Working Paper No. 159

Different Types of IT Skills in Occupational Training Curricula and Labor Market Outcomes

Fabienne Kiener, Ann-Sophie Gnehm, Simon Clematide and Uschi Backes-Gellner

Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

UNIVERSITÄT BERN

Working Paper No. 159

Different Types of IT Skills in Occupational Training Curricula and Labor Market Outcomes

Fabienne Kiener, Ann-Sophie Gnehm, Simon Clematide and Uschi Backes-Gellner

February 2019

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

Different Types of IT Skills in Occupational Training Curricula and Labor Market Outcomes[1]

Fabienne Kiener[2], Ann-Sophie Gnehm[3], Simon Clematide[4]
and Uschi Backes-Gellner[5]

February 2019

First draft. Preliminary results. Only cite with permission of the authors.

Abstract

Due to an increasing diffusion of information technologies (IT), the labor market requires more and more so-called "IT skills". Recent studies confirm that IT skills are relevant for individual labor market outcomes. However, so far researchers do not use a consistent definition and measure of IT skills. In our paper, we distinguish different types of IT skills that may lead to structurally different labor market outcomes like wages. To measure these different types of IT skills, we propose an innovative way of measuring skills based on training curricula of apprenticeship occupations. We use modern computational linguistics methods, i.e. a topic modelling algorithm called Non-Negative Matrix Factorization. By doing so, we identify different types of IT skills like e.g. implementing ICT (Information and Communications Technologies), developing applications, designing webpages or installing software, handling system technology, CNC (Computerized Numerical Control), CAD (Computer-Aided Design), and handling control technology. Our results show that although IT skills in general have a positive effect on wages, different types of IT skills are associated with differing labor market returns, e.g. general digital skills like ICT and developing applications relate to higher wages than technology-specific IT skills like handling system technology.

Keywords: IT skills, information technologies, apprenticeship, training, curricula

JEL Classification: I26, J24, O33

---

[1] This study is partly funded by the Swiss State Secretariat for Education, Research, and Innovation (SERI) through its Leading House on the Economics of Education, Firm Behavior and Training Policies. We would like to thank Simon Janssen, Edward Lazear, Samuel Muehlemann, and seminar participants at the University of Zurich for helpful comments and the Swiss Federal Statistical Office for data provision.

[2] Corresponding author. University of Zurich, Department of Business Administration. Email: fabienne.kiener@business.uzh.ch. Address: Plattenstrasse 14, 8032 Zurich, Switzerland

[3] University of Zurich, Institute of Sociology

[4] University of Zurich, Institute of Computational Linguistics

[5] University of Zurich, Department of Business Administration

1. Introduction

Driven by the increasing diffusion of information technologies (IT), the labor market requires more and more IT skills. Recent studies like e.g., Falck, Heimisch, and Wiederhold (2016) or Janssen and Mohrenweiser (2018) validate the relevance of IT skills for labor market outcomes. Analyzing different levels of sophistication of IT skills, a study by Borghans and ter Weel (2006) suggests that only the most advanced IT skills like programming are associated with a positive labor market return. However, when examining IT skills, the researchers do not use a consistent definition and measure of IT skills so far. We distinguish between different types of IT skills by applying machine learning tools on occupational training curricula. We then investigate the relationship of these different types of IT skills on labor market outcomes.

To do so, we use modern computational linguistics methods, i.e. a topic modeling algorithm called Non-Negative Matrix Factorization (NMF). The method requires no assumptions on skill categories in advance in contrast to other skill measures that are currently used for example in job posting research (Atalay, Phongthiengtham, Sotelo, & Tannenbaum, 2018; Deming & Kahn, 2017).

Our skills data are derived from Swiss occupational training curricula for apprenticeships, i.e. the 3 to 4-year programs of Vocational Education and Training (VET). 62% of all students on the upper secondary level participate in VET in the school year 2016/2017. Each VET curriculum defines the content of training, and extensive final examinations guarantee that workers possess the required skills (Eggenberger, Rinawi, & Backes-Gellner, 2018). We focus on curricula currently in use that offer a Federal Diploma of VET. Our skills data includes 166 occupational training curricula.

As labor market data, we use the Swiss Social Protection and Labor Market Survey (SESAM) from the period 2010 to 2016 since we can build a rolling panel for this period. SESAM includes data on socio-demographics, administrative wage data, employment status and education histories (e.g. VET on upper-secondary level and original training occupation).

We match the original training occupation from SESAM to VET occupations from the skills data. We restrict our sample on the working population (age 15-64, positive labor income) and workers with VET as the highest education level. Our sample comprises 94’761 observations with 59’072 individuals.

Our results show that although IT skills have a positive effect on wages, different types of IT skills are associated with differing labor market returns. General digital skills (ICT & application) have higher labor market returns in terms of wages than technology-specific types of IT skills (CNC & CAD skills, handling system technologies, handling control technologies).

2. (IT) Skills Measures

2.1 Skills Measures in General

One approach that allows to measure latent (underlying) cognitive and non-cognitive ability is to condition on various observable characteristics (Heckman, Stixrud, & Urzua, 2006). This approach is less suitable to directly measure skills.

A second approach to measure more detailed skills sets is to analyze occupational skills requirements using a well-established database like the Occupational Information Network O*NET or its predecessor that draw on expert opinion and self-reports of workers (Deming, 2017; Weinberger, 2014). However, the comparison between occupational skill requirements poses some challenges as their absolute levels are not on the same scale (Handel, 2017).

A third approach to also measure occupational skills requirements analyzes job postings with machine learning methods (Deming & Kahn, 2017). Keywords in job postings allow for categorizing skills to study the skill requirements of each job posting.

A fourth approach is based on occupational training curricula that define the content of training and thus offer insights into the acquired skills. This method is for example applied for apprenticeship systems in Germany and Switzerland (Eggenberger et al., 2018; Jansen, Grip, & Kriechel, 2017; Janssen & Mohrenweiser, 2018).

2.2 IT Skills Measures

IT skills are in general all skills that help workers to handle, apply and use information and communication technologies (Falck, Heimisch, & Wiederhold, 2016). To measure IT skills there are also different methods found in the literature. Earlier studies focus on computer use at the workplace to proxy workers' IT skills (DiNardo & Pischke, 1997). More recent studies use self-reported IT skills (Borghans & ter Weel, 2006), examine job postings and the mentioning of IT software, languages and hardware (Atalay et al., 2018) or assess basic IT skills by letting workers solve simple IT problems (Falck et al., 2016).
However, all previous measures for IT skills are rather limited in their breath, including only few predefined IT skills such as programming, software configuration, application development, system administration or web design. Novel or more specific IT-related skills are neglected in such approaches although they may be as or even more important than the standard IT skills that have been typically used in the past.

3. Computational Linguistics Method

To overcome issues of the previously used IT skills measures, we draw on computational linguistics that combines data science with linguistics and helps to automatically analyze large amounts of text. Automatic text analyses offer a high potential for economic research (Gentzkow, Kelly, & Taddy, forthcoming). Previously, text analyses were based on human manual coding where categories were defined before the texts are categorized (Quinn, L. Monroe, Colaresi, Crespin, & Radev, 2010). Human manual categorization of skills has been successfully applied to Swiss occupational curricula (Eggenberger et al., 2018). We apply computational linguistics to Swiss occupational training curricula as the texts contain skill description that can be automatically detected by suitable methods of computational linguistics.[6]

Moreover, methods like topic modeling exist where no skill categorization in advance is needed (Grimmer & Stewart, 2013). Topic modeling uses algorithms to detect hidden, i.e. latent, structures within texts called topics. For curricula, the topics contain skills and other information that are described in them.

Topic modeling has never been applied to occupational training curricula before. So far, applications of topic modeling analyze twitter or news data, speeches or scientific abstracts for political, communication or marketing research (Amado, Cortez, Rita, & Moro, 2018; Benites-Lazaro, Giatti, & Giarolla, 2018; Maier et al., 2018).

We use a topic modeling algorithm called Non-Negative Matrix Factorization (NMF). The algorithm has proven its worth in several studies (Lee & Seung, 1999; Quinn et al., 2010; Shahnaz, Berry, Pauca, & Plemmons, 2006; Tjioe, Berry, Homayouni, & Heinrich, 2008; Yan, Guo, Liu, Cheng, & Wang, 2013).

3.1 Non-Negative Matrix Factorization

Our topic modeling algorithm uses matrix factorization as the name "non-negative matrix factorization" implies (see Gillis (2014) for mathematical explanation). NMF generates two output matrices (Words by Topics and Topics by Documents) from one input matrix. Thus the algorithm factorizes the input matrix into the output matrices.

The input matrix shows the words by each curriculum where not only single expressions but also words occurring next to each other are considered. In the appendix, the preparation of the texts is explained in more detail. The cell contains a value for how relative important the word is for that curriculum (so-called term frequency-inverse document frequency, tf-idf, showing how important the word is in

[6] The computations were carried out in the Computational Linguistics Lab at University of Zurich.

the curriculum relative to all words occurring in the curriculum times a weight how important the word is over all curricula).

Table 1: Illustrative, partly fictional Input Matrix for "apprentice develops user-friendly applications"

Document \ *Words*	*{apprentice, develops}*	*{develops, user-friendly}*	*{user-friendly, applications}*
Information Technologist	0.1	0.3	0.2
Photographer	0.4	0	0

Source: Own illustrative, partly fictional example of the input matrix.

Based on the input matrix, the algorithm builds two output matrices. One output matrix, Words by Topic, shows the most important words according to the weight for each topic. The other output matrix, Topics by Document, represents the distribution of the topics over each curriculum.

Table 2: Illustrative, partly fictional Output Matrix Words by Topic

Topic	*Most important words*
Topic 1	apprentice develops (0.28), user-friendly application (0.14)
Topic 2	image editing (0.292), …

Source: Own illustrative, partly fictional example of the output matrix Words by Topics. In parentheses weights showing the relative importance of the words by topic.

Table 3: Illustrative, partly fictional Output Matrix Topics by Document

Document	*Topic 1*	*Topic 2*
Information Technologist	0.2	0.01
Photographer	0.1	0.3

Source: Own illustrative, partly fictional example of the output matrix Topics by Document. Each cell shows a weight representing the importance of the topic in the curriculum.

3.2 Skills Measure

The researcher only specifies the number of topics in advance when applying the topic modeling algorithm. The algorithm is run with different number of topics, and the researcher evaluates which number of topics is optimal by assessing if the words within one topic are semantically coherent (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). We assess that 12 topics are optimal for our analysis as the words within each topic is semantically coherent (see 5 Results of the Topic Modeling).

The biggest contribution of the researcher is the interpretation of the topics (Gentzkow et al., forthcoming; Quinn et al., 2010). Our first measure for IT skills is a topic that contains mostly IT vocabulary. This measure represents a broad definition of IT skill. Our second measure for IT skills aims at identifying IT vocabulary across all topics to possibly distinguish types of IT skills. We assess the 100 most important words in each topic and identify if a word originates from IT. We take the weight from the algorithm that represents the importance of the word for the topic if the word is identified as IT. Then, we sum up the weights of 'IT words' for each topic. The calculated value shows how IT intensive each topic is (see 5 Results of the Topic Modeling).

4. Skills Data

Our skills data includes all 166 Swiss vocational education and training (VET) curricula currently in use. 62% of all students on the upper secondary level in the school year 2016/2017 participate in VET. Each curriculum defines the content of training, and extensive final examinations guarantee that workers possess the required skills.

We use the current curriculum of each VET occupation with a Federal Diploma as published on the web directory of the Swiss confederation[7]. The curricula are not only extensive (a total of 8'102 pages, a median of 44 pages for each curriculum) but also rich in content (a total of around 1.5 million words, a median around 8'030 words for each curriculum).

[7] https://www.becc.admin.ch/becc/public/bvz/beruf/grundbildungen downloaded in summer 2018

5. Results of the Topic Modeling

If the results of the topic modeling algorithm, NMF, are suitable to identify IT skills, we assess by the two output matrices. One matrix is the Word by Topics showing the relevant words of each topic as Table 4 illustrates. The topics contain semantically coherent words, since we recognize that the words within one topic relate to the same subject. An important topic is topic 10 as it shows IT vocabulary like "ICT". We conclude that the output matrix Words by Topics is useful for our analysis.

Table 4: Resulting Output Matrix Words by Topics

Topic	*Most relevant words*
Topic 1	act (handeln), competent (fachgerecht), according to (gemäss), following (folgend), standard (vorgabe), environmental protection (Umweltschutz)
Topic 2	agriculture (Landwirtschaft), plant cultivation (Pflanzenbau), be in force (gelten), organic farming (Biolandbau)
Topic 3	occupational safety regulations environmental protection (Vorschrift Arbeitssicherheit Umweltschutz), comply with occupational health and safety (Arbeitssicherheit Umweltschutz einhalten)
Topic 4	course (Kurs), to name (nennen), technical building system (gebäudetechnische Anlage), building technical (gebäudetechnisch), system (Anlage)
Topic 5	explain (erklären), education and training (Grundbildung), vocational education and training (berufliche Grundbildung)
Topic 6	process-oriented economically (prozessorientiert wirtschaftlich), process-oriented economic thinking (prozessorientiert wirtschaftlich denken), act think economically (wirtschaftlich denken handeln)
Topic 7	explain understand (erklären verstehen), apply (anwenden), name know (nennen wissen), describe understand (beschreiben verstehen)
Topic 8	explain (erklären), act (handeln), competent (fähig), explain (erläutern), guest (Gast), course (Kurs), course (Kursus), meaning (Bedeutung)
Topic 9	learning strategy apprentices (Lernstrategie Lernende), apprentices explain (Lernende erläutern), work technique apprentices (Arbeitstechnik Lernende)

Topic 10	ICT (ICT), explain autonomously (erklären selbstständig), implement (implementieren), handle small project (Kleinprojekt abwickeln), application (Applikation), server (Server)
Topic 11	situation (Lage), public (öffentlich), explain (erklären), benchmark requirement (Massstab Voraussetzung), customer (Kunde), correct (korrekt), comprehensible (nachvollziehbar)
Topic 12	permanent mold (Dauerform), following (folgend), lost (verloren), lost mold (verloren Form), cutting tool (Schneidwerkzeug)

Source: Output matrix Words by Topics resulting from NMF of the current curricula. In parentheses original word in German.

The other output matrix of NMF, Topics over Documents, shows how important each topic is for each curriculum. Each occupation contains different topics with different weights. Thus we have a variation of topics by each occupation. To analyze labor market outcomes, the variation is crucial.

We assume that our IT skills measure (1) recognizes obvious IT-related occupations and (2) detects IT skills in many occupations. Table 5 shows that IT-driven occupations like Information Technologist contain topic 10 supporting our assumption (1). Topic 10 as IT skills measure fulfills also our assumption (2) as 54 of the 166 analyzed VET occupations include topic 10. We conclude that the topic modeling algorithm NMF is suitable to identify IT skills in occupational curricula.

Table 5: Eight occupations where topic 10 is important

Occupation	*Weight Topic 10*
Information Technologist (Informatiker)	0.59
ICT Expert (ICT-Fachmann/-frau)	0.45
Mediamatic (Mediamatiker/in)	0.09
Multimedia electronics technician (Multimediaelektroniker/in)	0.09
Architectural model maker (Architekturmodellbauer/in)	0.06
Photographer (Fotograf/in)	0.06
Geomatics Technician (Geomatiker/in)	0.05
Micro draughtsman (Mikrozeichner/in)	0.05

Source: Own calculations.

Our second measure for IT skills is computed with identified IT vocabulary in the 100 most important words by each topic. Table 6 shows (some) of the vocabulary identified as IT by each topic and the calculated measure for IT skills where each the weight of each 'IT word' in the topic is summed up. The identified 'IT words' differ between topics allowing us to distinguish types of IT skills.

In topic 3, we find CNC (Computerized Numerical Control) and CAD (computer-aided design) that require IT skills. VET occupations with a high weight for topic 3 are Polymechanic, Machine Operator and Automation Technician EFZ.

In topic 9, the IT vocabulary is about handling system technology. VET occupations with a high weight for topic 9 are Electrician, Electrical planner, Assembly electrician and Telematics.

In topic 12, handling control technology occur. VET occupations with a high weight for topic 12 are Casting technologist and Molder.

Topic 10 is the topic with clearly the most IT vocabulary. It contains ICT and applications. VET occupations with a high weight for topic 10 are already shown in Table 5.

The IT words in each topic refer to different types of IT skills. We name the types of IT skills CNC and CAD (topic 3), handling system technology (topic 9), ICT and applications (topic 10) as well as handling control technology (topic 12). The types of IT skills distinguish between general digital skills (ICT & application) and technology-specific types of IT skills (CNC & CAD skills, handling system technologies, handling control technologies).

Table 6: Measure IT skills differentiating between types of IT skills

Topic	*Measure IT skills*	*Number of identified IT words*	*IT vocabulary (ordered by weights representing importance of words for topic)*
Topic 1	0	0	-
Topic 2	0	0	-
Topic 3	0.175	2	CNC, CAD
Topic 4	0	0	-
Topic 5	0	0	-
Topic 6	0	0	-

Topic 7	0	0	-
Topic 8	0	0	-
Topic 9	0.558	9	Electrotechnical (elektrotechnisch), system documentation (Anlagedokumentation), electrical system engineering (elektrisch Systemtechnik)…
Topic 10	4.497	56	ICT (ICT), implement (implementieren), application (Applikation), server (Server), application development (Applikationsentwicklung), configure (konfigurieren)…
Topic 11		0	-
Topic 12	0.135	3	CAM (CAM), manufacturing technology machine technology handling control technology (Fertigungstechnik Maschinentechnik Steuerungstechnik)…

Source: Own calculations.

6. IT Skills and the Labor Market

Even though there is some evidence on the positive association between IT skills and labor market outcomes (Atalay et al., 2018; Falck et al., 2016; Janssen & Mohrenweiser, 2018), research on different types of IT skills is limited. Borghans and ter Weel (2006) use survey data on the self-assessed computer use at work with different sophistication levels (advanced, complex, moderate, simple, no computer use) and find links between only the most advanced IT skills and returns on the labor market.

We will show how different types of IT skills relate to labor market outcomes. First, we explain the labor market data and the matching to the skills data. Then, we discuss our sample and show some summary statistics. Finally, we show the results of the relationship between different types of IT skills and labor market outcomes.

6.1 Labor Market Data

We use SESAM for data on the labor market. The data set comprises detailed information of a representative sample of the Swiss population aged 15 or older. SESAM links data from the Swiss Labor Force Survey (SLFS) with information from different social insurance registers. SLFS includes household interviews where since 2010, each household is interviewed five times for one and a half years. It comprises around 50’000 interviews per year. Each household is interviewed around five times. The Federal Statistics Office aggregates the data on year level and connects the information from the SLFS to administrative data.

SESAM has a rolling panel structure (Federal Statistics Office, 2011). We build a panel between 2010 and 2016 (similar to Balestra and Backes-Gellner (2017) and Eggenberger et al. (2018) who constructed a panel for earlier years).

As an independent variable, we use administrative wage data on the gross income from employment per year. Dependent variables from the labor market data are time-variant characteristics like age, tenure in the firm currently working as well as marital status and nationality. Important variables for the matching and sampling procedure are the education variable showing VET on upper-secondary level and the variable on the original training occupation.

6.2 Matching of Skills Data and Labor Market Data

We match the skills data and labor market data on occupations. The skills data are curricula used in VET occupations. The labor market data, SESAM, includes the variable on the original training occupation.

The original training occupation in SESAM for persons with an education on VET level roughly corresponds to the VET occupations of the curricula. We link the two occupations codes whereas a one-to-one matching of the occupation codes do not exactly correspond. We match 134 of the VET occupations to 215 original training occupations in the labor market data.

6.3 Sample

Our sample includes the data from 2010 to 2016 of the working population, i.e. persons who receive a positive labor income as well as are aged 15 to 64. As we study curricula of VET occupations, we restrict the sample to persons with vocational education and training[8]. Moreover, we only include individuals with values for the control variables (age, tenure, marital status, swiss nationality).

Our sample comprises of 94’761 observations from 59’072 individuals, whereof 56% are men and 44% women. Slightly more than half of the individuals are married (57%). 69% have a Swiss citizenship. The annual average wage over all observations is CHF 70’373. Table 7 displays summary statistics of the IT skills measures.

Table 7: Summary statistics of the IT Skills Measures

	Variable	*Mean*	*St. dev.*	*Min*	*Max*	*N if Variable=0*	*N*
IT Skills Measure 1		*.028*	*.089*	*0*	*.59*	*45’803*	*94’716*
IT Skills Measure 2	*CNC & CAD*	*.008*	*.024*	*0*	*.103*	*45’824*	*94’716*
	handling system technology	*.015*	*.070*	*0*	*.368*	*72’719*	*94’716*
	ICT & application	*.126*	*.4074*	*0*	*2.653*	*45’803*	*94’716*
	handling control technology	*.0025*	*.0046*	*0*	*.093*	*55’759*	*94’716*

Source: Own calculations of the summary statistics from the sample.

[8] Vocational Education and Training Federal Diploma of VET (Berufslehre EFZ) and Vocational matura (Berufsmatura)

6.4 Results on IT Skills and Labor Market Outcomes

To analyze the relationship between wages and skills, we use a Mincer-type wage regression. Skills measures are based on original training occupations that should not change over time. To interpret the results, we standardize the skills measures. Time-invariant controls are gender and a swiss nationality. Time-variant controls are age, age squared, tenure and tenure squared, marital status.

Our labor market data allows for constructing a rolling panel structure (6.1 Labor Market Data). To utilize that we have observations of some individuals for more than one year, we use a Pooled Ordinary Least Squares (OLS) model and cluster standard errors on individual level:

$$\log\left(wage_{i,t}\right) = \beta_0 + \beta_1 skills_i + \beta_2 gender_i + \beta_3 age_{i,t} + \beta_4 age_{i,t}^2 + \beta_5 tenure_{i,t} + \beta_6 tenure_{i,t}^2 + \beta_7 married_{i,t} + \beta_8 swiss_i + \varepsilon_{i,t} \quad , \text{ where } t = 2010, \ldots, 2016;\ i = 1, \ldots, N.$$

Our first IT skills measure is the importance of topic 10 in the original training occupation. The regressions in Table 8 show a significant positive relationship between the first IT skills measure and wages, also when including controls.

The coefficient in regression 2 means that an increase of IT skills by one standard deviation (measured as the importance of topic 10 in the original training occupation) is associated with an 9.02% increase of the wage. Age and tenure have a positive effect on wages but with a decreasing marginal effect. We find negative coefficients for being Swiss and married.

Table 8: Regression Wage on First Measure of IT Skills

	(1)	(2)
VARIABLES	*log wage*	*log wage*
IT skills (topic 10)	0.0875***	0.0902***
	(0.00352)	(0.00339)
age		0.0500***
		(0.00191)
age^2		-0.000552***
		(2.31e-05)
tenure		5.23e-05***
		(2.44e-06)
tenure^2		-1.85e-09***
		(1.82e-10)
swiss		-0.0626***
		(0.00656)
married		-0.112***
		(0.00638)
Constant	10.89***	9.796***
	(0.00346)	(0.0365)
Observations	93,252	93,252
Number of individuals	58,124	58,124
R^2 overall	0.0111	0.0636
R^2 between individuals	0.0106	0.0675

Source: Own calculations. Panel data 2010-2016. Standard errors in parentheses clustered on individual level, *** p<0.01, ** p<0.05, * p<0.1 .

Our second measure differentiates between types of IT skills. We have different types of IT skills: CNC & CAD (topic 3), handling system technology (topic 9), ICT & applications (topic 10), handling control technology (topic 12). To analyze if the effect of IT skills varies with the types, we regress the log of wage on each type of IT skills separately, and then including all types of IT skills. Each type of IT skills is standardized by itself. Regression 5 in Table 9 show a statistically significant positive relationship of each type of IT skills and wages when including controls and all types. The statistically significant coefficients of the types are the highest for ‘ICT & application’, followed by ‘CNC & CAD’, followed by ‘handling system technology’, followed by ‘handling control technology’. The results also show that general digital skills (ICT & application) have higher labor market returns than technology-specific types of IT skills (CNC & CAD skills, handling system technologies, handling control technologies).

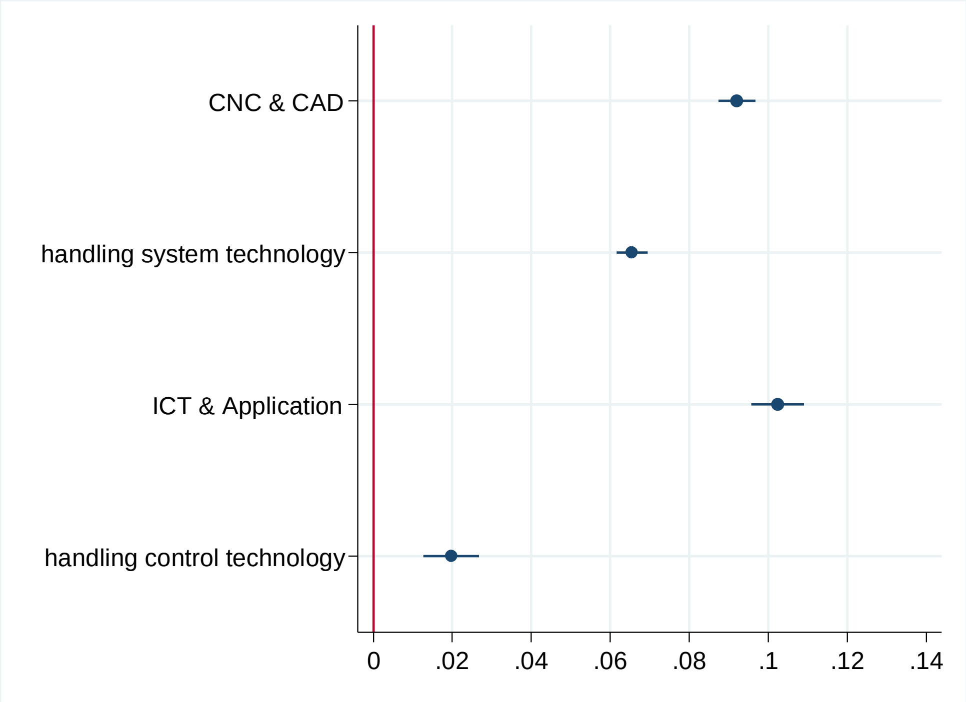
Table 9: Regression Wage on Different Types of IT Skills

VARIABLES	(1) *log wage*	(2) *log wage*	(3) *log wage*	(4) *log wage*	(5) *log wage*
CNC & CAD	0.0791*** (0.00234)				0.0920*** (0.00239)
handling system technology		0.0524*** (0.00198)			0.0655*** (0.00202)
ICT & application			0.0902*** (0.00339)		0.102*** (0.00339)
handling control technology				0.00524 (0.00349)	0.0197*** (0.00357)
Controls for age, age^2, tenure, tenure^2, swiss, married	Yes	Yes	Yes	Yes	Yes
Constant	10.89*** (0.00346)	9.818*** (0.0356)	9.862*** (0.0355)	9.908*** (0.0362)	9.796*** (0.0365)
Observations	93,252	93,252	93,252	93,252	93,252
Number of individuals	58,124	58,124	58,124	58,124	58,124
R^2 overall	0.0111	0.0238	0.0541	0.0559	0.0636
R^2 between individuals	0.0106	0.0247	0.0578	0.0595	0.0675

Source: Own calculations. Panel data 2010-2016. Standard errors in parentheses clustered on individual level. *** p<0.01, ** p<0.05, * p<0.1. Reading example regression (5) coefficient CNC & CAD: “an increase of 1 standard deviation of CNC & CAD leads to a 9.2% wage increase holding controls and other types of IT skills constant”

The coefficients of regression 5 for the different types of IT skills are statistically different from each other as the graph below shows. Only the confidence intervals on a 95% level of CNC & CAD and ICT & application are slightly overlapping.

Figure 1: Coefficients of different Types of IT skills as given in regression 5 above



Source: Own calculations. Panel data 2010-2016. Dots show the mean value of the coefficients and the line represents the confidence intervals on a 95% level.

6.5 Subsample Analysis

To further validate our results, we build a subsample of young workers that recently finished their VET. The subsample is more likely to have learned the skills described in the current curricula. We restrict the sample on workers older than 20 years old and younger than 26 who work full-time. Regression 1 in the table below includes our first measure of IT skills and shows a statistically significant coefficient for IT skills. Regression 2 including the different types of IT skills also shows positive statistically significant coefficients for each type of IT skills. As in the analysis of the full sample, the coefficient for general digital skills (ICT & application) is higher than the ones of the technology-specific types of IT skills.

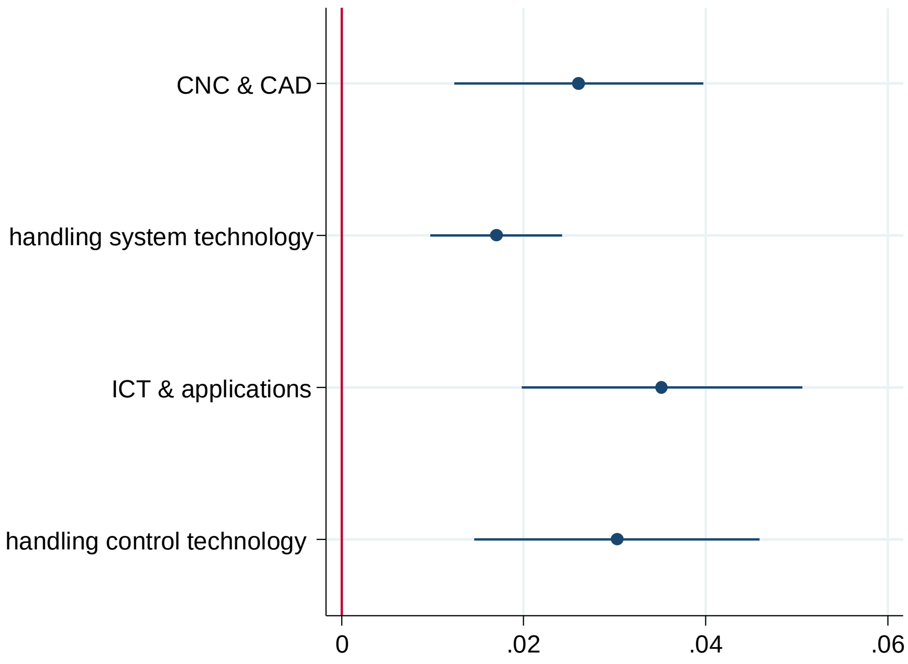
Table 10: Regressions Wage on Different Types of IT Skills

	VARIABLES	(1) *log wage*	(2) *log wage*
IT Skills measure 1 (topic 10)		0.0328*** (0.00782)	
IT Skills Measure 2	*CNC & CAD*		0.0261*** (0.00698)
	handling system technology		0.0170*** (0.00370)
	ICT & application		0.0352*** (0.00787)
	handling control technology		0.0303*** (0.00800)
Controls for age, age^2, tenure, tenure^2, swiss, married		Yes	Yes
Constant		8.455*** (1.751)	8.379*** (1.752)
Observations		4,938	4,938
Number of individuals		3,719	3,719
R^2 overall		0.0705	0.0781
R^2 between individuals		0.0734	0.0803

Source: Own calculations of the subsample (20<age<26, full-time workers). Panel data 2010-2016. Standard errors in parentheses clustered on individual level. *** p<0.01, ** p<0.05, * p<0.1.

The coefficients of the different types of IT skills in regression 2 above are graphically represented in the figure below. The confidence intervals of the coefficients are overlapping but that could also be due to less observations than in the full sample.

Figure 2: Coefficients of different Types of IT skills as given in regression 2 of the subsample above



Source: Own calculations of the subsample (20<age<26, full-time workers). Panel data 2010-2016. Dots show the mean value of the coefficients and the line represents the confidence intervals on a 95% level.

7. Conclusion

We distinguish between different types of IT skills using modern computational linguistics based on occupational training curricula. Our novel method allows for identifying types of IT skills like ICT and applications, handling system technology, CNC & CAD skills or handling control technology. We connect the identified types of IT skills with labor market outcomes using wage regressions.

Our results are statistically significant and show that (1) IT skills in general have a positive relationship with wages and (2) different types of IT skills are associated with differing labor market returns. The first result is in line with other studies analyzing IT skills and wages with limited measures (Atalay et al., 2018; Falck et al., 2016; Janssen & Mohrenweiser, 2018). The second result finds positive effects of all types of IT skills. We show that general digital skills (ICT & application) have higher labor market returns in terms of wages than technology-specific types of IT skills (CNC & CAD skills, handling system technologies, handling control technologies).

References

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, *24*, 1–7.

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2018). New technologies and the labor market. *Journal of Monetary Economics, 97*, 48–67.

Balestra, S., & Backes-Gellner, U. (2017). When a Door Closes, a Window Opens? Long-Term Labor Market Effects of Involuntary Separations. *German Economic Review*, *18*, 1–21.

Benites-Lazaro, L. L., Giatti, L., & Giarolla, A. (2018). Topic modeling method for analyzing social actor discourses on climate change, energy and food security. *Energy Research & Social Science*, *45*, 318–330.

Borghans, L., & ter Weel, B. (2006). Do We Need Computer Skills to Use a Computer? Evidence from Britain. *Labour*, *20*, 505–532.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 288–296). Curran Associates, Inc.

Deming, D., & Kahn, L. B. (2017). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, *36*, S337-S369.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, *132*, 1593–1640.

DiNardo, J. E., & Pischke, J.-S. (1997). The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too? *The Quarterly Journal of Economics*, *112*, 291–303.

Eggenberger, C., Rinawi, M., & Backes-Gellner, U. (2018). Occupational specificity: A new measurement based on training curricula and its effect on labor market outcomes. *Labour Economics*, *51*, 97–107.

Falck, O., Heimisch, A., & Wiederhold, S. (2016). *Returns to ICT Skills* (OECD Education Working Papers No. 134).

Federal Statistics Office. (2011). Syntheseerhebung soziale Sicherheit und Arbeitsmarkt (SESAM): Grundlagen, Methoden, konstruierte Variablen. Retrieved from https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/erhebungen/sesam.assetdetail.322180.html

Gentzkow, M., Kelly, B. T., & Taddy, M. (forthcoming). Text as Data. *Journal of Economic Literature.* Advance online publication.

Gillis, N. (2014). *The Why and How of Nonnegative Matrix Factorization* (Chapman & Hall / CRC machine learning & pattern recognition series).

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis, 21.*

Handel, M. J. (2017). Measuring Job Content: Skills, Technology, and Management Practices. In C. Warhurst, K. Mayhew, D. Finegold, & J. Buchanan (Eds.), *The Oxford handbook of skills and training.* Oxford, New York, NY: Oxford University Press.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics, 24,* 411–482.

Jansen, A., Grip, A. de, & Kriechel, B. (2017). The effect of choice options in training curricula on the demand for and supply of apprentices. *Economics of Education Review, 57,* 52–65.

Janssen, S., & Mohrenweiser, J. (2018). *The Shelf Life of Incumbent Workers during Accelerating Technological Change: Evidence from a Training Regulation Reform* (IZA Discussion Papers No. 11312).

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401,* 788–791.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures, 12,* 93–118.

Quinn, K., L. Monroe, B., Colaresi, M., Crespin, M., & Radev, D. (2010). How to Analyze Political Text With Minimal Assumptions and Costs. *American Journal of Political Science, 54.*

Shahnaz, F., Berry, M. W., Pauca, V.P., & Plemmons, R. J. (2006). Document clustering using nonnegative matrix factorization. *Information Processing & Management, 42,* 373–386.

Tjioe, E., Berry, M., Homayouni, R., & Heinrich, K. (2008). Using a literature-based NMF model for discovering gene functional relationships. *BMC Bioinformatics, 9,* P1.

Weinberger, C. J. (2014). The Increasing Complementarity between Cognitive and Social Skills. *Review of Economics and Statistics, 96,* 849–861.

Yan, X., Guo, J., Liu, S., Cheng, X., & Wang, Y. (2013). Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix. In J. Gosh (Ed.), *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 749–757). [Philadelphia, PA]: SIAM, Society for Industrial and Applied Mathematics.

Appendix

Preparing the texts

To prepare the texts to run the algorithm, we prepare the data according to computational linguistics standards. Words without meaningful content like articles (“the”, “their”) are excluded. Furthermore, the words are stemmed, such that different forms like plurals or verb conjunctions have the same stem.

The text is represented not as single expressions of words but as bag of words that contain several words that occur next to each other. The research can choose how many words the bag should contain (uni-, bi, trigram). We include single words (unigrams), two words together (bigrams) and three words together (trigrams). For example, the sentence “apprentice develops user-friendly applications” is in a trigram representation {apprentice, develops, user-friendly}, {develops, customer, applications} or in a bigram representation {the, apprentice}, {apprentice, develops}, {develops, user-friendly}, {user-friendly, applications} or in a unigram representation {the}, {apprentice}, {develops}, {user-friendly}, {applications}. The sequence of the words within a bag-of-word representation does not matter, meaning that a bag of words {apprentice, develops} is the same as {develops, apprentice}.

Additional Tables

Table 11: Correlations Wage and Topics

	Topic 1	*Topic 2*	*Topic 3*	*Topic 4*	*Topic 5*	*Topic 6*
Wage	-0.0306	-0.0264	0.0584*	0.0354*	0.0143	0.0131
	Topic 7	*Topic 8*	*Topic 9*	*Topic 10*	*Topic 11*	*Topic 12*
Wage	0.0163	-0.0619*	0.0343*	0.1033*	-0.0103	0.0390*

Source: Own calculations of pairwise correlations with a Bonferroni correction. Only data on 2016 included. * means statistical significance on the 0.01 significance level.

Table 12: Correlations Wage and different Types of IT Skills

	wage	*CNC & CAD*	*handling system technology*	*ICT & applications*	*handling control technology*
wage	1				
CNC & CAD	0.0584*	1			
handling system technology	0.0343*	-0.0660*	1		
ICT & applications	0.1033*	-0.0663*	-0.0650*	1	
handling control technology	0.0390*	-0.0810*	-0.1223*	0.0370*	1

Source: Own calculations of pairwise correlations with a Bonferroni correction. Only data on 2016 included. * means statistical significance on the 0.01 significance level.