rules such that they are easily programmable and can be performed by computers at economically feasible costs' (Spitz-Oener, 2006: 239) or define routine-intensity as 'the extent to which an occupation is automatable or codifiable' (Caines et al., 2017: 302).[1] Hence, conceptually, the labour economics literature focuses on codifiability and repetitiveness as the distinguishing features of routine tasks.

Existing operationalisations of occupational task content

This article engages with two of the most influential approaches to quantifying occupational task content.[2] The first, adopted by most labour economists, follows the pioneers of the task-based approach. Autor et al. (2003) identified five task dimensions in their empirical analysis which Autor and Dorn (2013) have consolidated and formalised into a framework which classifies tasks as routine, abstract or manual, and has become mainstream in the economics literature and beyond (see Sebastian and Biagi, 2018, for an overview). However, this dominance is not the result of rigorous debate on how to best measure occupational task content but mainly of convenience turned convention.

The second approach, represented by recent studies in economic sociology, retains the basic logic of the task framework. However, its proponents choose different task dimensions, variables, data sources, and units of analysis to operationalise task content (Fernández-Macías and Hurley, 2017). Further key differences from the Autor and Dorn (2013) approach are the use of workers' self-assessments, similar to Spitz-Oener (2006) and Autor and Handel (2013), and the level of analysis which is at the 'job' level (two-digit occupations in two-digit sectors). This article owes much to the work of Fernández-Macías and Hurley (2017) but goes beyond their important contribution.

Table C1 in Appendix C of the Supplemental File to this article[3] contrasts both operationalisations of the RBTC theory. The overview illustrates the proliferation of empirical studies based on the approach developed by Autor et al. (2003) and Autor and Dorn (2013). Their methodology has been adapted to a range of contexts beyond the United States (US), including studies of individual European countries and comparative studies. Some authors, like Spitz-Oener (2006), Acemoglu and Autor (2011) and Autor and Handel (2013), only adopt the analytical framework but use different data, while others, such as Goos et al. (2014), embrace the approach wholesale. The alternative approach of Fernández-Macías and Hurley (2017) has also been used in two Eurofound (2014, 2017) reports in which the same authors were involved. While other measures of occupational routine-intensity do exist (e.g. Salvatori, 2018), they do not generally make the same claim to generality as the Autor and Dorn (2013) and Fernández-Macías and Hurley (2017) measures. Therefore, my discussion focuses on these two prominent measures.

A critique of existing operationalisations of occupational task content

The thesis of this article is that the measures just described suffer from conceptual and empirical problems which can, however, be addressed. Conceptually, there are two major issues:

Hypothesis 1.1: Existing approaches use ill-defined auxiliary task dimensions (secondary task dimensions other than routine-intensity).

Hypothesis 1.2: The variables used in existing approaches do not capture the concepts they are purportedly measuring.

Empirically, I identify three main shortcomings:

Hypothesis 2.1: Existing approaches do not account for task variation within occupations by using expert-coded data.

Hypothesis 2.2: Existing approaches fail to account for change over time within occupations.

Hypothesis 2.3: Existing approaches do not account for differences between countries.

This section discusses these problems and how I propose to address them. I then introduce the new measures, analyse how they compare to existing indices and illustrate the arguments that I have discussed theoretically.

Ill-defined auxiliary task dimensions. The uniting feature of all approaches discussed above is their interest in occupational routine-intensity. However, all studies define one or more auxiliary axes of occupational tasks, for example, in Autor et al. (2003) that is the cognitive-manual axis and in Fernández-Macías and Hurley (2017) simply a cognitive task dimension. Caines et al. (2017) juxtapose routine and complex tasks. Usually, however, these auxiliary axes do not serve a well-defined purpose. Only Caines et al. (2017) formulate a theory of task complexity in relation to routine-intensity and technological change. In most other analyses, the auxiliary task dimension does not add much of substantive interest. For instance, Autor et al. (2003) do not posit any independent relationship between technological change and cognitive and manual task inputs. However, if the measure of routine-intensity is used to operationalise RBTC, whatever auxiliary measure is part of the analysis ideally should have some independent theoretical interpretation. In particular, since RBTC is an alternative to SBTC (skill-biased technological change), it would be eminently helpful to have a measure for the latter that is constructed in a similar manner as the measure of RBTC.

Variables that do not capture key concepts. The variables used to operationalise cognitive and manual routine tasks in Autor et al. (2003) and Autor and Dorn (2013) completely fail to capture key aspects of the notion of routine as defined above, most importantly, repetitiveness. For example, they measure cognitive routine-intensity with 'adaptability to situations requiring the precise attainment of set limits, tolerances and standards', a criterion which appears geared towards low-level clerical jobs and jobs in the manufacturing sector (Autor et al., 2003: 1323). Manual routine-intensity is measured with finger dexterity. The relationship between finger dexterity and codifiability and repetitiveness seems altogether questionable. It rests on the assumption that tasks which involve fine

movements and coordination are repetitive and can be automated – a shaky assumption, for example, with regard to musicians and artisans who often require a great deal of finger dexterity. In Autor and Dorn (2013), the measure for routine tasks is the simple average of those two variables. Thus, even though repetitiveness is at the core of the concept, it barely features in the variables used to measure routine-intensity. Overall, the variables used appear to have been chosen not so much with the abstract concept of routine in mind but rather with a preconceived set of purportedly routine occupations.

Like this study, Fernández-Macías and Hurley (2017) motivate their article with a desire to improve the match between concepts and operationalisations of task content. Their index comprises five items from the EWCS. Three questions on repetitive arm or hand movements and short repetitive tasks capture the repetitiveness dimension of routine, with the first identifying manual routine tasks and the second and third capturing repetitive tasks more broadly. A fourth question about monotonous tasks introduces the notion of 'boringness'. They also include a question on dealing with unforeseen problems which arguably is an inverted measure of codifiability – something that is unforeseen cannot be codified. However, above all, dealing with unforeseen problems requires creativity and problem-solving ability, two of the qualities measured by the complexity index. The Fernández-Macías and Hurley (2017) cognitive index focuses perhaps too much on computer usage with two out of four variables and abandons the task-based framework by including a worker characteristic (average education). Thus, while their measures undoubtedly get much closer to the core of the respective concepts, further improvements are necessary.

Expert-coded occupation-level data. The Autor and Dorn (2013) task measures are derived from the Dictionary of Occupational Titles (DOT), in which expert coders assign scores that characterise occupations in the United States. More recent studies often use the Occupational Information Network (O*NET) database which offers a wider range of indicators than the DOT but is also based on data from American workers and only provides aggregated data at the level of occupations (Caines et al., 2017). Survey data, by asking people what they actually do in their job, are conceptually closer to the idea of the task-based approach; furthermore, survey data can provide a sense of the variability of tasks within an occupation. Spitz-Oener (2006) moreover points out that experts tend to underestimate the true changes in task content. A drawback of the survey approach may be measurement error introduced by respondents understanding questions differently or having different reference points. Nevertheless, the best way to find out what people do at work seems to be to ask the workers themselves and to take the variability of their answers into account.

Failure to account for change within occupations. The data used by Autor and Dorn (2013), Fernández-Macías and Hurley (2017) and most subsequent studies make it impossible to account for within-occupation change over time.[4] However, changing job tasks are a crucial component of technological change, as numerous studies make clear. A case study by Fernandez (2001) details how job tasks changed at a plant that underwent modernisation. Spitz-Oener (2006) finds that in Germany, within-occupation changes account for most of the change in aggregate task requirements. It is therefore clear that as the

prevalence of occupations changes, so does their nature. A failure to account for this would result in underestimating the impact of technological change on the labour market.

Failure to account for differences between countries. Furthermore, all existing comparative studies fail to account for potential differences in task content between countries. Although Eurofound (2014) rightly point out that job tasks should be relatively similar across developed countries, the possibility that some occupations differ between countries should not be dismissed. For example, in lagging developed economies, limited access to computer capital may retard computerisation in routine cognitive occupations. Indeed, Eurofound (2014) find that there are differences across countries, albeit small, regarding the demand for routine and cognitive tasks in a job, relative to that country's task distribution. Yet, many studies use the measures of Autor et al. (2003) and Autor and Dorn (2013) outside of the US. Fernández-Macías and Hurley (2017), while they do not use data from a country that is not part of the analysis, still only calculate one measure of task intensity for all countries. Yet ideally, country-specific measures of task content should be used for more detailed analyses.

Towards better measures of task content

Meaningful auxiliary task dimension. Much of the RBTC literature lacks a well-defined auxiliary dimension to the routine dimension. This article proposes a measure with the aim of enabling an analysis of SBTC alongside RBTC. Following Caines et al. (2017), this is called the task complexity dimension and is defined as the demand for higher order skills such as effective communication, abstraction and decision making. Occupations that comprise many tasks requiring these skills are less likely to be replaced by technology, as machines and artificial intelligence (AI) cannot (yet) perform such tasks. On the contrary, these higher order skills are inclined to be complemented by modern technology: thanks to it, effective communicators can reach wider audiences, scientists have powerful tools at hand that facilitate abstraction and induction and so on. Hence, task complexity is suitable for measuring the prevalence of SBTC.

Routine task intensity (RTI) and task complexity are not just two sides of the same coin: some routine occupations also require performing a considerable number of complex tasks, for example, some health and clerical occupations. Thus, task complexity is expected to be negatively correlated with routine-intensity but is nevertheless analytically distinct.

Variables that capture key concepts. The most common operationalisation of RTI does not in fact measure RTI as it is defined in the same literature: the key concepts of repetitiveness and codifiability are insufficiently captured. Thus, there is a real need to better align the concept of routine-intensity and its measurement. Fernández-Macías and Hurley (2017) have taken an important step in this direction and this discussion of the issue relies heavily on their previous work, yet their measure of routine-intensity can be improved further.

My routine index includes the following items: whether a job involves (1) repetitive arm or hand movements, (2a) short repetitive tasks of less than 1 minute, (2b) short repetitive tasks of less than 10 minutes, (3) monotonous tasks and (4) meeting precise quality standards.[5] These five items are less occupation-specific than the ones used by Autor et al. (2003). At the same time, these items afford the notions of repetitiveness and codifiability their due importance. So, while this index departs almost completely from Autor et al. (2003), it closely resembles the RTI measure of Fernández-Macías and Hurley (2017).

The only difference to Fernández-Macías and Hurley (2017) is the inclusion of the item 'meeting precise quality standards' which is included instead of 'solving unforeseen problems on one's own'. In the appendix (see Supplemental File), I show that the latter is better suited as a component of the complexity index. Fernández-Macías and Hurley (2017) and Eurofound (2014: 48) argue explicitly against the inclusion of a quality control variable, on the grounds that this assigns relatively highroutine scores to higher skilled occupations which often include monitoring tasks. They are correct; however, there is a crucial difference between enforcing quality standards and being forced to meet them. Therefore, the requirement to meet precise quality standards is a suitable indicator of codifiability in an index which might otherwise focus too heavily on the repetitiveness aspect of routine-intensity.

The task complexity index has no direct counterpart in Autor et al. (2003) and is also where I depart further from Fernández-Macías and Hurley (2017) who focus on cognitive intensity rather than task complexity as the second dimension of occupational task content. It aims to measure the demand for higher order skills such as effective communication, abstraction and decision making. It includes the following items from the EWCS: whether a job entails (1) working with computers, tablets, smartphones and so on; (2) solving unforeseen problems on one's own; (3) complex tasks; and (4) learning new things.[6]

The only overlapping question with Fernández-Macías and Hurley (2017) is whether a job involves complex tasks. The included questions reflect the fact that on-the-job learning is a key characteristic of complex jobs, as is solving unforeseen problems on one's own. Moreover, rather than two separate questions whether a job involves the use of computers and use of the Internet, one question asking for the use of 'computers, tablets, smartphones, etc.' is included. This serves to avoid an undue emphasis on office jobs since it is unlikely that a job involves the use of computers but not the Internet, or vice versa. Overall, using variables that truly capture the prevalence of routine and complex tasks in line with their definitions is an important and long overdue advance in the literature on RBTC. It ensures that the indices measure what they claim to be measuring (criterion validity). The present indices are an important step in this direction.

Individual-level survey data. Using the EWCS helps to realise the advantages of individual-level survey data. Instead of one number assigned by an outsider, each score is the result of many (often thousands of) practitioners of an occupation evaluating what they do in their job and how they do it. It is worth noting that with O*NET replacing the DOT, there is a trend towards using survey data even in the US context. However, O*NET still only provides aggregated data at the occupational level. In contrast, a crucial benefit of using

survey data is that it provides an indication of the degree of variation within a job (Autor and Handel, 2013). Overall, the present approach follows the practice of a few scholars, mainly from Europe, who have used survey data for the analysis of occupational task content all along.

Accounting for change within occupations. With the EWCS, it is possible to analyse change within occupations in most European countries. In principle, any data source with consistent occupation-level data for several points in time can be used for this purpose. However, the DOT was only updated infrequently and was eventually replaced by O*NET. It is therefore impossible to develop a time series of occupational task content based on the variables in Autor et al. (2003). Overall, for comparative analysts, the EWCS is the most appropriate available data source. Eurofound (2014) and Fernández-Macías and Hurley (2017) only use the 2010 wave and ignore the temporal component of occupational change. In contrast, I use data from the four most recent waves (2000–2015) to investigate change within occupations over a period of 15 years. Furthermore, the dataset can easily be expanded to include future waves of the EWCS once they become available.

Accounting for differences between countries. The EWCS data can also be used to analyse differences between countries. The EWCS covered 35 European countries in its most recent wave, with a target sample size between 1000 and 4000 individuals depending on country size.[7] With these characteristics, a country-level analysis at the two-digit level for occupations is feasible. The entirely European nature of the sample restricts the scope for such comparative analyses to exclude countries such as the US, but compared to the prevailing practice of using task content measures based on US data for all countries, this broader approach promises more robust insights.

Table 1 provides an overview of how the measures differ. Harking back to the five thesis statements above, it is clear that the proposal in this article entails important conceptual and empirical improvements over both Autor et al. (2003) and Fernández-Macías and Hurley (2017).

Data and construction of the indices

The main data sources for this article are waves 3 to 6 of the EWCS, which provides workplace information for the years 2000, 2005, 2010 and 2015. In the following analyses, I work with samples comprising the EU-27 and EU-15 countries (see Table A1, Appendix A, Supplemental File). Total N amounts to 107,488 individuals in the EU-27 sample and 74,895 in the EU-15 sample. The EWCS allows me to characterise the task profile of occupations with a set of relatively objective questions (e.g. 'Does your main paid job involve repetitive hand or arm movements?') and a number of more subjective items (e.g. 'Does your main paid job involve complex tasks?'). The full wording of the questions as well as summary statistics and details on the construction of the indices can be found in Appendix B of the Supplemental File, showing that the measures are internally consistent (construct validity).

Table 1. Summary of indices with their differences and improvements.

Characteristics	Autor et al. (2003)/Autor and Dorn (2013)	Eurofound/Fernández-Macías and Hurley (2017)	This proposal
Task dimensions	Routine (measure of predominance over abstract and manual tasks)	Routine, cognitive → measures of prevalence	Routine, complex → dimensions capturing RBTC and SBTC
Variables	Severe mismatch between concepts and variables	Remaining problems: mis-classification of 'dealing with unforeseen problems'; mixing of worker and task characteristics	Coherent operationalisation of routine and complex dimensions
Data and unit of analysis	1977 DOT (O*NET in some later applications); US census occupations	2010 EWCS → contemporary individual-level survey data; ISCO-88 two-digit, NACE two-digit 'jobs'	2000, 2005, 2010, 2015 EWCS → contemporary individual-level survey data, repeated cross-section; ISCO-88 two-digit occupations
Treatment of change over time	Limited analysis in Autor et al. (2003), none in other articles	No analysis of change over time	Detailed analysis of change over time
Treatment of country differences	Index specific to the US, no analysis of country differences	No analysis of country differences	Analysis of country differences in EU-15

RBTC: routine-biased technological change; SBTC: skill-biased technological change; DOT: Dictionary of Occupational Titles; O*NET: Occupational Information Network; EWCS: European Working Conditions Survey; ISCO-88: International Standard Classification of Occupations 1988; NACE: Statistical Classification of Economic Activities in the European Community (Nomenclature statistique des activités économiques dans la Communauté européenne).

I calculate three versions of the indices. These overall indices use all available data and are pooled across countries and waves. They are available for the EU-27 and EU-15 groups of countries. In addition, I calculate a wave-specific version of the overall indices and a country-specific version with data for the respective country pooled over all available waves. Following the approach in most of the literature, the indices are calculated at the two-digit level for occupations and not for occupation-industry cells. This entails a certain loss of precision compared to Fernández-Macías and Hurley (2017) but in my view is warranted for two reasons. First, only by restricting the analysis to the 26, two-digit occupations, categorised by the International Standard Classification of Occupations 1988 (ISCO-88), is the analysis of country- and wave-specific scores feasible. Second, the method of Fernández-Macías and Hurley (2017), using a sample of about 43,000 to populate 3123 out of 3784 hypothetical job cells, means that for a large number of small jobs, the task scores are based on very few observations.[8]

Not accounting for sectoral differences may cause researchers to mistake differences in the sectoral composition of the economy for differences in occupational tasks and lead to a false positive finding of country differences in task content. Two graphs in Appendix C (see Supplemental File) detail the extent of sectoral differences by major occupational group. They show that the overall import of this is small: the outliers tend to be unlikely occupation-sector combinations such as clerks in agriculture or machine operators and assemblers in education which do not weigh heavily in the aggregate index. Goos and Manning (2007) likewise report very similar results whether they define jobs by occupation alone or by a combination of occupation and industry. Thus, since at present it is not feasible to address all potential sources of variation at once, I focus on the occupational level.

The indices are constructed by standardising the constituent variables to have a mean of 0 and a standard deviation of 1 following Acemoglu and Autor (2011) and then first averaging across individual survey respondents and subsequently across two-digit ISCO codes. Principal components analysis, which is sometimes used in the literature, is not useful in the present case because of the low number of items that make up each index. Thus, the routine index $rscore2d$ for occupation o is calculated as

$$rscore2d_o = \frac{\displaystyle\sum_{i \in I_o} \left(\frac{\displaystyle\sum_{j \in J} ewcs_{ji}}{J} \right)}{I_o}$$

where $ewcs_{ji}$ is the value for individual i on question j, J is the set of items used to calculate $rscore2d$ and I_o is the set of individuals with occupation o. Analogously, the complexity index $cscore2d_o$ is calculated as

$$cscore2d_o = \frac{\displaystyle\sum_{i \in I_o} \left(\frac{\displaystyle\sum_{l \in L} ewcs_{li}}{L} \right)}{I_o}$$

where L denotes the set of indicators used to calculate the index. For the wave- and country-specific indices, the respective subscripts w and c have to be added to the formula. In Appendix Table B1 of the Supplemental File, I provide detailed summary statistics for the EU-27 version of the indices.

Comparing measures of task content

This section describes the novel measures and compares them to the 'competitor' measures of Autor and Dorn (2013) and Fernández-Macías and Hurley (2017). It shows how the novel measures address the concerns formulated in Hypotheses 1.1, 1.2 and 2.1 and yield more plausible results for individual occupations.
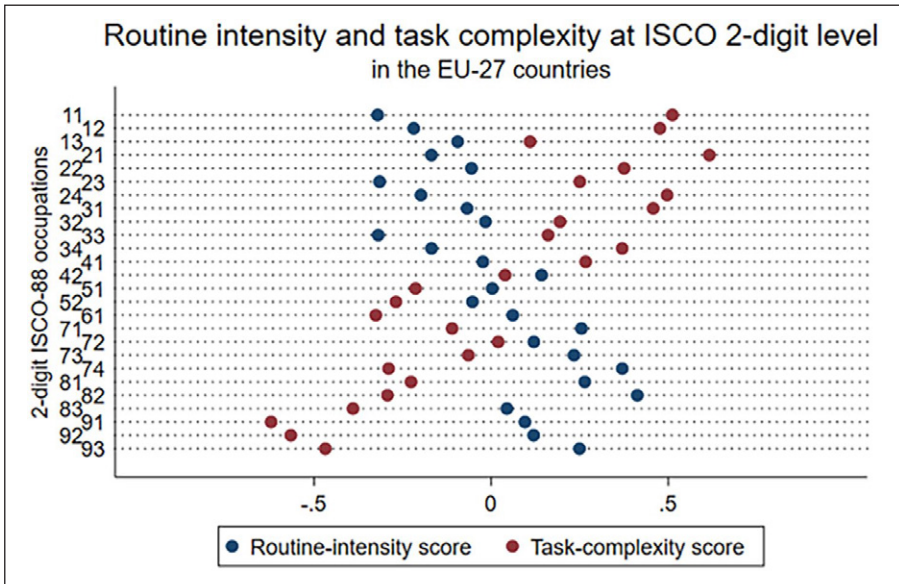
Figure 1. Routine-intensity and task complexity at ISCO two-digit level in the EU-27 countries.
See the appendix for the list of ISCO-88 codes at the two-digit level.

Describing the RTI and complexity indices

A plot of the routine and complexity indices for the EU-27 at the two-digit ISCO-88 level in Figure 1 shows a relatively linear increase of routine-intensity down the occupational hierarchy and a countervailing decrease of task complexity. Thus, at least based on this ordering of occupations, routine occupations do not cluster around the middle of the occupational distribution.

Echoing the findings of previous research on routine tasks, cognitive tasks and complexity, the analyses find an inverse relationship between RTI and complexity. The Spearman correlation between the two indices is −0.73 and the weighted Pearson correlation, at −0.66, is in the range reported in other studies for the correlation between RTI and cognitive task intensity. This shows that it remains a challenge to develop an index based on a priori considerations with dimensions that do not explain the same underlying variation.

Table 2 reports analyses of variance (ANOVAs), which suggest that about 24% of the variation in RTI is between occupations, while 76% is between workers within occupations. For task complexity, the numbers are 37% between-variation and 63% within-variation. The corresponding intraclass correlations confirm that within occupations, individual observations are more similar when it comes to task complexity: the intraclass correlation is 0.26 for the complexity indices, compared to 0.09 for the RTI indices. The finding of a large component of within-occupation variation is in line with previous research (e.g. Spitz-Oener, 2006) and illustrates the benefits of using survey data rather

Table 2. One-way analysis of variance.

Response variable	SS	df	F (sig.)	ICC	Estimated standard deviation
Routine – EU-27					
Between occupations	3464.5	25	399.19 (0.00)	0.0888	0.1839
Within occupations	37,305.3	107,462			0.5892
Total	40,769.8	107,487			
Complexity – EU-27					
Between occupations	12,685.7	25	1473.59 (0.00)	0.2649	0.3522
Within occupations	37,004.1	107,462			0.5868
Total	49,689.8	107,487			

The ratio of the estimated standard deviation over the combined estimated standard deviations gives the share of variation between and within occupations.
SS: Sum of squares; ICC: Intraclass correlation

than expert-coded measures. While broad occupational categories play an important role in explaining what people do in their jobs, they cannot capture the full complexity of individual workplaces – especially with regard to the extent of variation in routine tasks required from people in comparable occupations.[9]

Comparing routine indices

The RTI measure of Autor and Dorn (2013) has become the standard 'off-the-shelf' measure for RTI in the US and beyond. Fernández-Macías and Hurley (2017) improve on their measure with an approach which is more closely related to the method proposed here but requires further refinements. Therefore, comparing the new measure with these two RTI indices is paramount.

While Autor and Dorn (2013) work at the level of US census occupations, Goos et al. (2014) take their measure and map it onto ISCO-88, thus making the index applicable outside the US. The differences between Autor and Dorn (2013) and my RTI measure become visible in Figure 2. The markers are dispersed widely over the plot region and the line of best fit from a weighted linear regression meets the *y*-axis nowhere near the origin. This, and the relatively low adjusted R^2 of 0.35, implies that my operationalisation is substantively different from the approach commonly adopted in labour economics.

A look at the outliers is instructive to assess why the indices differ so much. Office clerks (group 41) have the highest RTI score of all occupations in Autor and Dorn (2013) but are just below the median according to my measure. This illustrates, in my view, the unrealistic characterisation of clerical occupations as far surpassing any other occupation in routine-intensity. While secretaries, finance clerks or librarians undoubtedly perform a fair share of routine tasks, it seems implausible that their job tasks are vastly more routine-intensive than those of printing machine operators, mechanical equipment assemblers or weavers. Time likely plays a role here, since by 2015 clerical occupations undoubtedly had become less routine-intensive compared to 1977, the year from which the task data in Autor and Dorn (2013) are taken.
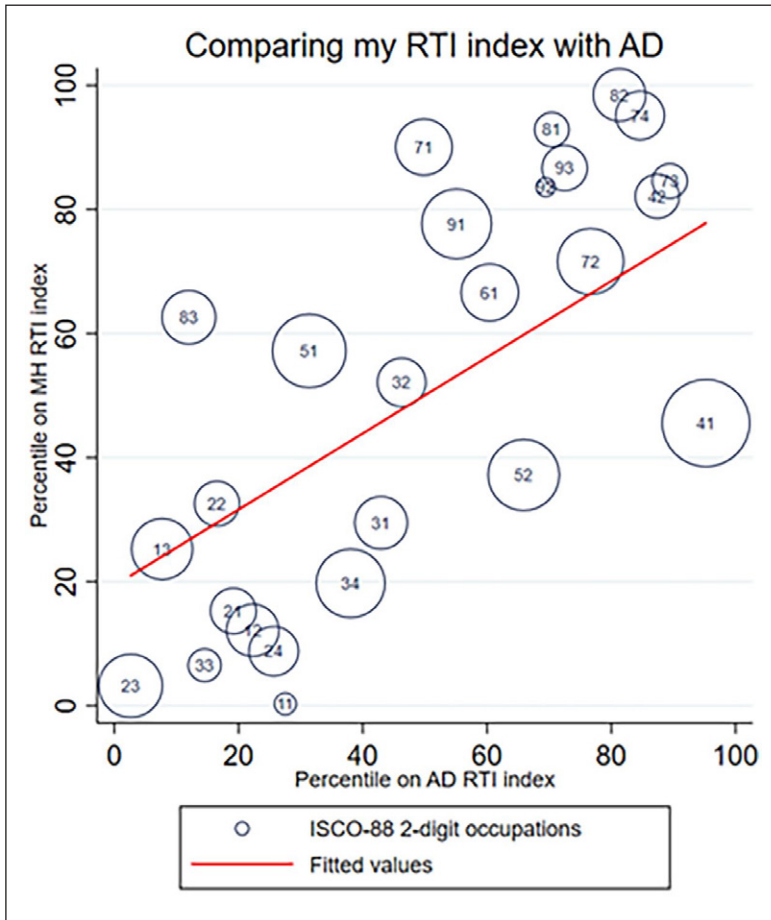
Figure 2. Comparing my RTI index with AD.
Adjusted $R^2 = 0.351$. Sample: EU-15, occupations weighted by employment in wave 3.

Overall, my measure tends to assign relatively higher routine-intensity scores to occupations in major groups 7, 8 and 9 (plant and machine operators, and assemblers; craft and related trades occupations; and elementary occupations). At the same time, managers, professionals and clerks tend to receive lower RTI scores than in Autor and Dorn (2013). Most of this accords with the classical RBTC hypothesis, but the finding that elementary occupations – which include many low-skilled service jobs – are relatively high routine contradicts the notion that displaced routine workers would move into such occupations that are classified as manual interactive in Autor and Dorn (2013).

Next is the comparison with the measures proposed by Fernández-Macías and Hurley (2017). I use data from their Table 1 which provides aggregated data at the two-digit ISCO-88 level that can be directly compared to my index (Fernández-Macías and Hurley,
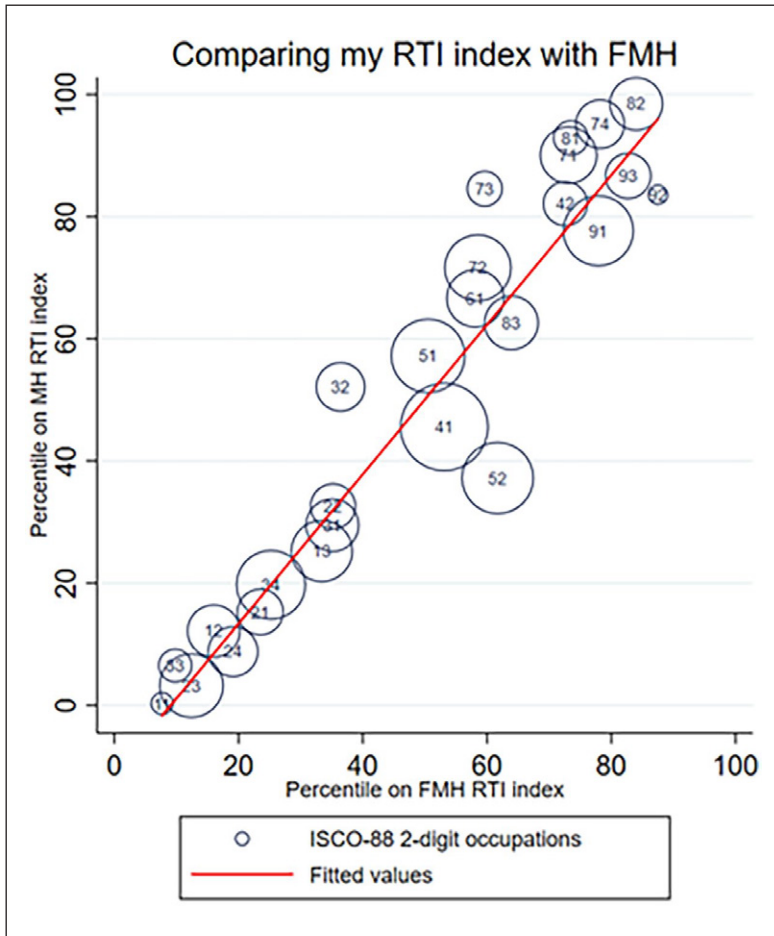
Figure 3. Comparing my RTI index with FMH.
Adjusted $R^2 = 0.866$. Sample: EU-15, occupations weighted by employment in wave 3.

2017: 574). Here, the occupations align much better; discrepancies are mainly visible with regard to medium- to high-routine occupations. Nevertheless, my findings qualify their stance somewhat. Salespersons and elementary occupations all are less routine-intensive according to my measure. All three elementary occupations are among the five most routine-intensive occupations according to Fernández-Macías and Hurley (2017), while only one (labourers in mining, manufacturing, construction and transport) makes the top five according to my method. Conversely, blue-collar occupations in major groups 7 and 8 are closer to the upper end of the RTI scale on my index. It seems plausible that crafts and manufacturing occupations which frequently require a significant degree of job-specific training would be more standardised and repetitive than unskilled elementary occupations in which workers often take over low value-added tasks from their more skilled colleagues (Figure 3).

Overall, this suggests that Fernández-Macías and Hurley (2017) are going too far when they claim that routine tasks are most frequent at the bottom of the 'skills-wage-cognitive tasks continuum' (p. 575). They underestimate the routine-intensity of crafts and manufacturing occupations, which are middling occupations in terms of wages and skills. The less skilled and lower paid elementary occupations arguably comprise a less standardised set of tasks and are consequently slightly less routine-intensive. My index reflects this.

Thus, the three routine measures have different implications for occupational hierarchies. All are unanimous that managerial and professional occupations are the least routine-intensive; however, differences are visible when considering medium- and high-routine occupations. Regarding clerical occupations, and some crafts and manufacturing occupations, as well as service and elementary occupations, there are large discrepancies. Some of these occupations account for substantial portions of total employment. This shows that better alignment of the concept of routine-intensity and its measurement, compared to Autor and Dorn (2013) and Fernández-Macías and Hurley (2017), results in a partial reshuffling of the list of high-routine occupations. This, in turn, may have far-reaching consequences for studies of RBTC and employment and wage changes.

Comparing cognitive and complexity indices

My complexity index provides a measure of skill bias as discussed in the 'Overview and critique of existing operationalisations' section. It has its counterpart in the cognitive index that is proposed by Fernández-Macías and Hurley (2017) and in Eurofound publications involving the same authors (Eurofound, 2014, 2017). However, they do not frame their index as a tool for analysing skill bias but as the 'other side of the same coin' as their routine measure. Recalling the 'Data and construction of the indices' section, the only direct overlap is that both measures ask whether a person's job involves complex tasks. Nevertheless, the resulting ordering of occupations is remarkably similar, as Figure 4 shows.

Indeed, despite substantial differences in the construction of the two indices, the only larger discrepancies are between life science and health associate professionals, and teaching professionals and agricultural occupations. However, there is no broader group of occupations that is systematically ranked differently on the complexity index compared to the cognitive index. Thus, even though the concept and operationalisation are different, the practical implications of operating with task complexity rather than cognitive intensity are likely to be small.

To summarise the comparison, Table 3 displays the rank order correlation between the various indices discussed and shows that the ordinal rankings of occupations are fairly similar. While some degree of similarity is to be expected, in the case of the complexity and cognitive indices it is striking how very different questions yield an almost identical ordering of occupations. The Autor and Dorn (2013) index, calculated as a measure of predominance of routine tasks, exhibits comparatively lower correlations with my and Fernández-Macías and Hurley's (2017) routine indices which measure the prevalence of the respective tasks. Furthermore, there are relatively strong negative
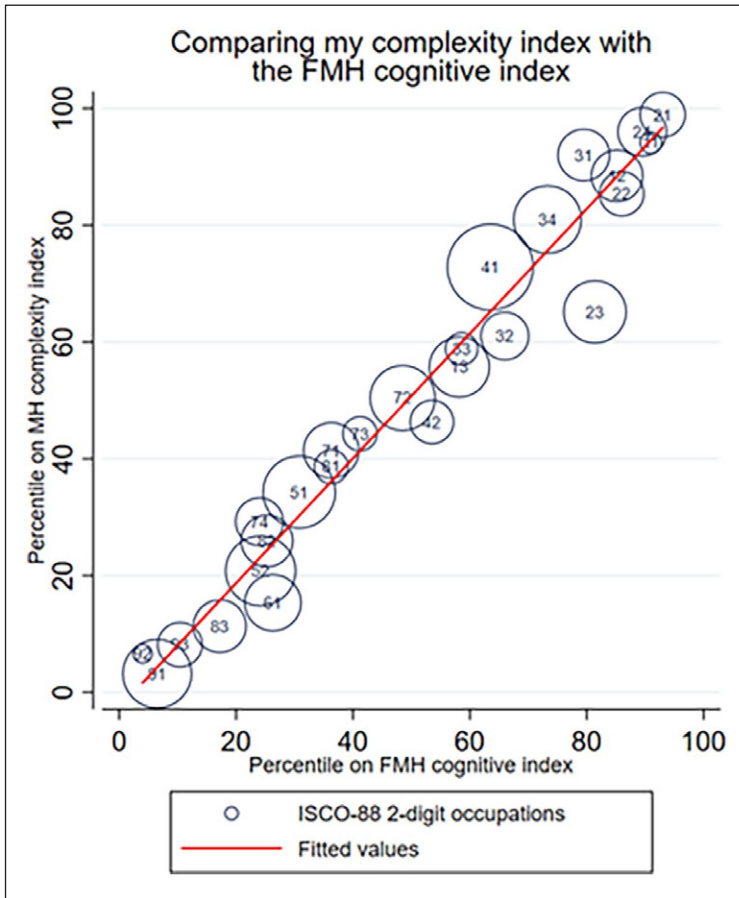
Figure 4. Comparing my complexity index with the FMH cognitive index.
Adjusted R^2 = 0.945. Sample: EU-15, occupations weighted by employment in wave 3.

Table 3. Rank order correlations between the indices.

Index	RTI (this art.)	AD RTI	FMH RTI	Complexity (this article)	FMH Cognitive
RTI (this art.)	I	0.74	0.94	−0.73	−0.76
AD RTI		I	0.69	−0.40	−0.47
FMH RTI			I	−0.83	−0.86
Comp. (this art.)				I	0.98
FMH Cog					I

Calculated based on employment shares in the EU-15 in 2000 for my indices and the Autor and Dorn (2013, AD) RTI index, and taken from Table 1 in Fernández-Macías and Hurley (2017, FMH) for their RTI and cognitive indices. RTI: routine task intensity.

correlations between the routine indices and the auxiliary measures of complexity and cognitive intensity. This is expected, yet it does not imply that they are two sides of the same coin, as there are several high-routine complex occupations and vice versa. Robustness checks using different samples yield similar results, as shown in the Appendix Supplemental File. Overall, the comparisons show that the routine-intensity and task complexity indices do not simply replicate previous research. The complexity index is more conceptually meaningful, and the more theoretically informed operationalisation of both measures leads to a reappraisal of the task content of some occupations. Furthermore, the importance of using survey data to analyse within-occupation variation is confirmed.

New opportunities for research

In addition to the improvements highlighted in the previous section, the measures developed in this article rely on superior data and so create exciting new opportunities for research into task change over time and differences between countries, thus addressing the concerns formulated in Hypotheses 2.2 and 2.3.

Task content over time

It is widely accepted that not only does the prevalence of occupations change in response to technological advances but also the tasks they entail (Spitz-Oener, 2006). Yet, in the literature on RBTC, this facet of technology has received scant attention. The few studies that do look at within-occupation change over time find contradictory results. My index makes it possible to analyse within-occupation changes in the EU-15 countries over four points in time over a 15-year period from 2000 until 2015, thus covering a wider geographical area and a more recent time period than other studies. I find that there was no consistent trend towards less routine-intensive work and a noticeable increase in overall occupational complexity. At the level of individual occupations, wave-on-wave increases in complexity are associated with a decline in routine-intensity.

Figure 5 illustrates the different patterns for RTI and complexity. While for RTI there is no clear movement in one direction and the 15-year differences are not statistically significant in most cases, there is an almost uniform increase in task complexity, with statistically significant increases in 19 of the 26 two-digit occupations. The increases in task complexity tend to be larger in the more complex occupations, giving rise to further divergence between simple and complex occupations. These takeaways are reinforced if one considers the changing sizes of occupational groups. In Appendix C (Supplemental File), it is shown that occupations with above- and below-median routine-intensity converged in terms of RTI while the employment share of routine occupations declined. Contrary trends at the extensive and intensive margins therefore result in trendless fluctuation of overall RTI. Concerning complexity, compositional and task changes have reinforced each other and contributed to strong overall upskilling.

Finally, I consider to what extent changes in RTI and task complexity are related. Plotting the wave-on-wave changes in RTI and task complexity for each occupation and regressing one on the other, a moderate negative relationship between changes in the two task dimensions is visible and depicted in Figure 6. More precisely, a reduction of the
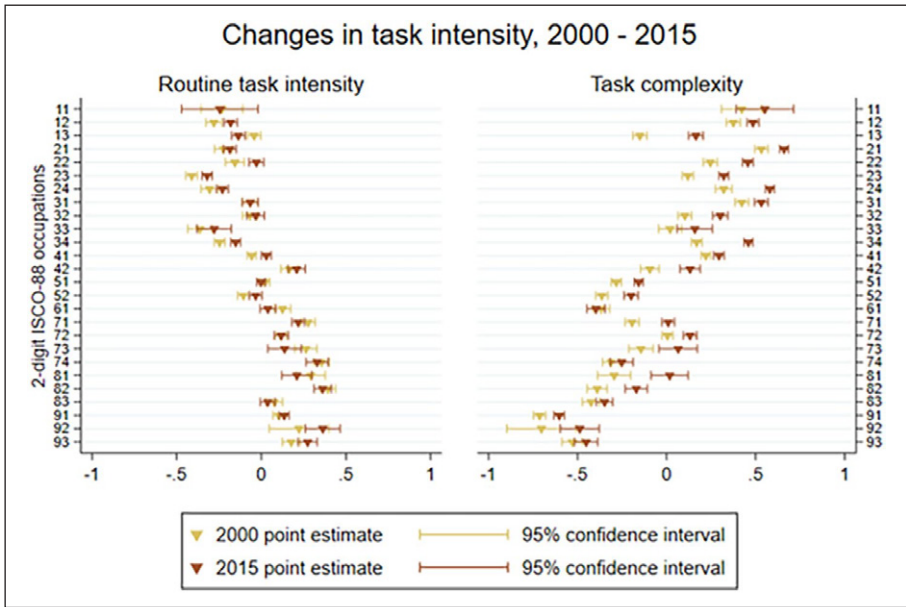
Figure 5. Changes in task intensity, 2000–2015.
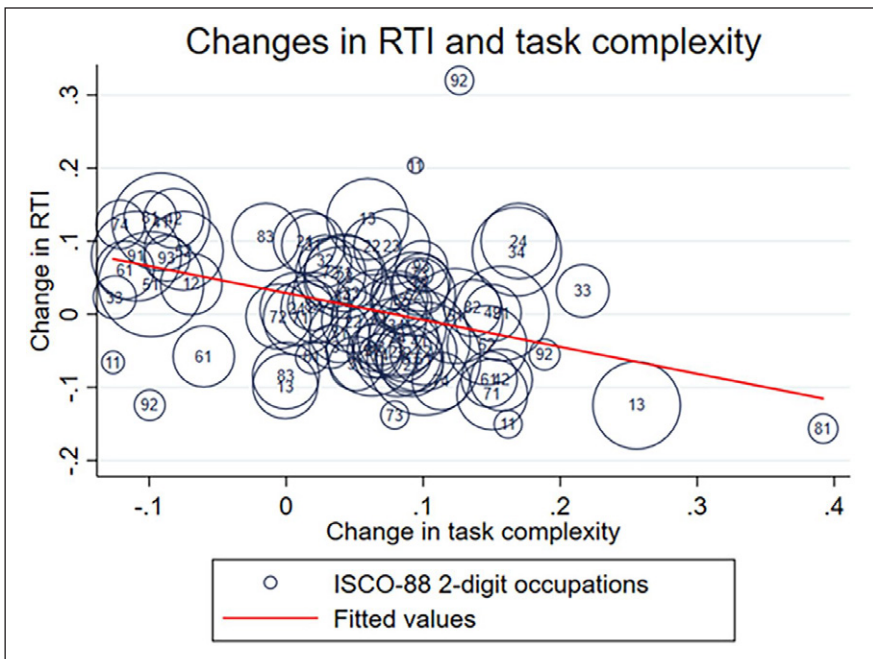Sample: EU-15 countries, scores for waves 3 and 6 based on the pooled dataset.

Figure 6. Changes in RTI and task complexity.
Adjusted $R^2 = 0.201$. Sample: EU-15, occupations weighted by employment in each wave.

Table 4. Rank order correlation of the country-specific RTI and complexity rankings.

	AT	BE	DE	DK	ES	FI	FR	GR	IE	IT	LU	NL	PT	SE	UK	EU-15
AT		0.96	0.94	0.96	0.92	0.94	0.93	0.92	0.95	0.93	0.91	0.97	0.94	0.95	0.96	0.97
BE	0.74		0.95	0.94	0.96	0.92	0.97	0.94	0.95	0.95	0.92	0.96	0.95	0.96	0.95	0.98
DE	0.84	0.87		0.95	0.91	0.92	0.94	0.96	0.94	0.94	0.94	0.92	0.90	0.94	0.93	0.96
DK	0.80	0.83	0.87		0.92	0.92	0.93	0.92	0.97	0.95	0.93	0.94	0.95	0.97	0.95	0.97
ES	0.95	0.75	0.84	0.84		0.86	0.94	0.77	0.89	0.92	0.80	0.95	0.95	0.84	0.81	0.97
FI	0.83	0.86	0.86	0.90	0.79		0.86	0.74	0.86	0.90	0.86	0.90	0.90	0.85	0.79	0.92
FR	0.82	0.82	0.82	0.81	0.85	0.87		0.94	0.96	0.92	0.86	0.95	0.92	0.92	0.95	0.97
GR	0.68	0.66	0.72	0.72	0.66	0.82	0.75		0.89	0.85	0.87	0.90	0.78	0.90	0.89	0.81
IE	0.76	0.82	0.85	0.84	0.77	0.84	0.83	0.71		0.94	0.89	0.95	0.89	0.94	0.94	0.93
IT	0.92	0.82	0.89	0.90	0.95	0.79	0.89	0.68	0.84		0.89	0.92	0.94	0.92	0.86	0.95
LU	0.77	0.82	0.73	0.69	0.71	0.60	0.77	0.41	0.65	0.73		0.88	0.81	0.95	0.90	0.83
NL	0.82	0.79	0.83	0.91	0.84	0.85	0.89	0.69	0.76	0.89	0.76		0.93	0.95	0.95	0.97
PT	0.73	0.65	0.83	0.66	0.64	0.69	0.57	0.63	0.74	0.65	0.51	0.57		0.85	0.80	0.97
SE	0.74	0.87	0.85	0.89	0.69	0.91	0.80	0.74	0.81	0.72	0.61	0.86	0.69		0.95	0.87
UK	0.81	0.86	0.86	0.88	0.76	0.88	0.79	0.65	0.94	0.83	0.68	0.79	0.79	0.86		0.85
EU-15	0.90	0.86	0.90	0.92	0.94	0.89	0.93	0.75	0.88	0.96	0.74	0.92	0.70	0.84	0.88	

RTI: routine task intensity.
Upper triangle: complexity index – lower triangle: RTI index.
AT: Austria; BE: Belgium; DE: Germany; DK: Denmark; ES: Spain; FI: Finland; FR: France; GR: Greece; IE: Ireland; IT: Italy; LU: Luxembourg; NL: Netherlands; PT: Portugal; SE: Sweden; UK: United Kingdom

RTI measure by 0.1 points is associated with an increase in complexity by roughly 0.3 points. Of course, there is no reason to assume that changes in complexity cause changes in routine-intensity or vice versa; rather, it stands to reason that an omitted variable – technological change – affects both simultaneously. Thus, even though there have been conflicting trends in low-routine and high-complexity occupations, broadly speaking, occupations that have become more complex have also become less routine-intensive, and vice versa. A more detailed investigation of this relationship is provided in Appendix C (Supplemental File).

Task content across countries

This article is also the first to account for country differences by calculating country-specific versions of the routine and complexity indices. I find that there are indeed non-negligible differences between countries in the ordering of occupations, validating the argument that researchers should avoid applying one country's task data to another country, wholesale.

Supplementary analyses in Appendix C (see Supplemental File) show that the country-specific measures exhibit a range which suggests that country-specific data are to be preferred unless small sample sizes make measurement imprecise. However, most importantly, the differences in the ordinal ranking of occupations across countries are considered. The analysis shows that there are indeed strong reasons to prefer country-specific or pooled data for cross-country analyses. Tabulating the rank order correlations of the country-specific measures in the EU-15 countries shows that the ordering of

occupations in terms of routine-intensity and task complexity is by no means identical even in countries as similar as the EU-15.

The correlations are displayed in Table 4, with the grey lower triangle containing the correlation between RTI rankings and the white upper triangle that of the complexity indices. It shows that the average correlation between two countries' RTI ranking is of a similar magnitude as the correlation between my pooled RTI index and the Autor and Dorn (2013) index, at 0.78. At the same time, the average correlation of each individual country's RTI ranking with the pooled index is substantially higher, at 0.87. Only 18% of the 105 country dyads show a higher correlation, and not a single country has a higher average correlation with the 14 other countries in the sample, driving home the point that it is highly problematic to assume that occupational routine-intensity is constant across countries. The differences are less pronounced, with regard to the complexity index, with an average correlation of 0.93 between countries and the pooled index, and 0.91 between country dyads, of which a full 52% have a correlation greater than 0.93. Overall, because of the clear advantages associated with the country-specific measures of routine-intensity, analysts should use such measures wherever that is feasible. If it is not, this analysis shows that an index pooled over many countries will, on average, still be closer to a country's true ranking than an index based on data from a single country, at least when it comes to RTI. These statistics show that differences in task content over time and between countries deserve greater attention as argued in the 'Overview and critique of existing operationalisations' section, and that my indices provide a tool for performing such analyses. This constitutes a significant improvement over the indices of not only Autor and Dorn (2013) but also Fernández-Macías and Hurley (2017).

Conclusion

The literature on occupational task content has long relied on just a few off-the-shelf measures without giving much thought to how key concepts are theorised and operationalised. Hence, cognitive tasks were defined without a clear purpose, routine tasks operationalised with unsuitable variables and task variation within occupations, ignored. Further ignored were change over time and diversity across countries. Yet, for a meaningful comparative analysis of the effects of technological change, conceptually and empirically sound measures are crucial. This article identifies and discusses the shortcomings of existing measures and proposes new indices which address the problems and offer researchers flexible tools for analysing occupational task content.

The new indices entail the following improvements. First, both the routine-intensity and the complexity index have a clear theoretical interpretation: they capture the task characteristics that the RBTC and SBTC theories focus on, respectively. Second, the variables used to operationalise them really quantify the essence of the underlying concepts. This places some occupations very differently in the routine hierarchy. Third, the indices use survey data rather than expert-coded task data which is important for understanding within-occupation variation.

Furthermore, with the wave- and country-specific indices, a much more detailed analysis of the impact of RBTC and SBTC on occupational change becomes possible for the first time. This represents a significant improvement over the measures developed by Autor and Dorn (2013) and goes beyond the important contribution of Fernández-Macías

and Hurley (2017). My descriptive analysis of the novel measures shows trends and differences that are too substantial to ignore. This calls into question the practice of applying task measures from one country or year in very different contexts. Further research into the reasons for and consequences of these differences may have important implications for the understanding of occupational and technological change.

The main limitations of this study are the sample size of the EWCS which prevents more disaggregated analyses at the occupation-sector level, and concerns about potential measurement error introduced by individuals understanding the survey questions differently. Furthermore, other authors have also proposed alternative approaches to Autor and Dorn (2013) and Fernández-Macías and Hurley (2017). For example, Salvatori (2018) develops a routine-intensity index using UK data based on Autor and Dorn's (2013) methodology. However, Salvatori's and similar contributions lack a comprehensive methodological discussion and a generalisable and flexible set of measures. Thus, the measures proposed here will be useful for future research on the nature of contemporary technological change. Much remains to be learned about how work is changing and what role technology plays in the process.

Declaration of conflicting interests

Funding

ORCID iD

Matthias Haslberger [iD] https://orcid.org/0000-0003-3056-6630

Supplemental material

Supplemental material for this article is available online. Appendices A to C can be found at: http://journals.sagepub.com/doi/suppl/10.1177/10353046211037095

Notes

1. While the emphasis on codifiability is sensible, there is a danger of circularity: if routine tasks are defined as codifiable tasks that are being replaced by machines, technological change is by construction routine-biased.
2. A third approach is to classify routine occupations based on census one-digit occupational codes as in Acemoglu and Autor (2011). However, this very coarse method is clearly inferior to either expert-coded scores or worker self-assessments (Salvatori, 2018).
3. http://journals.sagepub.com/doi/suppl/10.1177/10353046211037095
4. Sebastian and Biagi (2018) provide an overview of the years from which task data in various studies are taken.
5. These items correspond to question numbers 30e, 48a, 48b, 53a and 53d in the European Working Conditions Survey (EWCS).
6. These items correspond to question numbers 30i, 53c, 53e and 53f in the EWCS.
7. There are a few exceptions: the target sample size was 500 in Luxembourg in 2000, and 600 in Cyprus, Estonia, Luxembourg, Malta and Slovenia in 2005.

8. For the same reason, ISCO (International Standard Classification of Occupations) group 62 has been merged with group 61. This change affects 0.4% of observations in the original dataset.
9. The logic of the ISCO approach to ordering occupations is closely related to the logic behind the complexity dimension, as it is based on similarity in the skill level and skill specialisation of the tasks that make up a job. Routine-intensity, on the contrary, is less directly linked to skills.

References

Acemoglu D and Autor D (2011) Skills, tasks and technologies: implications for employment and earnings. In: Ashenfelter O and Card D (eds) *Handbook of Labor Economics*, vol. 4b. Amsterdam: Elsevier, pp. 1043–1171.

Autor D and Dorn D (2013) The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review* 103(5): 1553–1597.

Autor D and Handel MJ (2013) Putting tasks to the test: human capital, job tasks, and wages. *Journal of Labor Economics* 31(2): S59–S96.

Autor D, Levy F and Murnane RJ (2003) The skill content of recent technological change: an empirical exploration. *The Quarterly Journal of Economics* 118(4): 1279–1333.

Caines C, Hoffmann F and Kambourov G (2017) Complex-task biased technological change and the labor market. *Review of Economic Dynamics* 25: 298–319.

Eurofound (2014) *Drivers of Recent Job Polarisation and Upgrading in Europe: European Jobs Monitor 2014*. Luxembourg: Publications Office of the European Union.

Eurofound (2017) *Occupational Change and Wage Inequality: European Jobs Monitor 2017*. Luxembourg: Publications Office of the European Union.

Fernandez RM (2001) Skill-biased technological change and wage inequality: evidence from a plant retooling. *American Journal of Sociology* 10(2): 273–320.

Fernández-Macías E and Hurley J (2017) Routine-biased technical change and job polarization in Europe. *Socio-economic Review* 15(3): 563–585.

Goos M and Manning A (2007) Lousy and lovely jobs: the rising polarization of work in Britain. *The Review of Economics and Statistics* 89(1): 118–133.

Goos M, Manning A and Salomons A (2014) Explaining job polarization: routine-biased technological change and offshoring. *American Economic Review* 104(8): 2509–2526.

Handel MJ (2017) Measuring job content: skills, technology, and management practices. In: Buchanan J, Finegold D, Mayhew K, et al. (eds) *The Oxford Handbook of Skills and Training*. Oxford: Oxford University Press. Available at: https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199655366.001.0001/oxfordhb-9780199655366-e-5 (accessed 17 July 2021).

Salvatori A (2018) The anatomy of job polarisation in the UK. *Journal for Labour Market Research* 52(8): 1–15.

Sebastian R and Biagi F (2018) The routine biased technical change hypothesis: a critical review. JRC Technical Report, European Commission, Luxembourg: Publications Office of the European Union.

Spitz-Oener A (2006) Technical change, job tasks, and rising educational demands: looking outside the wage structure. *Journal of Labor Economics* 24(2): 235–270.

Author biography

Matthias Haslberger is currently a Doctor of Philosophy student in Social Policy at the University of Oxford researching how technological and institutional change influence employment and wage inequality. In particular, he is interested in the potential of labour market institutions to contain rising inequality in advanced economies facing the effects of technological change.