ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Rohrbach-Schmidt, Daniela; Tiemann, Michael

Article

Changes in workplace tasks in Germany—evaluating skill and task measures

Journal for Labour Market Research

Provided in Cooperation with:
Institute for Employment Research (IAB)

Suggested Citation: Rohrbach-Schmidt, Daniela; Tiemann, Michael (2013) : Changes in workplace tasks in Germany—evaluating skill and task measures, Journal for Labour Market Research, ISSN 2510-5027, Springer, Heidelberg, Vol. 46, Iss. 3, pp. 215-237, https://doi.org/10.1007/s12651-013-0140-3

This Version is available at:
https://hdl.handle.net/10419/158798

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

J Labour Market Res. (2013) 46:215–237
DOI 10.1007/s12651-013-0140-3

ARTICLE

Changes in workplace tasks in Germany—evaluating skill and task measures

Daniela Rohrbach-Schmidt · Michael Tiemann

Published online: 21 August 2013
© Institut für Arbeitsmarkt- und Berufsforschung 2013

Abstract The task approach is attracting increasing attention and recognition among scholars in economics, sociology and related fields. However, measurement still presents an important challenge to the task approach. This paper studies the comparability of task measures in the commonly used German BIBB/IAB-BIBB/BAuA employment cross-sections on qualification and working conditions from 1979, 1985/86, 1991/92, 1998/99, and 2006. We hypothesize that findings on task-biased technological change are sensitive to variable choice. The task data differ substantially from other task data, and task items are not readily comparable between survey years. As a result, classifying single task items into distinct domains leads to a number of problems. To test our hypothesis, we use different strategies for classifying tasks into task domains, and analyze whether different operationalizations lead to different conclusions about task change in Germany. Our results show that this indeed is the case. Our paper provides readers with a broader understanding of German task data and gives recommendations for applying the task approach to German employment data.

Wandel der Tätigkeiten am Arbeitsplatz in Deutschland – Analysen von Skill und Task-Maßen

Zusammenfassung Der Task-Ansatz findet zunehmende Beachtung unter Ökonomen, Soziologen und verwandten Disziplinen. Allerdings stellen Operationalisierung und Messung eine besondere Herausforderung des Ansatzes dar. Dieser Artikel untersucht die Vergleichbarkeit von unterschiedlichen Operationalisierungen des Task-Ansatzes anhand der BIBB/IAB-BIBB/BAuA Erwerbstätigenbefragungen 1979, 1985/86, 1991/92, 1998/99 und 2006. Unsere Hypothese ist, dass Ergebnisse zum task-biased technological change in Deutschland von der Operationalisierung der Tätigkeitsgruppen abhängig sind. Die Tätigkeitsdaten der Erwerbstätigenbefragungen unterscheiden sich grundlegend von anderen der im Rahmen des Task-Ansatzes verwendeten Datensätzen; einzelne Tätigkeitsitems sind zudem nicht ohne weiteres zwischen den einzelnen Erhebungswellen vergleichbar. Dementsprechend ist die Zuordnung einzelner Tätigkeitsitems zu Gruppen mit einer Reihe von Schwierigkeiten verbunden. Um unsere Annahme zu prüfen, nutzen wir unterschiedliche Klassifizierungsstrategien für die Zuordnung einzelner Tätigkeitsitems zu Tätigkeitsgruppen und untersuchen, ob die alternativen Operationalisierungen zu unterschiedlichen Schlussfolgerungen zum Tätigkeitswandel in Deutschland führen. Unsere Ergebnisse bestätigen das. Der Artikel beschreibt die deutschen Tätgkeitendaten der Erwerbstätigenbefragungen und gibt Empfehlungen für die Anwendung des Tätigkeitsansatzes mit diesen Daten.

D. Rohrbach-Schmidt (✉) · M. Tiemann
Robert-Schuman-Platz 3, 53175 Bonn, Germany
e-mail: Rohrbach@bibb.de

M. Tiemann
e-mail: Tiemann@bibb.de

1 Introduction

In 2003, Autor, Levy and Murnane (ALM) proposed a task-based framework to investigate how the adoption of computer technology changes job tasks and employer demand for human skills. This approach focuses on determining the tasks that computers are best suited to perform and whether computer-performed tasks serve as complements or substitutes for human job skills (Autor et al. 2003, p. 1280). The authors affirm that computers substitute for cognitive and manual tasks that have large routine components. At

the same time, computers also complement complex non-routine problem-solving tasks and complex communications tasks. Conversely, low-skilled, manual, non-routine jobs are not directly affected by computerization. Therefore, their hypothesis is that computerization leads to a decline in the demand for middle-education workers and leads to an increase in the relative demand for both the most educated and least educated workers. Studies for the U.S., Great Britain and certain European countries suggest that the relative demand trends for skilled workers have been polarized since the 1990s (Autor et al. 2006, 2008; Goos and Manning 2007; Spitz-Oener 2006; Dustmann et al. 2009 and Goos et al. 2009).

Autor and Handel (2009) argue that one of the challenges to the task approach is measurement. To date, researchers have mostly employed the task approach with job task measures at the occupational level. By contrast, Autor and Handel (ibid.) use both occupational and worker level data. With both types of data, they show that job tasks vary significantly among workers in a given occupation, that variance in job tasks is systematically related to the worker's race, gender and education and that job tasks at the worker level are significant predictors of wages.

The German BIBB/IAB- and BIBB/BAuA employment cross-sections on qualification and working conditions from 1979, 1985/86, 1991/92, 1998/99, and 2005/2006[1] (also known as the "Qualification and Career Surveys") are rare examples of databases providing information on job tasks at the worker level over a long time period. Spitz-Oener (2006) first used this data to apply the task approach to the German case. Following her approach, many subsequent publications have made important contributions to the debate on task-biased technological change in Germany (Gathmann and Schoenberg 2010; Spitz-Oener 2006, 2008; Black and Spitz-Oener 2010; Antonczyk et al. 2009).

Unquestionable, the German data are an immensely valuable source for research on the task approach. Importantly, however, the task items reported in these cross-sections were not originally intended to measure the ALM task domains, and the data differ substantially from other task data. Moreover, as we will demonstrate, the data have considerably changed over the years. In conclusion, the process of classifying single tasks into distinct domains leads to a number of problems which we will address here.

Given the background of past research, this paper aims to evaluate the comparability of task measures in the German data and to explore in how far results on task change in Germany are robust to variable choice. We hypothesize that findings depend on the measures specified. To test our hypothesis, we apply different strategies to classify tasks into task domains, and we subsequently analyze how results on task change for these different operationalizations compare. Our results demonstrate that different operationalizations lead to different conclusions about task change in Germany.

The paper is organized as follows. We start by briefly reviewing the literature on the task approach, focusing on how the relevant task domains are determined and how they are measured. The third section focuses on the data. We give an overview of our methodologies and describe the general content of the data with a particular emphasis on task items. In the fourth section, we apply three different strategies of operationalizations to our data. The fifth section compares the three operationalizations and explores in how far results depend on the measures specified. In the last section, we summarize our recommendations for applying the task approach to German data.

2 What are relevant tasks, and how are they measured?

Changes in job skill requirements, workplace tasks, technology use and new management practices are debated intensively in the labor market literature. However, most available data do not include measures that address the current research questions and approaches. The task framework is a prominent example for this. In their paper, ALM differentiate five domains of job tasks: routine manual tasks, routine cognitive tasks, non-routine manual tasks, and non-routine cognitive tasks, i.e., analytical and interactive tasks (Autor et al. 2003, p. 1286). The most relevant differentiation is between routine and non-routine tasks.

In their reading,

> a task is "routine" if it can be accomplished by machines following explicit programmed rules. Many manual tasks that workers used to perform, such as monitoring the temperature of a steel finishing line or moving a windshield into place on an assembly line, fit this description. Because these tasks require methodical repetition of an unwavering procedure, they can be exhaustively specified with programmed instructions and performed by machines. (ibid., p. 1283).

As Autor et al. (2003) conclude, the substitution of routine manual tasks is not new; however, computerization is now able to replace routine cognitive tasks, namely repetitive information-processing tasks, as well.

On the contrary, "non-routine" tasks are tasks,

> for which the rules are not sufficiently well understood to be specified in computer code and executed

[1] In 1979, 1985/86, and 1991/1992 data were surveyed by the Federal Institute for Vocational Education and Training (BIBB) in cooperation with the Institute of Employment Research (IAB) of the German Federal Labour Office (BA), and in the years 1998/99 and 2005/2006 together with the Federal Institute for Occupational Safety and Health (BAuA). Between October 2011 and March 2012, a sixth Qualification and Career Survey has been conducted by BIBB and BAuA.

> by machines. Navigating a car through city traffic or deciphering the scrawled handwriting on a personal check-minor undertakings for most adults-are not routine tasks by our definition (...). The reason is that these tasks require visual and motor processing capabilities that cannot at present be described in terms of a set of programmable rules [...]. (ibid., p. 1283).

The central argument of their seminal paper, therefore, is that computers substitute for cognitive and manual routine tasks at the same time as they complement cognitive non-routine tasks. (This is referred to as the "routinization hypothesis" in Goos et al. 2009.) As explained in the quote above, non-routine manual tasks such as navigating a car or cleaning are not directly affected because they require situational, visual, or language capabilities and interpersonal communication.

Moreover, ALM assume that different domains of job tasks are typically performed by different groups of skilled workers: cognitive non-routine tasks would be typical for (high-) skilled professional and managerial jobs, routine manual and cognitive tasks constitute most middle-education jobs, and manual non-routine tasks are mostly performed by unskilled workers. As a result, the task model predicts an increase in the relative demand for the most-skilled and least-skilled workers and a decline in the relative demand for middle-skilled workers.

ALM use the Fourth (1977) and Revised Fourth Edition (1991) of the U.S. Department of Labor's Dictionary of Occupational Titles (DOT) to apply the task model to the data. These databases include detailed evaluations of more than 12,000 occupations along with 44 objective and subjective job characteristics. As ALM note, these evaluations are based on first-hand observations of workplaces by Department of Labor examiners (Autor et al. 2003, p. 1292). ALM aggregate these evaluations at the level of three-digit Census Occupation Codes. These DOT data are merged with individual worker observations and aggregated at the level of 140 Census industries. To measure changes in job task requirements, these researchers inspect changes in the occupational distribution of employment (while holding task content within occupations at the 1977 DOT level constant). They also examine changes in task content measures within occupations by matching occupations between the two DOT editions (ibid.).

ALM select the DOT measure of "general educational development in mathematics" ("GED Math") as a measure for non-routine analytic tasks. The DOT measure of "adaptability to accepting responsibility for the direction, control, or planning of an activity" (DCP) is used to measure non-routine interactive tasks. For routine cognitive tasks, these researchers select the DOT measure of "adaptability to situations requiring the precise attainment of set limits, tolerances, or standards" (STS). As a measure of routine manual activity, ALM choose the DOT variable measuring the "ability to move fingers, and manipulate small objects with fingers" ("Finger Dexterity"). Finally, these researchers select the variable "Eye Hand Foot Coordination" as a measure for non-routine manual tasks. According to the DOT definition, this variable measures the "ability to move the hands and foot coordinately with each other in accordance with visual stimuli."[2]

Their analyses affirm changes in job task requirements toward non-routine tasks. They demonstrate that these changes were a relevant factor in U.S. demand shifts favoring educated labor beginning in the 1970s. ALM are sensitive to the concern that the choice of variables could matter; therefore, they reanalyze the data using alternative composite variables computed based on the results of a principal components analysis. They conclude that while their results are generally robust to variable choice, variable choice matters nevertheless. (ALM, p. 1306).

Following their approach, Goos and Manning (2007) use the DOT data to demonstrate pervasive job polarization into "lovely" (high-paid) and "lousy" (low-paid) jobs in Britain in the period from 1975 to 1999.[3] In their 2009 paper "Job polarization in Europe" Goos, Manning and Salomons use 96 variables from the US Occupational Information Network (O*NET) database to apply this approach to a cross-national comparative analysis of selected European countries. This paper documents similar polarization trends in these countries as well (for a more detailed reflection on the O*NET data, see Handel 2009). Following Autor and Dorn (2009), Goos et al. (2009, p. 60) use principal component analysis to construct measures for "abstract", "routine" and "service" tasks ("manual" tasks in Autor and Dorn (2009), using the Fourth DOT edition).

In their 2009 MIT working paper, Autor and Handel review the task approach literature and highlight that measurement presents an important challenge for the task approach (p. 3). They conclude that DOT and O*Net have substantial limitations for tasks measurement, most significantly because they only "provide information on job characteristics at the level of occupations, not workers" (Autor and Handel 2009, p. 3–4).

Autor and Handel combine task measures from O*Net with self-reported job tasks at the worker level from the Princeton Data Improvement Initiative (PDII) and classify nine single tasks items in measures for "cognitive", "interpersonal", and "physical" dimensions of job demands, cor-

[2] For single task pooled in these measures see Autor et al. (2003, p. 1323).

[3] On applying U.S. DOT job measures on British data the authors state: "The main data in this paper comes from Britain, but we would expect the task composition of occupations and the impact of technology to be very similar to that observed in the United States." (Goos and Manning 2007, p. 120).

responding to the Data, People, and Things classification used in the DOT (ibid., p. 12).[4] Thus, in this study, the authors do not differentiate between routine and non-routine tasks. Instead, the "physical" dimension is a combined measure of physical and routine tasks (for a similar approach, see Green 2012[5]). Thus, this paper's use of "routine tasks" with regards to the first item overlaps with tasks that are described as non-routine manual tasks in the 2003 ALM paper. The second item (repetitive tasks) however, seems to correspond with ALM's theoretical definition of routine tasks better than the operationalization in their 2003 paper; "Finger Dexterity" presumably includes also non-routine tasks. This has probably resulted in confusion about definition and following matching between manual working activities and routine tasks in different papers.[6]

Autor and Handel show that job tasks vary significantly among workers within a given occupation, that this variance is systematically related to the worker's race, gender and education and that job tasks at the worker level are significant predictors of wages, even when occupation-level job tasks are included in the model. Altogether, their analysis presents evidence for the usefulness of job task measures at the person level.

Spitz-Oener (2006) applies the task approach to Germany and therefore uses the BIBB/IAB- and BIBB/BAuA cross-sections. Spitz-Oener does not use any statistical method of classification, but she selects items from the first four waves to classify job tasks into the five domains proposed in the ALM paper.[7] This author's task index measures the share (between zero and hundred) of tasks a worker performs among all activities (tasks) in one domain. Based on this operationalization, and the aggregation of task shares at the occupation-level, Spitz-Oener (ibid.) finds evidence for changes in job tasks and skill demand in West Germany, similar to ALM's findings for U.S. data. This author demonstrates that the shift toward analytical and interactive activities and away from cognitive and manual routine tasks accompanies changes in skill demand and employment polarization.

Following this researcher's approach,[8] many important subsequent contributions to the debate on task-biased technological change in Germany have been published (Spitz-Oener 2008; Antonczyk et al. 2009; Black and Spitz-Oener 2010; Gathmann and Schoenberg 2010; Geel and Backes Gellner 2011; Geel et al. 2011). Among these studies, Antonczyk et al. (2009) make simultaneous use of the 2006 and the 1999 surveys. To operationalize the task domains, these researchers follow Spitz-Oener (2006) "as closely as possible" and aggregate the 14 different tasks reported in the data to five task measures.[9] Due to the different task items available for each survey year, these authors' measures vary, particularly in the domains of routine and non-routine manual tasks. In addition to the task index derived from Spitz-Oener, these researchers report results for a job complexity task index that measures the share of reported tasks in one domain among all tasks reported by the worker.

Interestingly, for the Spitz-Oener index, the findings for 1999 to 2006 deviate strongly from Spitz-Oeners findings for the period from 1979 to 1998/99. Instead of an increase in all types of non-routine tasks and a decrease of manual

[4]These single job tasks measuring the three domains are (cp. Autor and Handel 2009, p. 13): "cognitive job demands: (1) the length of longest document typically read as part of the job (ranging from one page or less to more than 25 pages); (2) frequency of mathematics tasks involving high-school or higher mathematics (algebra, geometry, trigonometry, probability/statistics, or calculus); (3) frequency of problem solving tasks requiring at least 30 minutes to find a good solution; and (4) proportion of work day managing or supervising other workers. (...) interpersonal job demands: (1) interactions with customers or clients; (2) interactions with suppliers or contractors; and (3) interactions with students or trainees (...) physical and routine job tasks: (1) proportion of the work day spent performing physical tasks such as standing, operating machinery or vehicles, or making or fixing things by hand; and (2) proportion of the work day spent performing short, repetitive tasks."

[5]Green (2012) studies changes in job skills using task data from the U.K. Skill Survey of 1997, 2001, and 2006 and the 1992 Employment Survey in Britain. He points out that in the empirical testing of the ALM model there is some potential for misclassification and that the classification of tasks as either routine or nonroutine is especially problematic.

[6]We thank one of the anonymous reviewers for this hint.

[7]The single tasks are (Spitz-Oener 2006, p. 243): Non-routine analytic: "Researching, analyzing, evaluating and planning, making plans/constructions, designing, sketching, working out rules/prescriptions, and using and interpreting rules"; Non-routine interactive: "Negotiating, lobbying, coordinating, organizing, teaching or training, selling, buying, advising customers, advertising, entertaining or presenting, and employing or managing personnel"; Routine cognitive: "Calculating, bookkeeping, correcting texts/data, and measuring length/weight/temperature"; Routine manual: "Operating or controlling machines and equipping machines"; Non-routine manual: "Repairing or renovating houses/apartments/machines/vehicles, restoring art/monuments, and serving or accommodating".

[8]As for the O*NET data, researchers applied factor or principal component analysis to construct highly reliable, parsimonious indicators for "cognitive", "manual" or "interactive task" out of the single task items in the German database. This strategy was applied to single waves, and authors used a wider range of skill (requirement) items in addition to the task questions (e.g. Nedelkoska and Neffke 2010; Bublitz and Noseleit 2012 with the 2006 data).

[9]Non-routine analytic: developing, researching, designing and gathering information, investigating, documenting; Non-routine interactive: informing, advising and training, teaching, tutoring, educating and organizing, planning/preparing working processes and promoting, marketing, public relations and buying, providing, selling and to be supervisor; Routine cognitive: measuring, controlling, quality checks; Routine manual: fabricating, producing goods and supervising, controlling machines and transporting, stocking, posting; Non-routine manual: repairing, patching and nursing, serving, healing (Antonczyk et al. 2009, p. 27).

and cognitive routine tasks, the authors find a large decrease in non-routine manual tasks, an increase in routine cognitive tasks, and no change in the level of non-routine interactive tasks. The authors conclude: "This discovered opposing trend in the development of the routine cognitive and non-routine manual task domains is very notable." (ibid., p. 9).

To sum up, the literature review shows that both the data itself and researchers' methods to classify task items into distinct domains vary substantially. To some extent this is a natural consequence of different research questions (occupational or workplace studies). However, especially with regards to the domain of routine tasks, operationalizations and (interpretations of) results differ. Though this domain is theoretically well defined, it is very hard to identify codifiable routine tasks in whatever datasets on ex ante grounds (see also Green 2012, p. 41). Altogether, this should sensitize us to how the operationalization affects how we interpret task change. Our hypothesis is that this especially applies to the German case.

The next chapter is devoted to describing the methodologies and general content of the German task data with particular emphasis on task items. We systematically discuss the idiosyncrasies of the German data's task measures and thereby demonstrate that difficulties as regards operationalization might be enforced by the fact that the data are not readily comparable between survey years.

3 Job task measures in the German BIBB/IAB and BIBB/BAuA Surveys from 1979–2006

3.1 Data collection and survey content

The BIBB/IAB- and BIBB/BAuA employment cross-sections from 1979, 1985/86, 1991/92 and 1998/99 and 2005/2006 are worker-level probability samples and include information on employees' qualification and career history, as well as detailed job-related information (e.g., organizational information, job tasks, job skill requirements, working conditions, health). With large sample sizes of 20,000 to over 30,000, the database was designed for research on special social groups (such as old-age, female, non-formally qualified workers, workers with different national backgrounds) and structural developments within specific occupations, industries and vocational fields.

Over the years of survey implementation, there have been changes to the study's population, modes of data collection, and content. Table 1 gives a brief overview of the differences in population and sample size. Up to 1998/99, the survey collected information from a 0.1 percent sample of Germany's workforce population (thus being a "small Mikrozensus" (Parmentier and Dostal 2002, p. 39)). Across all years of survey data, the concept of core-employment ("Kernerwerbstaetigkeit") is applied to the working population: employees and workers aged 15 or older working at least 10 hours per week are included, as well as individuals attending a qualification scheme who are also in a paid employment program (e.g., students or apprentices with side jobs of at least 10 hours/week) and helping family members or other employees whose employment is interrupted for a maximum of three months. In contrast, employees whose employment constitutes a compulsory part of their vocational training (e.g., an apprenticeship training in the dual system) are not included in the survey data. Similarly, the dataset does not include information on "paid voluntary work" (i.e. voluntary work one gets expense allowances for), employees in compulsory military work, community service, or voluntary work in the social or environmental sector. Beginning with the 1991/92 survey, the dataset includes the resident population of East Germany and workers with qualifications from the former German Democratic Republic. In 1991/92, unemployed individuals were also included in the study-population (Rohrbach 2009; Zopf and Tiemann 2010).

Data collection methods also changed over different years of the survey: the mode changed from face-to-face personal interviews with standardized Paper-and-Pencil questionnaires (1979–1991/92) to Computer-Aided Personal Interviews (1998/99) to Computer-Aided Telephone Interviews (2005/06). The questionnaire itself was changed accordingly.

To date, no cumulative, synchronized data file of all five cross-sections exists. Yet, a first synopsis of selected variables of the first four cross-sections (Hartmann 2000) could be utilized. Based on a program for merging comparable samples of respondents and some core variables (including weights) over the first four waves, we added selected variables (and cases) from the cross-section 2005/06. We thus reduce the study population to comparable cases and so narrow the scope for effects resulting from different samples.

The final dataset for the analyses presented in this paper includes information collected over the period from 1979 to 2006 and, for each survey year, comprises representative samples of employees from the western part of Germany, aged 15 to 65, who belonged to the labor force (defined as having a paid employment situation) with a regular working time of at least ten hours per week. The total sample size is 118,105 (see Table 1).

3.2 Task and skill items

The German database differs from other datasets used in the task approach literature in several ways. Table 2 summarizes important characteristics of the German task data as compared to other datasets discussed or mentioned in this paper.

Table 1 Population and sample sizes for the cross-sections

Year	Population and sample sizes	
1979	German labor force (not included: apprentices and other people in qualification schemes, armed forces, foreign workers) age 15–65	
	active	28,828
	unemployed	992
	Comparable sample	*28,088*
1985/86	German active labour force age 15–65	26,361
	Comparable sampe	*25,933*
1991/92	West-German active labor force	
	German	23,476
	German speaking foreign workers	614
	Comparable sample	*22,900*
	New Laender (east Germany)	
	active	7,851
	unemployed	1,880
	people in retraining schemes	456
1998/99	German active labor force (including German speaking foreign workers) age 15 and over, regular employment of 10 hrs/week	34,343
	Comparable sample	*25,513*
2005/06	German active labor force (including German speaking foreign workers) age 15 and over, regular employment of 10 hrs/week	20,000
	Comparable sample	*15,671*
Total	Population 1979–2006	144,801
	Comparable sample 1979–2006	*118,105*

Notes: Comparable Sample comprises West-German active labor force (defined as having a paid work) aged 15 to 65, with a regular working time of at least ten hours per week

In the German data the individual worker represents both the survey response unit and the unit of analysis (and thus allows for the study of within occupational heterogeneity). The data include self-reported job tasks for workers' main jobs at the worker level. Broadly speaking, in all survey years, respondents are asked to select which of the listed tasks correspond to their jobs.

This "subjective method" differs from "objective" measures in the DOT and O*NET data, which (also) rely on expert ratings for job titles in an occupational classification. The primary advantage of objective measures is that they are based on clear definitions and detailed measurement instructions. However, the person-level measure captures within occupational heterogeneity in job tasks and thus avoids aggregation bias. Moreover, in the literature, the subjective method is said to correspond better with real job requirements and is said to be more reliable and valid (McGuinness 2006 and Hartog 2000).[10]

With regards to over-time comparisons, both methods suffer from comparability problems. E.g. in the fourth edition of the DOT around two-thirds of ratings of job titles were taken over from earlier editions, so that change might be underestimated (Spenner 1990). Cain and Treiman (1981, p. 273) even come to the conclusion that "the DOT cannot legitimately be used to study changes in the job content of the economy over time." With both methods it is unclear to what extent results on trends are artifacts since the 1970s

[10]McGuinness (2006) and Hartog (2000) give useful overviews of advantages and drawbacks of both methods with regard to required schooling levels and overeducation. For a discussion of some methodological properties of the DOT measures as compared to self-report measures see Spenner (1990) and Cain and Treiman (1981).

Table 2 Selected characteristics of task datasets used in the task approach literature

DATASET	BIBB/IAB-BIBB/BAuA Surveys	PDII Survey	STAMP Survey	UK Skill Surveys	O*NET	DOT
Name	The BIBB/IAB- and BIBB- BAuA Surveys of the Working Population on Qualification and Working Conditions in Germany	Survey of the Princeton Data Improvement Initiative (PDII)	Survey of Skills, Technology, and Management Practices (STAMP)	UK Skill Surveys 1997, 2001, 2006 (Employment in Britain 1992; Social Change and Economic Life Initiative, 1986)	U.S. Department of Labor's Occupational Information Network	U.S. Department of Labor's Dictionary of Occupational Titles
Dates of data collection	1979, 1985/86, 1991/92, 1998/99, 2005/06, 2011/2012	2008	2006–2008	1986, 1992, 1997, 2001, 2006, 2011/2012	Ongoing	1949, 1965, 1977, 1991
Population	German (active) labor force	U.S. labor force	U.S. labor force	British (active) labor force	Standard Occ. Class. (SOC)-defined O*NET occupations	U.S. labor force[a]
Unit of Analysis	Worker	Worker	Worker	Worker	Occupation	Job/Occupation
Response Unit	Job incumbents	Job incumbents	Job incumbents	Job incumbents	Job incumbents, occupational experts, and occupational analysts	Occupation expert
Number of cases	Around 20,000 to 30,000 each	2,513	2,304	Around 4,000 to 5,000 each	800	12,000 (around 450 3-digit Census Occ. Code (OCC) Occupations)
Data collection method	1979–1991/92: Face-to-face PAPI, 1998/99: Face-to-face CAPI, 2005/06, 2011/2012: CATI	CATI	CATI	Face-to-face	PAPI, online	On-site observations (and official job descriptions)
Method of task measurement	Subjective method: self-report ratings of job incumbents	Subjective method: self-report ratings of job incumbents	Subjective method: self-report ratings of job incumbents	Subjective method: self-report ratings of job incumbents	Objective and subjective method: expert ratings of jobs /occupation and self-reports of job incumbents	Objective method: expert ratings of jobs /occupation
Detailed study information (and data access)	Zopf and Tiemann 2010, Rohrbach 2009; BIBB-FDZ 2012	Autor and Handel 2009; Princeton Data Improvement Initiative 2012	Handel 2008	Felstead et al. 2007; Felstead et al. 2002; Cardiff School of Social Sciences 2012	Handel 2009; Research Triangle Institute 2012 National Center for O*NET Development 2012	Spenner 1990, Cain and Treiman 1981; Coutsoukis and Information Technology Associates 2011

Notes: [a]Cain and Treiman (1981) report an undercoverage of service jobs in the 4th edition and sex bias in the 3rd edition

cohort of respondents and raters interprets (the meaning of items) differently from those in 2006.

The tasks surveyed in the German database do not represent multiple task items for larger factors or constructs, such as in DOT or O*NET. There are no items with an anchored rating scale response format[11] to reflect the complexity level of the task item (for example, ALM use the DOT "GED Math" measure as a proxy for analytic tasks).

Finally, the task items are not based on what Handel calls an "explicit scaling approach"[12] (Handel 2008, p. 12–13), such as in the survey of Skills, Technology, and Management Practices (STAMP, see Handel 2009, and Table 2).

Thus, the database's task measures differ substantially from the DOT and O*NET data also with regards to their rating scale response format. It is, however, not easy to assess consequences of these differences for the study of task change; while the more advanced response formats might generally raise the reliability of measures, their validity, i.e. the degree to which the selected task items survey what they are supposed to measure, to the best of our knowledge has not been compared systematically so far.

In addition to its striking differences to other measures in the task approach literature, the BIBB/IAB-BIBB/BAUA task measures themselves differ considerably between survey rounds: (a) the wording of questions, (b) the response categories, (c) the number of task items surveyed, and (d) the content and wording of the surveyed task items have changed. The wording of questions changed as the mode of collection changed.[13] That is, in 1979, respondents were asked to mark with a cross "all tasks that accrued recently in your job", in 1985/86 and 1991/1992, respondents were asked to select from a list handed to the respondent "all tasks that belong to your job," and in 1998/99 and 2006, respondents were asked to indicate whether and how frequently the tasks read in random order by the interviewer "appear at your job". Thus, from 1979 to 1991/92 it was surveyed whether a task was performed or not;[14] from 1998/99 on, respondents were asked to indicate the frequency of tasks they perform ("frequently", "sometimes" (2006) / "seldom" (1998/99)[15] or "never"[16]).

To illustrate (c), in 1979, there were a maximum of 121 single task items, whereas by 1999, there were only 13 single task items available to respondents. An analysis of the content (and number) of changes in task items surveyed would lead researchers to conclude the importance of non-routine analytic and interactive tasks increased over time. To use a concrete example, the single task item of "manufacturing goods" in 2006 was represented by more than 30 single task items, such as "spinning, weaving, knitting," or "tanning, conserving" in 1979. On the contrary, until the data collection in 2005/2006, the number of non-routine cognitive tasks increased disproportionally.

As a consequence of the large reduction in items, principally there is a much higher probability that respondents mark one item in 1979 than in 1998/99. However, mostly the decline in items is not to the disadvantage of the scope of tasks covered. Rather, as illustrated by "manufacturing" above, more disaggregated job tasks in earlier waves are aggregated into broader groups of tasks in later ones.[17]

Naturally, the task items in the dataset were not originally intended to measure the ALM domains of "routine" and "non-routine" tasks. The employment surveys were intended to gather information not available in the German Mikrozensus (see Parmentier and Dostal 2002 for a comprehensive overview). Regarding questions on specific tasks there is no reference to a single theoretical framework. Tasks

[11]This technique uses concrete examples for selected response categories of a scale to give respondents common reference points and help them selecting an answer (Handel 2009, p. 12). For instance, the O*Net questionnaire uses a 7-point scale for number facility that is anchored by three behavioural examples: Add 2 and 7 (level 1), Balance a checkbook (level 3) and Compute the interest payment that should be generated from an investment (level 5; see ibid.).

[12]This includes questions and response options that are "objective, specific, correspond directly to researchers' objects of interest, and have absolute meanings for respondents. (...) Questions are phrased in terms of facts, events, and behaviors, rather than attitudes, evaluations, and holistic judgments." (Handel 2008, p. 12–13). Handel expects them to have lower measurement error because the scope for respondents' subjective interpretations of the meaning of items is as low as possible.

[13]To the best of our knowledge, there is no study on effects of changing data collection modes on task composition. From the literature on method changes in general (Roberts 2007; Jäckle et al. 2008), we do not expect relevant differences in the answering behaviour (measurement error); job tasks as surveyed here are not sensitive questions and thus should not provoke social desirability. All five rounds are probability samples of the German active labor force and we use comparable samples of respondents, i.e. we do not expect serious effects due to differences between respondents attracted by different modes (non-response error) or coverage error. For the first time, the 2011/12 survey includes a mode comparability study so that mode effects on selected variables can be systematically studied in the near future.

[14]In 1985/86 there is a consecutive question about which task is performed predominantly (question 13B). It was only asked when more than one task were carried out (as stated in question 13A). Data on question 13B was not used for the described analyses.

[15]The original synopses of the cross-sections conflates frequencies of "sometimes" or "seldom" with "never". Especially in 1998/99 this was sensible, "seldom" in German is less often than "sometimes" and almost "never". Additionally the frequencies of appearance of single tasks seem to have been taken into account. With "sometimes" added to "frequently", numbers would have grown substantially. When adding the 2005/06 data we perpetuated the original operationalization.

[16]It is obvious that ordinal and nominal scales have large disadvantages when computing additive scales. Because the relevant task domains are unevenly covered by task items, the indices regularly vary in their measures of central tendencies and variance purely by construction (see Sect. 3.3).

[17]Results discussed later in this paper however show that there is an increase in the number of tasks performed over rounds which we interpret as an increase in overall job complexity.

were first surveyed in the German Mikrozensus of 1969 (Statistisches Bundesamt 1971). As far as we can assess, the task items then were based on a translation of the "Work Performed Manual" (Statistisches Bundesamt 1967). Preiser (1979, pp. 4 f.) describes the discussion between Federal Statistical Agency and Federal Labour Office that lead to the surveying of tasks. There seem to have been two main reasons: to be able to describe occupations, supplementing and possibly improving occupational classifications (occupational classifications were seen to come to their limits as occupations evolved) and to gather information for the description of occupational positions used in the employment services. Being able to describe the structure of tasks and its change over time was also part of the discussion. Given this, these questions cannot be said to be based on theory. Rather, the first inclusion of task item questions in 1979 dates back to what had been done and discussed in the late 60's and 70's. The changes in the number and content of questions signals that survey researchers have tried to adapt the measurement instrument to structural changes in the workforce (which is stated explicitely for the items of cross-section 1998/99 in Parmentier and Dostal 2002, p. 37f.).

Changes in the content and wording of single items take place gradually over the cross-sections. In many cases, a group of related tasks are condensed into fewer items in the next wave, or a small number of other tasks were added or removed from the task item. For instance, in 1979, there were nine single task items for "overseeing, operating plant and machines". These items were subsumed into two task items in 1985/86 and finally into one task item in 1998/99.[18] On the contrary, some few items are split up and confounded with new task items; as a result, these tasks cannot be observed separately over the years.

One strategy to handle different task item measures would be to classify items from each survey wave, even if they appear only once, into the five ALM domains wave for wave. Then, in principle, one domain could include different items for each survey wave. However, this strategy makes it difficult to differentiate between substantial changes and changes that result simply from changing measures. Also, it is possible that the interpretation of single task items changed over time. In a principal component analysis of the task items it became clear that over time some of the items had changed their meaning for the surveyed persons.[19] Therefore, we identified comparable task items in each wave for which survey data exists for at least two of the five cross-sections having interpretations as clear as possible. We thereby assume maximal comparability to earlier waves, as well as comparability between 1998/99 and 2006, when we count "frequently" performed tasks. As a result, we end up with 22 task variables (Table 3). Table 3 includes variable names[20] (col. 2), the single task items combined in each variable (col. 3), and their percentages in the study population in each survey round (col. 4). In the following, when using the variable names we refer to all task items combined in each variable.

The 22 items show some interesting developments over time. For example, only for manufacturing can one see a decline in shares. However, even this decline follows a more or less u-shaped curve and seems to be ascending again. Looking at a long timespan ranging from 1979 to 2006, for certain tasks, we could have expected more distinct developments or clear increasing or decreasing trends. There are also some changes that seem implausible (like the share for *consulting* rising from 17 to 54 percent between 1991/92 and 1998/99 or for *storing* rising from 18 to 46 percent between 1985/86 and 1991/92). Most of these reflect societal discussions like the one on the impact of services in the 1990s or they reflect organisational changes across industries at the end of the 1980s.

However, it appears that almost all single tasks are increasingly performed by growing numbers of employees. Of course, this property may be a feature of the questionnaire because the number of tasks for respondents to choose from steadily declined from over 120 in the first survey to less than 20 in the latest survey. Still, according to the task approach, tasks such as *operating* or *repairing* should have followed a more distinctive pattern of decline because they are (at least partially) programmable.

For non-routine manual tasks (such as *caring*, *cleaning* and *protecting*), the trends follow patterns of increasing shares, as we would have expected under the task approach. Additionally, *programming*—even though this does not yet represent high-level computer-usage—gains shares. *Organizing* and *consulting* show similar growth. The other tasks either keep constant shares, show no trend, or move in larger shifts around the mean. All in all, the composition of single tasks shows that insofar as tasks represent the components of jobs, one could argue for a growth in complexity of jobs as far as the number of tasks performed are concerned.[21] While this is only a weak hint on growing complexity of jobs, also content analysis of jobs point in the same direction (cf. Tiemann 2012). Of the tasks that could be categorized as routine, only *manufacturing* loses shares.

[18] However, even if the content did not changed much, it is not clear whether respondents from later rounds interpret the items' meaning in the same way as respondents from earlier ones.

[19] This principal component analysis covered only task items that were surveyed comparably in all cross-sections. Here some items changed their correlations to the first three principal components and all other items over time. Their relative positions changed which can be interpreted as changes in their attributed meaning. Results upon request.

[20] Naturally, these are not the original variable names. As names we chose one characteristic task item of those combined in the variable.

[21] Antonczyk et al. (2009) give a short overview on quantitative job-complexity measures.

Table 3 Variable names, variable content and their percentages 1979–2006

Variable number and names		Variable content	1979	1985/86	1991/92	1998/99	2005/06
1	*investigating*	Gathering information, investigating, documenting	N.av.	N.av.	N.av.	28.8	49.6
2	*organizing*	Organizing, making plans, working out operations, decis. making	11.5	15.8	27.1	37.6	36.1
3	*researching*	Researching, evaluating, developing, constructing	10.1	19.7	18.0	N.av.	11.1
4	*programming*	Working with computers, programming	5.1	9.8	19.5	N.av.	64.1
5	*applyinglaw*	Applying law, notarizing	4.5	12.7	11.5	N.av.	N.av.
6	*teaching*	Teaching, training, educating	7.0	17.0	14.8	17.6	21.0
7	*consulting*	Consulting, advising	21.1	18.8	17.2	53.6	58.2
8	*buying*	Purchasing, procuring, selling	21.1	27.7	24.6	25.3	22.9
9	*promoting*	Promoting, marketing, public relations	2.7	8.4	17.4	11.9	12.0
10	*managing*	Managing personnel, leading, employing	14.4	20.8	16.2	N.av.	N.av.
11	*negotiaing*	Negotiating, lobbying	5.1	N.av.	N.av.	23.1	N.av.
12	*repairing*	Repairing, renovating, restoring	11.8	22.8	16.9	17.0	17.9
13	*accomodating*	Accommodating, preparing food, serving, (entertaining)	4.9	3.4	3.8	N.av.	8.1
14	*caring*	Taking care, healing	3.9	6.6	7.0	7.3	8.1
15	*cleaning*	Cleaning, recycling, waste disposal	8.1	7.0	12.0	N.av.	25.6
16	*protecting*	Protecting, guarding, observing, controlling traffic	7.1	4.3	4.0	N.av.	21.0
17	*measuring*	Measuring, testing, quality control	18.9	N.av.	N.av.	33.9	43.9
18	*writing*	Clerical work	34.3	37.2	37.1	N.av.	N.av.
19	*calculating*	Calculating, bookkeeping	20.5	19.9	20.6	N.av.	N.av.
20	*operating*	Operating, controlling machines	10.5	24.3	23.5	22.3	25.2
21	*manufacturing*	Manufacturing of goods, planting	34.0	20.4	16.7	15.3	17.7
22	*storing*	Storing, transporting, shipping, stocking, posting	23.2	18.1	46.3	N.av.	25.7

Notes: N.av. (not available) marks survey years in which the respective variable has not been surveyed. Sample is restricted to people from the western part of Germany belonging to the labor force (defined as having a paid work) aged 15 to 65 with a regular working time of at least ten hours per week. Source: BIBB/IAB—BIBB/BAuA Employment Surveys 1979–2006, weighted values, own calculations

Furthermore, the database includes other skill items surveyed in all cross-sections that had fairly comparable wording and content of questions. In our analyses, we use comparable task items for job conditions ("Arbeitsbedingungen") and job requirements ("Arbeitsanforderungen") and recode items into a common metric, i.e., a dichotomous scale. For the sensitivity analyses, we also use education variables whose measures are highly consistent over the five waves. We recode degrees according to the International Classification of Education (ISCED 97).

3.3 Methods

In the previous sections, we argue that findings on task change based on the German database might be sensitive to variable choice. As demonstrated in the first part of this section, the German data differ substantially from their counterparts, and the data have changed considerably over the years. Altogether, the process of classifying single tasks into the widely used ALM scheme is not straightforward.

To test our hypothesis, we vary our method of pooling tasks into domains (Sect. 4) and compare the values for task change over time (Sect. 5).

The question, which task items should be pooled into the ALM domains can be understood as a problem of evaluating the quality of multiple measures of the same concept, or, even more general, as a problem of the quality of survey measures. Corresponding to the three quality criteria objectivity, reliability and validity, we use three methods to solve the problem: The first, and nonstatistical method is to select items by reviewing the literature, i.e. pool items in the same way as other authors have done before. This "research practice approach" simply asserts the belonging of items to domains by theory; no empirical operations accompany the operationalization.

Our second method is factor analysis (FA). FA is a standard multivariate technique that is used for identifying the dimensional structure of a number of variables, the construction of scales, for data reduction and the orthogonalisation of variables (Gorsuch 1983; cit. by Wolff and Bacher 2010, p. 355). We use this technique, first, to analyse to what extent the task variables in our data are dimensionally structured by factors that can be interpreted as the ALM domains. Second, we pool tasks with factor loadings above .5 and a low uniqueness into the five ALM task domains. FA requires a correlation or covariance matrix. However, our task items are nominally scaled (respondent's are asked whether or not the tasks appear at their job). We follow Kolenikov and Angeles (2004) in their advise to perform a factor analysis from a polychoric correlations matrix instead of from the nominal variables in the dataset.[22] To find the numbers of factors to retain, we conduct a parallel analysis (O'Connor 2000). From the results (on request) we follow that five factors (with eigenvalues greater 1 (Kaiser 1958)) should be extracted. The model equation of the FA is

$$y_{ij} = z_{i1}b_{1j} + z_{i2}b_{2j} + \cdots + z_{iq}b_{qj} + e_{ij} \quad (1)$$

where y_{ij} is the value of the ith observation on the jth variable, z_{ik} is the ith observation on the kth common factor, b_{kj} is the set of linear coefficients called factor loadings, and e_{ij} is the jth variable's unique factor (StataCorp 2007, p. 288). To improve the interpretation of results, we then applied the default orthogonal varimax rotation without normalization, i.e. Kaiser off (Kaiser 1958, see also Gorsuch 1983). We validate the pooling of items into domains as suggested by the FA results by means of a reliability analysis (using Cronbach's α).

Objective and reliable measures must not necessarily be valid measures. Therefore, in a third step we analyze the criterion validity of task items to pool only tasks into the ALM domains that are of high validity (see Saris and Gallhofer 2007, Chaps. 9 and 10). The criterion validity describes the degree of correlation of a measure with an external criterion. The criterion (predicted) variables used for this validation are external measures for manual, routine and computer tasks. We regress the criterion variables on single task items and evaluate the explanatory power and the signs (−/+) of the linear and (in the case of computer) the logistic association.

Our criterion variables are based on the comparable items for specific job conditions (criterion variable *manual*) and job requirements (criterion variables *routine* and *computer*). Both, job conditions and requirements, refer to what people are required to do at their workplaces. We compute a variable, *manual*, that aims to measure whether the current job is characterized by manual (or cognitive) skill requirements, a variable *routine* that reflects the degree of routine skill requirements, and a variable *computer* that measures the degree of computer use in the job. We also attempted to find a measure for interactive tasks; however, there is no comparable information available in all survey waves. *Manual* is an additive index of ten items indicating whether the respondent has to "work standing", "carry heavy stock" and other job conditions (see Table 8 in the Appendix). *Routine* is an additive index that scores high if the current job is "stipulated in the minutest details" and "one and the same work cycle / process is repeating in the minutest details" and if respondents disagree that they "are confronted with new problems that remain to be understood / familiarized with" and disagree that their "tasks include process optimization

[22] We use the *tetrachoric* command and the factormat command in Stata. The Kolenikov and Angeles paper not only examines how discrete data can be used in this framework but also gives an overview of factor and principal component analysis and references to relevant literature in economics (ibid., p. 3).

or trying out new things". Alternative versions of *manual* and *routine* based on selection of other response questions lead to similar results (results upon request). *Computer* simply measures whether the job requires skills in standard computer software.[23] Consistent with the ALM-hypothesis, *manual* and *routine* decrease and *computer* increases from 1979 to 2006.

In the fourth section, we will show that the resulting operationalizations differ between the three methods, which is a central finding of this exercise.

The fifth section presents a sensitivity analysis of whether they all lead to similar conclusions on task change in Germany or rather to what extent the results depend on the operationalization.

To this end, we apply the three operationalizations to the data using the formula for the computation of task indicators as it was introduced by Spitz-Oener (2006):[24]

$$Task_{ijt} = \sum(T_{ijt}) / \sum(T_{jt}) * 100 \quad (2)$$

which is the percentage of task activities in domain j performed by i in cross section t of all task activities in domain j at time t where $t = 1979$, 1984/85, 1991/92, 1998/99, and 2005/06, and $j =$ nonroutine analytic tasks, nonroutine interactive tasks, routine cognitive tasks, routine manual tasks and nonroutine manual tasks. In addition to this additive index, we calculate a non-additive version which equals 1 if the respondent performs at least one of the tasks within one domain j. The reason is that the relevant task domains are unevenly covered by task items, so the indices regularly vary in their central tendencies and variances purely by construction. We choose these simple indices for several reasons. First, this approach follows methods in the available literature. Second, we aim to analyze changes in job tasks as "purely" as possible; we prefer simple computation of job complexity measures, rather than more advanced methods (for example, factor analysis) because we want to avoid detrending our results by use of an artificially complex index.

For each of the three operationalizations we calculate the additive and non-additive indices' percentages for ALM's task domains over the period 1979 to 2006 based on the above discussed comparable sample of West German workers aged 15 to 65.

[23] Alternatively, *computer* could be measured by work tools used (see Spitz-Oener 2006).

[24] This index is used exemplary here. Other indices of this sort are being calculated differently (like the one of Antonczyk et al. 2009) but mostly get similar results (e.g. Antonczyk et al. (2009, p. 223) find their own index "indicates similar changes" to the Spitz-Oener index although there are partial differences. It has to be noted that an index like the one of Antonczyk et al. (2009, p. 222), summing up to one for each individual leads to the obligation of leaving one of the index variables out as a covariate in a regression analyses).

4 Classifying tasks into "task-approach" domains

4.1 Classification based on existing theory and research practice

By examining the 22 items, one can identify tasks with notably similar wording to items in the DOT or O*Net databases. Examples include "operating machines" (DOT) and "managing and supervising" (O*NET, Sect. 2). In these cases, we classify them according to current research practice. However, we cannot assess whether these German measures are functionally equivalent to the U.S. measures. For most of the measures in the German database, there is no direct counterpart in the U.S. data. In these cases, the researcher must determine the domain. The potential for misclassification is rather small with respect to the manual versus cognitive dimension. In contrast, the identification of routine, programmable tasks within the two dimensions is especially challenging.

Reviewing Spitz-Oeners' classification, Green notes that "calculating" might be misclassified as a cognitive routine task, given that the GED Math measure for non-routine cognitive tasks in the ALM paper includes the items "adds and subtracts 2-digit numbers". Likewise, he notes that "selling" is a partly automated activity (e.g., Internet sales) and thus could alternatively be classified as a cognitive routine task. In addition, some items seem to include both routine and non-routine tasks. For instance, "measuring" includes measuring, testing, and quality control tasks. Whereas "measuring" can be considered a manual routine task (and the similar—though more specific—item "measures dimensions of bottle," is classified as such by ALM), "testing" and especially "quality control" might also include non-routine job activities. As a rule, in these ambiguous cases, i.e., in the case of *measuring*, *writing*, *calculating*, *operating*, *manufacturing*, *storing*, *repairing*, *accommodating*, *caring*, *cleaning*, and *protecting*, we follow the classifications in the literature (i.e., Spitz-Oener 2006; Gathmann and Schoenberg 2010, and Antonczyk et al. 2009). The resulting classification is shown in Table 4, column 2.

4.2 Statistical classification

Table 5 shows the rotated factor loadings (using the varimax procedure) of task items on the five extracted factors and their uniqueness. For each variable, the highest loading (usually over .5, only for *protecting* at .43) is highlighted.

The factors account for 67.51 percent of the common variance in the tetrachoric correlation matrix. Looking at the uniqueness for each factor, it becomes clear that this factor solution does not yield satisfying results for certain task items. The first factor, labeled "non-routine-analytic", explains 18.95 of the total shared variance and consists of

Table 4 Task items and their operationalization by the three classification methods

Variable number and name		Task categorization		
		Theory/research practice	FA/reliability analysis	Criterion validation
1	*investigating*	analytic	analytic	non-rout. cognitive
2	*organizing*	analytic	analytic	non-rout. cognitive
3	*researching*	analytic	analytic	non-rout. cognitive
4	*programming*	analytic	routine cognitive	non-rout. cognitive
5	*applyinglaw*	analytic	routine cognitive	non-rout. cognitive
6	*teaching*	interactive	analytic	non-rout. cognitive
7	*consulting*	interactive	analytic	non-rout. cognitive
8	*buying*	interactive	interactive	non-rout. cognitive
9	*promoting*	interactive	interactive	non-rout. cognitive
10	*managing*	interactive	analytic	non-rout. cognitive
11	*negotiaing*	interactive	interactive	non-rout. cognitive
12	*repairing*	non-rout. manual	routine manual	non-rout. cognitive
13	*accomodating*	non-rout. manual	non-routine manual	non-rout. cognitive
14	*caring*	non-rout. manual	non-routine manual	non-rout. cognitive
15	*cleaning*	non-rout. manual	non-routine manual	routine manual
16	*protecting*	non-rout. manual	routine manual	non-routine manual
17	*measuring*	routine cognitive	routine manual	non-rout. cognitive
18	*writing*	routine cognitive	routine cognitive	non-rout. cognitive
19	*calculating*	routine cognitive	routine cognitive	non-rout. cognitive
20	*operating*	routine manual	routine manual	routine manual
21	*manufacturing*	routine manual	routine manual	routine manual
22	*storing*	routine manual	routine cognitive	routine manual

6 items. The factor has relatively high loadings of *organizing*, *investigating*, *researching* and also *managing*, while *teaching* and *consulting* show the highest loadings. Factor 3 (13.20 percent of total var.), labeled "non-routine interactive", is defined by *buying*, *negotiating* and *promoting*. Besides *investigating* and *organizing*, *storing* and *calculating* also show high correlations here. This phenomenon leads to confounding between the two non-routine cognitive factors and also makes it possible to interpret Factor 1 as "high-qualification tasks" and Factor 3 as "commercial" or "mercantile" tasks. Factor 2, here labeled "routine-cognitive", holds high loadings of *programming, writing, calculating* and *applyinglaw*. While it seems sensible for *manufacturing* to have a strongly negative correlation, it is not clear how we should interpret the high positive loading of *storing* on this cognitive factor. Factors 4 and 5 are both described as manual tasks; while the former is regarded as routine, the latter is viewed as non-routine. On factor 4 *measuring, operating, repairing* and *manufacturing* show the highest loadings, on factor 5 these are *accomodating, caring* and *cleaning*. Tasks such as *investigating, programming, writing* and *calculating* show negative loadings on both manual factors. Why this does not hold for *researching, consulting* and *buying* raises further questions, just as the negative loading of *manufacturing* with non-routine-manual also bears further investigation. A close examination of the factor loadings matrix raises a number of questions about what the task items really measure. The implicit assumption is that what is really measured and extracted here is a broader view of shaping or structuring the form of occupations and in that way reflecting structural changes in the occupational organization of the German labour market (Tiemann 2012). Inspecting the reliability of indices based on this factor solution yields values for Cronbach's α between .2 and .7 with means over cross-sections of around .5 for factors 2 to 5, and of .65 for the non-routine analytic factor. So, except for this scale the values for internal consistency are below social science standards, indicating that the internal consistency of the indices built around the factor analysis is rather low. Still, we compute indices for the five factors with the highest loading task items (highlighted).

The resulting operationalization differs from the operationalization of task domains by existing theory and research practice (see col. 3 in Table 4).

4.3 Criterion validation of task items

One result of the sections above is that classifying the single task items into the distinct task domains proposed by ALM

Table 5 Rotated factor loadings and unique variances. Highest loadings of each variable highlighted

Variable number and name		Factor 1 "non-routine-analytic"	Factor 2 "routine-cognitive"	Factor 3 "non-routine-interactive"	Factor 4 "routine-manual"	Factor 5 "non-routine-manual"	U[a]
6	*teaching*	.76	.13	−.09	−.01	.24	.34
7	*consulting*	.76	.16	.23	−.09	.21	.29
17	*measuring*	.26	.35	.02	.73	.21	.23
20	*operating*	−.00	−.09	−.11	.77	−.02	.39
12	*repairing*	−.08	−.28	−.01	.77	−.12	.31
8	*buying*	.03	.17	.80	−.05	.14	.31
11	*negotiating*	.61	.22	.63	.04	.00	.18
9	*promoting*	.28	.06	.85	−.07	.00	.20
1	*investigating*	.73	.01	.32	−.09	−.17	.33
2	*organizing*	.67	.30	.37	.12	.09	.31
3	*researching*	.53	.12	.07	.31	−.17	.58
21	*manufacturing*	−.20	−.45	−.03	.59	−.01	.41
13	*accomodating*	.05	−.06	.18	−.04	.86	.22
14	*caring*	.25	−.04	−.17	−.16	.80	.24
10	*managing*	.63	.26	.27	.20	.06	.42
22	*storing*	−.29	.58	.36	.23	.18	.36
16	*protecting*	.39	.35	−.15	.43	.33	.41
4	*programming*	.40	.54	.10	−.05	−.20	.50
15	*cleaning*	−.11	.00	.08	.31	.77	.29
18	*writing*	.22	.83	.25	−.17	−.08	.17
19	*calculating*	.20	.59	.49	−.10	−.17	.34
5	*applyinglaw*	.49	.64	−.12	.03	.00	.33
% Var. exp.[b]		18.95	13.20	12.22	12.00	11.13	
Cum. % Var. exp		18.95	32.16	44.38	56.38	67.51	
Cronbach's α[c]	Mean	.65	.49	.51	.51	.43	
	Range	.58–.73	.20–0.74	.32–.70	.39–.62	.29–.71	

Notes: [a]Uniqueness. [b]Percentage share of total shared variance explained by each factor (upper row) and cumulated share (lower row). [c]Due to the data's structure reported measures for α were calculated for each cross-section and apply to 2 to 6 variables. Sample is restricted to people from the western part of Germany belonging to the labor force (defined as having a paid work) aged 15 to 65 with a regular working time of at least ten hours per week. *Source*: BIBB/IAB—BIBB/BAuA Employment Surveys 1979–2006, weighted values, own calculations. $N = 118,105$

is not at all straightforward. As demonstrated in the previous section, the two strategies identified in the literature lead to different operationalizations. For this reason, we empirically evaluate single task items by analyzing the items' criterion validity. Table 6 gives the results.

First, we find that task variables 1 through 11 (for numbers see Table 5, column 1) are negatively signed and highly correlated (statistically significant if not stated otherwise) with manual and routine skill requirements but positively related to computer use.[25] Thereby, the tasks explain comparably high shares of variances in the routine measures (up to 10,5 percent in the case of *investigating*). Thus, it is highly valid to classify these tasks as non-routine cognitive. Unfortunately, we do not have measures for separating analytic from interactive tasks. However, it seems reasonable to classify variables *investigating*, *organizing*, *researching*, *programming*, and *applying law* (no. 1–5) as "analytic" and variables *teaching*, *consulting*, *buying*, *promoting*, *managing*, and *negotiating* (no. 6–11) as "interactive." Second, consistent with the classification by existing theory and research practice for the non-routine manual domain, task variables *repairing*, *accomodating*, *caring* and *protecting* (no. 12–14 and 16) are positively related to manual skill requirements (with comparatively high shares of explained variances) and negatively correlated with routine requirements. On the contrary, *cleaning* (no. 15) is positively correlated with our external measure for routine activities. *Cleaning* and the other four items are negatively linked to computerization, signaling that these activities are not characterized by computer activities. This hints that computers might at least not complement for these tasks as their usage decreases with the occurance of the items.

Third, in contradiction to the classification developed by existing theory and research practice, *measuring*, *writing*, and *calculating* (no. 17–19) are not valid measures for routine cognitive tasks; rather, our results suggest that they are measures for non-routine activities. Admittedly, in the case of *measuring*, the explained variance for routine measures is very low (below .01). *Writing* and *calculating* are surveyed only until the 1991/92 data collection. Their positive correlation with computer activities could be interpreted as either complementary or substitutionary effects. *Measuring* is not only inconsistently correlated with computer tasks, but also positively related to manual skill requirements. Altogether, the criterion validation demonstrates that the meaning of *measuring* changes throughout the rounds; classifying this task variable as a measure for routine cognitive tasks—as it is suggested by existing theory—is not supported by the criterion validation. An inspection of educational groups reveals (see Appendix Fig. 1) that *measuring* is more often performed by workers with high-skilled vocational degrees (ISCED 5B), suggesting that it rather measures a non-routine cognitive task.

Fourth, *operating*, *manufacturing*, and *storing* are supposed to be measures of routine manual tasks, according to research practice. Except for *operating* and *storing* in 1991/92, these task variables are actually comparatively strongly associated with manual skill requirements but only weakly with routine and computer skills. Again, the inconsistent direction of their correlation with *routine* and *computer* indicates that the meaning of the task measures may be changing over time.

Over and above these results, the criterion validation reveals that with regard to the problematic classification into routine and nonroutine tasks, only cognitive (analytic and interactive) nonroutine tasks are validly measured. For all other task variables the explained variance in the routine variable is low.

In sum, the criterion validation leads to a slightly different operationalization than the statistical and theory-based classifications (see last column in Table 4). In the next section, we explore the differences between the three operationalizations empirically.

5 Task change in Germany

This section compares the three operationalizations through a sensitivity analysis. We elaborate on whether all three lead to similar conclusions on task change in Germany or whether the results depend on the task measures used in the analysis.

Table 7 outlines the results for task-based change from 1979 to 2006 within the West German population aged 15 to 65, measured as additive and non-additive indices based on the three different operationalizations, respectively.

First, as suggested by the ALM hypothesis, analytic tasks consistently increase over each cross-section for all three measures. Similarly, interactive tasks increase when items are classified according to existing theory or the criterion validation, and when the indices are computed additively. The statistical classification of interactive tasks (i.e. factor 3 in Table 3) only includes *buying*, *promoting* and *negotiating* (but excludes *teaching*, *consulting*, and *managing*) and *negotiating* is only sporadically observable (see Table 3).

Second, the routinization hypothesis predicts an increase in non-routine manual tasks. In fact, these tasks consistently grow under the theory-driven additive index and for most of the other indices. However, the trend is not present in 1998/99, due to an especially low number of task items available to respondents in this survey year (only *caring* and *repairing*).

[25] The only exception is that *buying* is negatively related to *computer* in 2006.

Table 6 Criterion validation of task items

Variable number and name		Sign and average variance explained[a] of task variable on[b] ...					
		manual		routine		computer	
		Sign of assoc.	Var. explained	Sign of assoc.	Var. explained	Sign of assoc.	Var. explained
1	*investigating*	−	.046	−	.105	+	.043
2	*organizing*	−	.013	−	.094	+	.027
3	*researching*	−	.008	−	.063	+	.021
4	*programming*	−	.078	−	.030	+	.208
5	*applying law*	−	.018	−	.022	+	.013
6	*teaching*	−	.007	−	.061	+	.003
7	*consulting*	−	.028	−	.082	+	.011
8	*buying*	−	.019	−	.010	+ ('06−)	.001
9	*promoting*	−	.013	−	.020	+	.004
10	*managing*	−	.006	−	.071	+	.018
11	*negotiaing*	−	.022	−	.062	+	.025
12	*repairing*	+	.146	−	.003	−	.008
13	*accomodating*	+[3]	.004	− ('06+)	.000	−	.006
14	*caring*	+[c]	.005	− ('06+)	.004	−	.007
15	*cleaning*	+	.049	+	.021	−	.018
16	*protecting*	+	.013	−	.003	'79+; '06−	.002
17	*measuring*	+	.057	− ('06 n.s.)	.007	'79+,'99 n.s., '06−	.005
18	*writing*	−	.087	−	.031	+	.056
19	*calculating*	−	.038	−	.033	+	.047
20	*operating*	+	.093	+ ('79 n.s.)	.004	− ('79, 85/6+)	.005
21	*manufacturing*	+	.125	+	.004	−	.021
22	*storing*	+	.036	+ ('92−)	.018	− ('92+)	.017

Notes: [a]The table summarizes results of separate survey year regressions of task measures on the criterion variables. If not explicitly stated for single years, the association are at least statistically significant at the $p \leq .05$-level and have the reported signs in all separate regressions. "−" signifies a stat. sig. negative association with the criterion variable, "+" signifies a stat. sig. positive association with the criterion variable. [b]For detailed variable description see Sect. 3.3 and Table 8 in the Appendix. [c]Deviating from the operationalization as shown in Table 7, *manual* only includes "working standing". Sample is restricted to people from the western part of Germany belonging to the labor force (defined as having a paid work) aged 15 to 65, with a regular working time of at least ten hours per week. *Source*: BIBB/IAB—BIBB/BAuA Employment Surveys 1979–2006, weighted values, own calculations

Third, challenging the routinization hypothesis, none of the indices show a constant decline in routine cognitive tasks. Rather, the domain of cognitive routine tasks increases over time. Based on these results, we could (probably) misleadingly conclude that workers in Germany today perform more routine tasks than in the 1970s. This finding strongly deviates from the outcome suggested by the routinization hypothesis. However, our criterion validation suggests that researchers must be careful in classifying cognitive routine tasks in the German database. In our view, the criterion validation instead supports the view that the data does not sufficiently measure routine cognitive tasks over time. At least, we should be careful not to overemphasize the importance of observed changes in routine cognitive tasks.

Fourth, as with the contradicting results for cognitive routine tasks, manual routine tasks do not decline consistently over the observed period. Only when items are classified by criterion validation for the routine manual index, the overall picture until 1998/99 shows a decline. For all indices, however, there is a clear increase from 1998/99 to 2006. Again, this outcome does not support the routinization hypothesis, from which we would expect to see a constant decline. As with routine cognitive tasks, this could result from changes in how the task items are measured over time. As demonstrated in the section above, single items' meanings could have changed as well. We assume that the large increase until 2006 cannot (completely) be explained by changes in measures; however, we do not propose any alternative causes as an explanation. We inspected whether results differ if we alter the sample selection, but there are no changes for selected subgroups, e.g., for male full-time workers aged 25 to 55 (results upon request).

Table 7 Task-based change by different measures

Year	Statistical class.		Hypothetical C.		Criterion validation	
	Additive	Non-additive	Additive	Non-additive	Additive	Non-additive
1979						
Analytic	12.8	38.0	7.8	22.8	15.0	52.0
Interactive	9.6	23.4	11.9	35.1	11.9	35.1
Routine-.c	17.5	52.4	24.6	46.8	–	–
Routine-m	16.5	55.4	22.6	55.5	19.0	58.8
Nonr.-m.	5.7	13.2	7.2	27.2	6.9	24.0
1985/86						
Analytic	18.4	43.3	14.5	39.1	19.2	54.7
Interactive	17.2	28.2	17.4	48.5	17.4	48.5
Routine-.c	19.5	57.1	28.6	41.5	–	–
Routine-m	17.9	44.5	20.9	46.1	17.4	49.4
Nonr.-m.	5.7	14.6	8.8	37.5	9.3	33.7
1991/92						
Analytic	18.6	43.3	19.0	46.4	22.3	57.2
Interactive	21.0	29.0	18.0	45.2	18.0	45.2
Routine-.c	27.0	55.2	28.8	41.6	–	–
Routine-m	15.3	39.6	28.8	69.0	17.4	40.0
Nonr.-m.	7.6	18.8	8.8	35.5	7.9	28.7
1998/99						
Analytic	34.4	65.5	33.2	46.0	33.4	61.6
Interactive	20.1	38.8	26.3	63.2	26.3	63.2
Routine-.c	–	–	33.9	33.9	–	–
Routine-m	22.1	47.8	18.8	29.0	18.8	29.0
Nonr.-m.	7.3	7.3	12.2	23.8	12.2	23.8
2005/06						
Analytic	35.2	77.2	40.2	79.0	40.2	86.7
Interactive	17.4	29.4	28.5	67.1	28.5	67.1
Routine-.c	44.9	75.7	43.9	43.9	–	–
Routine-m	25.2	61.8	22.8	47.5	23.5	55.5
Nonr.-m.	14.0	28.2	16.2	48.0	19.5	33.3

Notes: Sample is restricted to people from the western part of Germany belonging to the labor force (defined as having a paid work) aged 15 to 65, with a regular working time of at least ten hours per week. *Source*: BIBB/IAB—BIBB/BAuA Employment Surveys 1979–2006, weighted values, own calculations

Altogether, our sensitivity analysis demonstrates that different classifications of tasks into the domains proposed by ALM lead to different conclusions on task change in Germany. To a certain extent, findings depend on the operationalization of task variables.

6 Summary and conclusions

The paper demonstrates that the German task data differ in a substantial way from DOT, O*NET and PDII, databases that have been used for analyses on the task approach primarily in the U.S. but also in other countries. Beyond these analyses, our paper describes broad changes in the data collection, survey population, and study content between survey waves. These changes also pertain to the task items; the content and wording of survey questions, the response categories, and the number of task items surveyed changed throughout the cross-sections. In conclusion, the process of classifying single tasks into distinct domains leads to a number of problems.

We apply different strategies to classifying tasks into the ALM task domains and perform a sensitivity analysis to compare these operationalizations; i.e., we analyze whether they all lead to similar conclusions on task change in Germany or whether results depend on measures. Our results

show that different operationalizations of task variables lead to different conclusions on task change in Germany.

From these results, we draw several conclusions. Our results suggest that for researchers using task items from multiple cross-sections, focusing empirically on the increase in non-routine tasks could be a reasonable strategy. For this group of tasks, the number of single task items is sufficiently high, these items are regularly observable throughout the cross-sections, their content did not change significantly from year to year, and measurement validity is comparatively strong. If routinization is nevertheless to be measured directly (and not indirectly by the increase in non-routine tasks), then we would suggest that researchers focus on manual routine tasks because, in our view, routine cognitive tasks are not sufficiently measured over time. An alternative to studying trends in routinization with the data would be to use our routine measures (or measures similar to ours) that rely on time-consistent information for several job (skill) requirements. Moreover, regarding the different classifications, we would prefer to classify items according to the results of our criterion validation. The information used to validate single task items proved a fruitful resource for deciding into which task cetegory single task items should be classified. Data on routine job contents could possibly also be used to construct a routine index apart from the single task items, which cover the other four ALM task domains. One exception might be *cleaning*, which could reasonably be classified as a non-routine manual task. We do not encourage researchers to categorize items by means of a statistical classification. First, the principal investigators never intended to measure the ALM-scheme, and second, the correlations reflect structural changes in the occupational organization of the German labor market. The classification based on existing literature and research practice may or may not be inferior to the other classifications; our results indicate that at least for routine tasks, the task items' meaning should be interpreted with caution. Finally, researchers comparing additive and non-additive measures should consider the number of items within each cross-section, as well as the availability of items between waves. For example, the non-additive indices are not practical if the number of available items changes significantly between waves or is particularly low in a given wave. However, more sophisticated ways of computing this index could be relevant in this instance but are beyond the scope of this paper.

Despite these limitations the German data nevertheless remain the single most comprehensive source of over-time task data for Germany.

In the near future, we plan to take two further steps with this research. First, we will include and synchronize data from the recently surveyed 2011/12 round. Second, we are interviewing a subsample of respondents from the 2011/12 cross-section again to learn more about the time spent on specific tasks at a regular working day and about task changes at the individual level. As a result of our work, we plan to precisely document our methods for compiling and synchronizing several years' worth of survey data and to allow interested researchers to access our programs and data files. The measurement and the operationalization of tasks are important current research questions (contributions from Autor and Handel on the T.A.S.K.S. II workshop in 2012 made this conclusion explicit). With our approach, we hope to contribute to this field of research in the task approach literature.

Executive summary

With the task approach skill-biased technological change was profoundly refined. To date, it was shown for a number of countries that differences in task content of occupations could explain changes in shares of employment as well as the wage structure. Beside this economic perspective the framework of the task approach proved useful for explaining developments in mismatching and outsorcing and was thus adopted in sociological research as well. Although the framework is in itself widely accepted, there are some issues to be solved.

One is that of a practical definition of routine-cognitive tasks, another that of measurement. During an international conference on "Technology, Assests, Skills, Knowledge and Specialization (T.A.S.K.S. 2)" in 2012 the problem of valid measurements of tasks was raised. Our paper shows, for the case of the commonly used German Employment Survey data 1979 to 2006, where these problems with measurement lie and what the consequences are.

The Employment Surveys provide a rich source of data for the analyses of task change in Germany. They hold information on skills, qualifications, tasks, held jobs and much more for more than 20.000 respondents in each cross-section. Starting in 1979, roughly every six to seven years a cross-section was surveyed, with data from 2005/06 being the most recent regarded for the analysis at hand. Another cross-section was sampled in 2011/12, which could not be regarded in our analysis. We demonstrate that the German task data differ in a substantial way from DOT, O*NET and PDII, databases that have been used for analyses on the task approach primarily in the U.S. but also in other countries. Beyond these analyses, our paper describes broad changes in the data collection, survey population, and study content between survey waves. These changes also pertain to the task items; the content and wording of survey questions, the response categories, and the number of task items surveyed changed throughout the cross-sections. In conclusion, despite being used frequently, the data does not lend itself automatically for over-time comparisons and the process of classifying single tasks into distinct domains leads

to a number of problems. Against that background we hypothesize that findings on task-biased technological change are sensitive to variable choice. To test our hypothesis, at first, we select 22 variables suitable for describing tasks over time. Then, we apply three different strategies for classifying these tasks into task domains (existing theory and research practice, factor analysis and criterion validation), and analyze whether different operationalizations lead to different conclusions about task change in Germany. All three approaches lead to slightly different operationalizations of task domains. We show that, due to the data's nature, single tasks regarding routine cognitive aspects are sparse. Nonroutine tasks, both analytic and manual are covered well by the data. Criterion validation reveals that some single tasks seem to have changed their meaning from the respondents' perspective. Altogether, each of the three approaches has drawbacks and advantages. While the approach of classifying tasks by existing theory and research practice allows for more single tasks to be included in the analysis, it is often not clear in which group to put single tasks due to changes in measurement and perceived meaning. Factor analytic approaches seem to overcome these problems at a first glance, but it is shown that it is not at all clear whether underlying factors actually are task groups (which we think they are not). The criterion validation approach uses alternative variables and measures to operationalize routineness, cognitive and manual aspects of occupations and therefore also provides a possible solution for the problem of routine manual tasks not being covered very well with items. Conforming to our hypothesis results on task change in Germany are very sensitive to variable choice. According to the task approach literature, we find large increases in nonroutine cognitive tasks. In contrast to it, we can not observe strong decreases in routine cognitive tasks. Rather, the domain of cognitive routine tasks increases over time.

Against the background of our results we provide several recommendations for applying the task approach to the German Employment Surveys. First, we advise not to overemphasize the importance of observed changes in routine cognitive tasks and instead empirically focus on the increase in nonroutine tasks. For this group of tasks, the number of single task items is sufficiently high, these items are regularly observable throughout the cross-sections, their content did not change significantly from year to year, and measurement validity is comparatively strong. Regarding the different classifications, we would prefer to classify items according to the results of our criterion validation. The information used to validate single task items proved a fruitful resource for deciding into which task category single task items should be classified. Data on routine job contents could possibly also be used to construct a routine index apart from the single task items, which cover the other four ALM task domains.

Kurzfassung

Der Ansatz eines qualifikationsverzerrten technischen Fortschritts (skill-biased technological change) wurde durch den Tätigkeitsansatz (task-approach) grundlegend weiterentwickelt. Mit diesem Ansatz, der nicht nur Qualifikationen, sondern auch die Art der Tätigkeit berücksichtigt, lassen sich die Lohn- und Beschäftigungsdynamiken in einer Reihe von Staaten gut erklären. Der Ansatz lässt sich auch etwa für die Erklärung von Fehlqualifizierungen und Outsourcing fruchtbar machen und wird daher zunehmend auch in der Soziologie verwendet. Wenngleich der Ansatz als solcher weitgehend akzeptiert ist, sind noch offene Fragen zu beantworten.

Dazu gehört eine praktikable Definition von kognitiven Routinetätigkeiten und die Frage valider Messungen von Tätigkeiten. Unsere Arbeit zeigt für die im Rahmen des Tätigkeitsansatzes häufig genutzten Daten der deutschen Erwerbstätigenbefragungen 1979, 1985/86, 1991/92, 1998/99 und 2006, worin die Probleme bei der Messung liegen und welche Konsequenzen daraus folgen.

Die Erwerbstätigenbefragungen stellen eine reichhaltige Datenbasis zur Analyse des Wandels von Tätigkeiten in Deutschland dar. Sie enthalten Informationen zu Fertigkeiten, Kenntnissen, Qualifikationen und Tätigkeiten für mindestens 20.000 Erwerbstätige je Welle.

Wir zeigen, dass sich diese Daten substantiell von denen des DOT, O*NET and PDII unterscheiden; Datenquellen, die genutzt wurden, um vergleichbare Analysen zum Taskansatz vor allem in den USA aber auch in anderen Ländern durchzuführen. Darüber hinaus beschreiben wir die Veränderungen in der Erhebungsmethode, der Grundgesamtheit und den Studieninhalten über die einzelnen Befragungen. Diese Veränderungen betreffen auch die Tätigkeitsitems, deren Inhalt und Formulierung, Antwortkategorien und Anzahl der Items sich im Zeitverlauf immer wieder geändert haben. Wir arbeiten heraus, dass die Daten der einzelnen Erhebungswellen nicht unmittelbar vergleichbar sind und die Zuordnung einzelner Tätigkeitsitems zu Gruppen mit einer Reihe von Schwierigkeiten verbunden ist. Vor diesem Hintergrund ist unsere Annahme, dass Ergebnisse zum task-biased technological change entscheidend von der Operationalisierung abhängen.

Um unsere Annahmen zu prüfen, wählen wir zunächst 22 Variablen aus, mit denen Tätigkeiten über die Zeit beschrieben werden können. Wir nutzen dann drei unterschiedliche Klassifizierungsstrategien für die Zuordnung einzelner Tätigkeitsitems zu Tätigkeitsgruppen (wissenschaftliche Praxis, Faktorenanalysen und Kriteriumsvalidierung). Schließlich untersuchen wir, ob die dabei resultierenden Operationalisierungen zu unterschiedlichen Ergebnissen im Hinblick auf den Tätigkeitswandel in Deutschland führen.

Zunächst stellen wir dar, dass Items mit denen kognitive Routinetätigkeiten abgebildet werden können, in den Daten so gut wie nicht enthalten sind. Demgegenüber sind Nichtroutinetätigkeiten, sowohl analytische als auch manuelle, gut abgedeckt. Anhand der Kriteriumsvalidierung zeigen wir dann, dass sich die Bedeutung einiger Items über die Erhebungswellen verändert. Im Vergleich der drei Klassifizierungsstrategien finden wir Vor- und Nachteile für jede Strategie. Während die Zuordnung nach Theorie und Forschungspraxis es erlaubt, mehr Items in die Analysen aufzunehmen, bleibt hier doch oft wegen Veränderungen der Messung unklar, in welche der fünf Gruppen sie klassifiziert werden sollten. Faktorenanalysen scheinen diese Probleme zu lösen, aber es zeigt sich, dass nicht klar ist, ob die zugrundeliegenden Faktoren tatsächlich Tätigkeitsgruppen abbilden (wovon wir nicht ausgehen). Bei der Kriteriumsvalidierung werden alternative Variablen genutzt, um die Routinehaftigkeit, kognitive und manuelle Tätigkeitsinhalte von Berufen zu messen und zu operationalisieren. Sie bietet damit ebenfalls eine Möglichkeit, das Problem der weniger gut abgebildeten manuellen Routinetätigkeiten zu lösen.

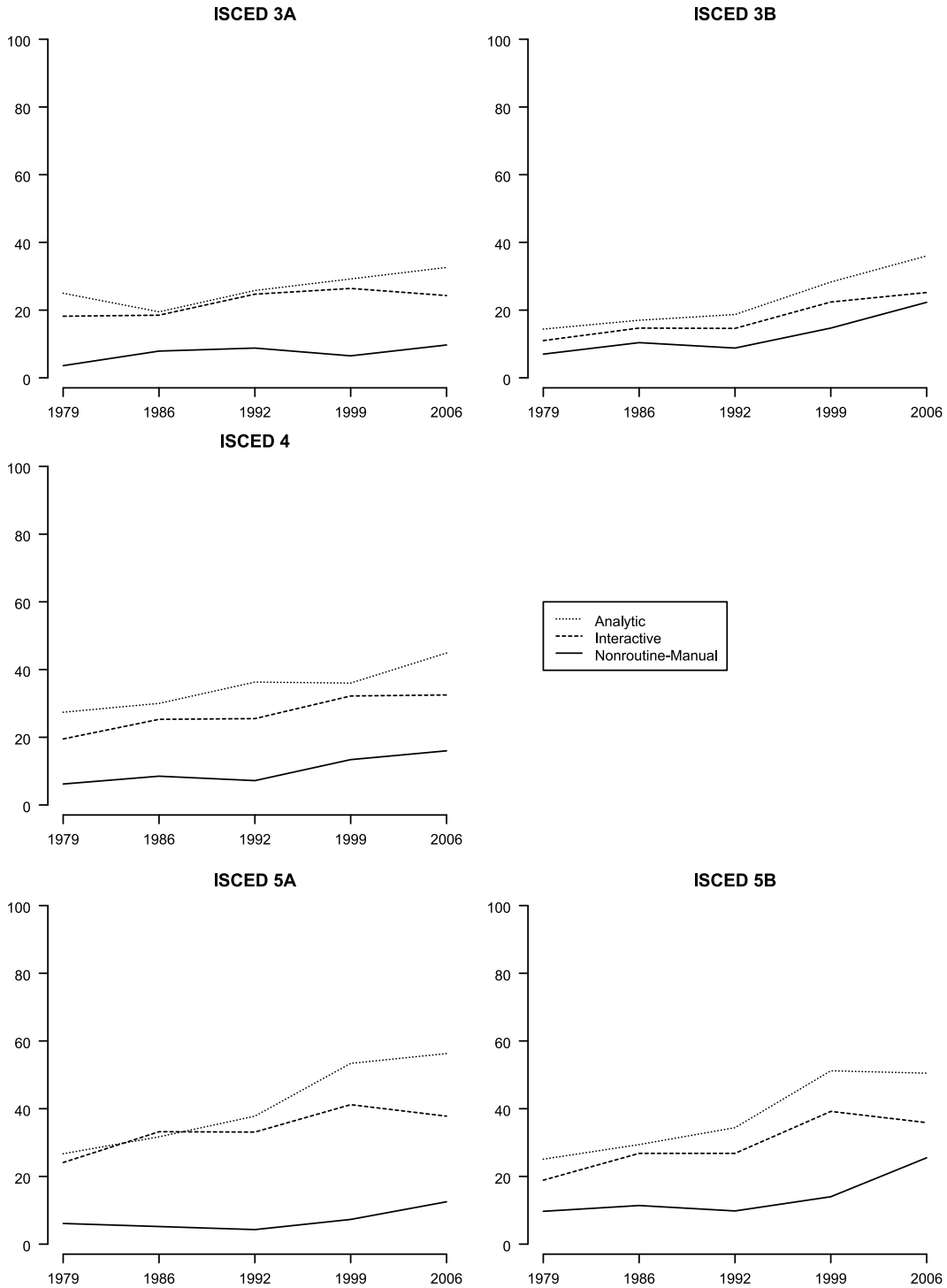
Unsere Annahme, dass Ergebnisse zum Wandel von Tätigkeiten in Deutschland von der Wahl der Variablen und der Operationalisierung abhängig sind, bestätigt sich. Gemäß einer der Grundhypothesen des Tätigkeitsansatzes finden wir übereinstimmend eine starke Zunahme von kognitiven Nichtroutinetätigkeiten. Entgegen den Erwartungen des Tätigkeitsansatzes können wir keine deutlichen Rückgänge für kognitive Routinetätigkeiten feststellen. Eher scheint die Gruppe der kognitiven Routinetätigkeiten über die Zeit an Bedeutung zu gewinnen.

Vor dem Hintergrund unserer Ergebnisse geben wir eine Reihe von Empfehlungen für die Anwendung des Tätigkeitsansatzes mit den deutschen Erwerbstätigenbefragungen. Zunächst folgern wir aus unseren Analysen, dass der Tätigkeitswandel valider anhand der Veränderungen in Nichtroutinetätigkeiten als in Routinetätigkeiten zu messen ist. Für die Gruppe der Nichtroutinetätigkeiten sind viele Items in allen Wellen vorhanden, ihr Inhalt bleibt über die Zeit vergleichbar und die Validität der Messungen ist hoch. Im Hinblick auf die drei Klassifizierungsstrategien empfehlen wir eine Operationalisierung, die die Bedeutung der Items anhand externer Kriteriumsvariablem validiert. Die Kriteriumsvariablen zu Routineinhalten von Berufen können auch genutzt werden, um einen Routineindex jenseits der Tätigkeitsitems zu erstellen.

Appendix

Table 8 Operationalization of criterion variables

Index	Variables
Manual	Work standing up (not in 91/2); Work sitting down; Carry heavy stocks; Working being exposed to smoke, dust, gases, vapors; Working being exposed to coldness, heat, wet, dampness, draft; Working with oil, dirtiness, mud (not in 79); Working crouched down, kneed, reclined, overarm; Working being exposed to strong tremor, crushes, vibrations (not in 91/2, 79); Working being exposed to dazzling light, or poor visibility / bad lightening (not in 91/2, 79; Working being exposed to hazardous substances, radiance (not in 79); Working being exposed to noisiness
Routine	Work is stipulated in the minutest details; One and the same work cycle / process is repeating in the minutest details; You are confronted with new problems that remain to be understood / familiarized with; Your tasks include process optimization or trying out new things;
Computer	Computer skills in application software;



Source: BIBB/IAB – BIBB/BAuA Employment Surveys 1979-2006, weighted values, own calculations.

Fig. 1 Task demand change by ISCED levels

References

Antonczyk, D., Fitzenberger, B., Leuschner, U.: Can a task-based approach explain the recent changes in the German wage structure? J. Econ. Stat. **229**(2+3), 214–238 (2009)

Ashton, D., Davies, B., Felstead, A., Green, F.: Work Skills in Britain. Oxford, SKOPE, Oxford and Warwick Universities (1999)

Autor, D., Dorn, D.: This job is "getting old": measuring changes in job opportunities using occupational age structure. Cambridge, Mass. eng. (2009)

Autor, D.H., Handel, M.J.: Putting tasks to the test. Human capital, job tasks and wages. Cambridge, Mass (2009)

Autor, D.H., Levy, F., Murnane, R.J.: The skill content of recent technological change: an empirical exploration. Q. J. Econ. **118**(4), 1279–1333 (2003)

Autor, D.H., Katz, L.F., Kearney, M.S.: The Polarization of the US Labor Market. National Bureau of Economic Research, Cambridge (2006)

Autor, D.H., Katz, L.F., Kearney, M.S.: Trends in U. S. wage inequality. Revising the revisionists. Rev. Econ. Stat. **90**(2), 300–323 (2008)

Black, S., Spitz-Oener, A.: Explaining women's success: technological change and the skill content of women's work. Rev. Econ. Stat. **92**(1), 187–194 (2010)

Bublitz, E., Noseleit, F.: The skill balancing act: determinants of and returns to balanced skills. Presentation at T.A.S.K.S. 2 workshop, Bonn (2012)

Cain, P.S., Treiman, D.J.: The dictionary of occupational titles as a source of occupational data. Am. Sociol. Rev. **46**, 253–278 (1981)

Cardiff School of Social Sciences (2012)

Coutsoukis, Photius and Information Technology Associates (2011)

Dustmann, C., Ludsteck, J., Schoenberg, U.: Revisiting the German wage structure. Q. J. Econ. **114**(2), 843–882 (2009)

Felstead, A., Gallie, D., Green, F.: Work Skills in Britain, 1986–2001. DfES Publications, Nottingham (2002). Skills Survey 2001

Felstead, A., Gallie, D., Green, F., Zhou, Y.: Skills at Work in Britain, 1986 to 2006. ESRC Centre on Skills, Knowledge and Organisational Performance, Oxford (2007)

Gathmann, C., Schoenberg, U.: How general is human capital? A task-based approach. J. Labor Econ. **28**(1), 1–50 (2010)

Geel, R., Backes Gellner, U.: Occupational mobility within and between skill clusters: an empirical analysis based on the skill-weights approach. Empir. Res. Vocat. Educ. Train. **3**(1), 21–38 (2011)

Geel, R., Mure, J., Backes-Gellner, U.: Specificity of occupational training and occupational mobility: an empirical study based on Lazear's skill-weights approach. Educ. Econ. **19**(5), 519–535 (2011)

Goos, M., Manning, A.: Lousy and lovely jobs. The rising polarization of work in Britain. Rev. Econ. Stat. **89**(1), 118–133 (2007)

Goos, M., Manning, A., Salomons, A.: Recent changes in the European employment structure: the roles of technology and globalization. Open Access publications from Katholieke Universiteit Leuven, Katholieke Universiteit Leuven (2009)

Gorsuch, R.L.: Factor Analysis. Erlbaum, Hillsdale (1983)

Green, F.: Employee involvement, technology and evolution in job skills: a task-based analysis. Ind. Labor Relat. Rev. **65**, 36–67 (2012)

Handel, M.J.: Measuring Job Content: Skills, Technology, and Management Practices. Discussion Paper, vol. 1357-08. Institute for Research on Poverty, University of Wisconsin, Madison (2008)

Handel, M.J.: The O*NET content model: strengths and limitations (2009)

Hartmann, J.: Trenddatensatz der BIBB/IAB-Erhebungen 1979–1999, München

Hartog, J.: J. Econ. Lit. **38**(4), 965–966 (2000). eng

Jäckle, A., Roberts, C., Lynn, P.: Assessing the Effect of Data Collection Mode on Measurement. Institute for Social and Economic Research, Essex (2008)

Kaiser, H.F.: The varimax criterion for analytic rotation in factor analysis. Psychometrika **23**, 187–200 (1958)

Kolenikov, S., Angeles, G.: The use of discrete data in principal component analysis with applications to socio-economic indices. WP-04-85 (2004)

LP, StataCorp (ed.): Stata statistical software: release 10. Texas: College Station (2007)

McGuinness, S.: Overeducation in the labour market. J. Econ. Surv. **20**(3), 387–418 (2006). eng.

National Center for O*NET Development (2012)

Nedelkoska, L., Neffke, F.: Human capital mismatches along the career path. Jena Econ. Res. Pap. 2010-051 (2010)

O'Connor, B.P.: SPSS and SAS programs for determining the bumber of components using parallel analysis and Velicer's MAP test. Behav. Res. Methods Instrum. Comput. **32**, 396–402 (2000)

Parmentier, K., Dostal, W.: Qualifikation und Erwerbssituation in Deutschland – Konzeption und inhaltliche Schwerpunkte der BIBB/IAB-Erhebungen. In: Kleinhenz, G. (ed.) IAB-Kompendium Arbeitsmarkt- und Berufsforschung. Beiträge zur Arbeitsmarkt- und Berufsforschung, vol. 250, pp. 31–44 (2002)

Preiser, K.: Analyse der Berufs- und Tätigkeitsmerkmale aus den Mikrozensen 1973 und 1976. Forschungsgruppe Arbeit und Gesundheit, Dortmund (1979)

Roberts, C.: Mixing Modes of Data Collection in Surveys: A Methodological Review. NCRM Methods Review Papers, vol. 008. ESRC National Centre for Research Methods, Southampton (2007)

Rohrbach, D.: The BIBB/IAB- and BIBB/BAuA surveys of the working population on qualification and working conditions in Germany. BIBB-FDZ Daten- und Methodenberichte 1 (2009)

Saris, W.E., Gallhofer, I.N.: Design, Evaluation, and Analysis of Questionnaires for Survey Research. Wiley, Hoboken (2007)

Spenner, K.I.: Skill: meanings, methods, and measures. Work Occup. **17**, 399–421 (1990)

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

Spitz-Oener, A.: The returns to pencil use reviseted. Ind. Labor Relat. Rev. **61**(4), 502–517 (2008)

Statistisches Bundesamt (ed.) Handbuch der Tätigkeitselemente. Wiesbaden (1967)

Statistisches Bundesamt: Tätigkeitsmerkmale der Erwerbstätigen. Ergebnis der Mikrozensus-Zusatzbefragung Juli 1969. Wirtsch. Stat. **9**(1971), 552 (1971)

Tiemann, M.: Die Entwicklung von Beruflichkeit im Wandel der Arbeitswelt. In: Bolder, A., Dobischat, R., Kutscha, G., Reutter, G. (eds.) Beruflichkeit zwischen institutionellem Wandel und biographischem Projekt. Springer, Wiesbaden (2012)

Wolff, H.-G., Bacher, J.: Hauptkomponentenanalyse und explorative Faktorenanalyse. In: Wolf, C., Best, H. (eds.) Handbuch der sozialwissenschaftlichen Datenanalyse, pp. 333–365. VS Verlag für Sozialwissenschaften, Wiesbaden (2010)

Zopf, S., Tiemann, M.: BIBB/BAuA-employment survey 2005/06. Schmollers Jahrb. **130**(5), 409–420 (2010)

Daniela Rohrbach-Schmidt Senior Researcher at the German Federal Institute for Vocational Education and Training (BIBB), Research Data Centre since 2008. M.A. in Sociology, Information Science and Media Science, University of Duesseldorf, 2004. Dr. rer. pol. in Sociology,

Faculty of Management, Economics and Social Sciences at the University of Cologne in 2008. Research Areas: Labor Market Sociology, Sociology of Education, Social Stratification, Quantitative Methods, Cross-National Comparisons.

Michael Tiemann Senior Researcher at the German Federal Institute for Vocational Education and Training (BIBB), Section "Qualification, Occupational Integration, Employment" since 2007. M.A. in Philosophy, University of Leeds, GB, 2000. M.A. in Sociology, University of Duesseldorf, 2006. Research Areas: BIBB/BAuA Employment Surveys, Changes in Occupational Contents, Knowledge in Work.