Learning the Right Skill: Vocational Curricula and Returns to Skills*

Femke Cnossen
University of Groningen and GLO

Matloob Piracha
University of Kent, GLO and IZA

Guy Tchuente
Purdue University and GLO

September 6, 2023

Abstract

Using text data from curricula of the entire Dutch vocational education system, we extract verbs and nouns to measure basic cognitive and cross-functional skills, such as social, technical and resource management skills. This method allows us to uncover, at a fine-grained level, the skills middle-educated students learn at school. We then estimate the returns to these skills. Results show that resource management and social skills have the highest return. Importantly, the returns to skills are strongly conditional on sorting mechanisms. Our results fit in the literature that highlights the demand for complex interactive skills in environments that require more coordination.

*We thank Max Boiten, Steven Brakman, Elwin Buisman, Welmoed Cnossen, Robert Inklaar, Milena Nikolova, Olena Nizalova, Noemi Peter, Christian Siegel, Juliette de Wit, Klaus Zimmermann, participants from the GLO Virtual Young Scholar program, ESPE, SOM Conference, EEA and Untangled Conference for helpful comments and advice. Cnossen gratefully acknowledges funding from the National Program Groningen, under the project "Talent in the Region". Contact the corresponding author at f.d.cnossen@rug.nl. Any remaining errors are our own.

1 Introduction

Technology, globalization and structural change have worsened labour market outcomes of middle-skilled workers, and especially for the relatively recent entrants to the labour market. This has been related to changing demand in skills and tasks, yet we know little about which skills middle educated students are currently learning in school, and which of these help students achieve a good start on the labour market. This paper addresses two questions: (i) how focused are curricula on certain skills? and (ii) what are the returns to those skills? Based on recent literature, we focus on basic cognitive and cross-functional skills, such as social, technical and resource management skills. We develop a new measure of the relative weight of these skills in curricula by labelling extracted texts from the full set of training curricula in the Dutch middle education system to O*NET skill descriptions. By linking the curriculum data to register data on wages for the graduates, we can estimate skill returns on a fine-grained, curriculum-content level.

The existing literature describes how technology created poorer labour market trajectories of middle-skilled workers. New technologies and offshoring caused a decline in the labour demand for routine tasks, which are historically often executed by middle-skilled workers (Autor, Levy, and Murnane, 2003; Autor and Dorn, 2013; Goos, Manning, and Salomons, 2014; Goos and Manning, 2007). This is accompanied by increased demand for complex, analytical and social tasks, which are more complementary to technology (Autor, Katz, and Kearney, 2006; Spitz-Oener, 2006). At the same time, there is a general upskilling of the workforce, where more workers than ever are highly educated (Goldin and Katz, 2010). This upskilling has also led to higher skilled workers taking up less skilled occupations than before, increasing the average skill level within occupations (Beaudry and Green, 2003; Spitz-Oener, 2006). These features of the contemporary labour market make it increasingly difficult for middle-skilled workers to have fulfilling careers as routine occupations are disappearing, the remaining occupations are becoming more complex, and the set of occupations that do match their skill level are increasingly filled by higher skilled workers. For middle-educated labour market entrants, this has resulted in a declining probability of employment, lower starting wages, and slower wage growth than cohorts since the turn of the century (Reinhold and Thomsen, 2017; Dauth, Findeisen, Suedekum, and Woessner, 2021; Bisschop, Zwetsloot, Ter Weel, and Van Kesteren, 2020).

Despite this decline in employment, Autor (2015) expects that a 'significant stratum of middle-skill jobs combining specific vocational skills with foundational middle-skill levels of literacy, numeracy, adaptability, problem solving, and common sense will persist in the coming decades' (p. 27). He also points out that this prediction strongly depends on the education system being able to teach the current generation of middle-educated workers the "right" skills. This notion is the key motivation for this paper.

Besides job-specific and foundational skills, the extant literature shows several cross-functional skills that are increasing in importance. The most prominent of these are the so-called people skills; there is growing evidence of relative employment and wage growth for higher-educated occupations that require social skills. These are jobs that require high levels of coordination, persuasion, negotiation, social perceptiveness, influencing, and decision-making (Felstead, Gallie, Green, and Zhou, 2007; Borghans, Ter Weel, and Weinberg, 2014; Deming, 2017, 2021; Edin, Fredriksson, Nybom, and Öckert, 2022). Hansen, Ramdas, Sadun, and Fuller (2021) use a sample of executive occupations to show the growing importance of social skills, which not only includes social skills as persuasion, social perceptiveness or negotiation, but also human resource management.

In addition to interpersonal skills, there is evidence for a need of skills related to working with specific technologies. For instance, graduates from STEM-degrees earn more than other graduates in the first years after graduation (Deming and Noray, 2020), which can be explained by the fact that such programs are relatively intensive in technology-specific skills that complement technologies in a unique way. The fact that the bundle of skills in STEM degrees is intensive in technical skill might increase its returns, but this may also imply that non-STEM degrees might benefit from learning technical skills in different contexts, e.g., learning to use a computer in a business degree. Decomposing degrees into technical and non-technical skills may help us understand the extent to which wage differences between middle-educated graduates can be explained by differences in the share of technical skills taught.

Most of the current literature on the technology-induced changes in demand for different types of skills focus on the higher-earning end of the labour market. For instance, the findings of Deming (2017) are based on a sample restricted to professional occupations that employ predominantly college-educated workers. This is because vacancy data (in this case Burning Glass) has the most representative coverage for this group of occupations. Hansen et al. (2021) use a sample of executive occupations to show the growing importance of social skills, which might be explained by the fact that there is specific complementarity between cognitive and social skills, or that social skills are specifically required in leadership positions (Edin et al., 2022). Nevertheless, as the main sample of these analyses are college-educated workers, their patterns need not be the same for middle-educated graduates.

We use data from the Foundation for cooperation on Vocational Education, Training and Labour Market (SBB) to measure the skills that are described in the training curricula of the Dutch middle education system. More precisely, we extract verb-noun combinations (e.g., "read manual") which we map to O*NET skill categories. This allows us to categorise curricula as bundles of basic or cross-functional skills. Basic skills comprise those that relate to content (e.g., reading comprehension or mathematics) and those that relate to processing skills (e.g., critical thinking or monitoring). Cross-functional skills are either social (e.g., persuasion or social perceptiveness), technical (e.g., equipment maintenance or programming) or related to resource management (e.g., human or material resource management). Skills that do not fall in any of the O*NET skill categories or their subcategories are either too job-specific (e.g., milk cow) or too general (e.g., use context), and we label them as such. We merge the curriculum-based skill data to labour market outcomes of students that graduated in the period 2012 to 2018, using non-public microdata from Statistics Netherlands. We use this to estimate Mincer (1974) equations, where we regress log hourly wages in the years after graduation on our skill measures.

Our paper is closely related to Girsberger, Koomen, and Krapf (2022) who examine how Swiss vocational training in interpersonal, cognitive, and manual skills affect employment and wages. They highlight that manual skills are more beneficial to low-ability workers than social or cognitive skills. Eggenberger, Rinawi, and Backes-Gellner (2018) also perform content analysis of training curricula and focus on the relative returns to each skill. Unlike both papers, we use a broader set of O*NET skill categories to group the skills similar to Deming and Kahn (2017) and Hansen et al. (2021). However, our target population is the set of middle-educated workers and the textual analysis is carried out on a unique data set of Dutch middle school curricula. We estimate a Mincer wage equation on five major skill categories, in addition to uncovering the return to underlying fine-grained skills. We deem the returns to these detailed skills as one of the most important novelties of our paper. Furthermore, unlike the existing literature on vocational and middle-educated schooling systems (e.g. Malamud and Pop-Eleches, 2010; Golsteyn and Stenberg, 2017; Hanushek, Schwerdt, Woessmann, and Zhang, 2017; Tchuente, 2016; Eggenberger et al., 2018), our focus is not on the specific versus general skills but rather the type of skills that students learn.[1]

In terms of returns to skills, we find that social (e.g. service orientation, coordination, persuasion) and resource management skills (e.g. human or material resource management) have higher returns than basic content skills (e.g. reading, writing, mathematics). More precisely, a 10 percentage point increase in the share of resource management and social skills relative to basic content skills relates to, respectively, a 5.5% and 4.0% increase in wages. Furthermore, we find that the return to social skills depends on the field of study and the industry in which vocational graduates find employment. Further analysis of the mechanism that could explain some of the results shows that there are indeed differences within subcategories of skills. For instance, even though technical skills are overall not different from basic content, tasks like installation are valued more in the labor market, potentially because middle-educated are likely to be assigned those roles. Similar differences appear within subcategories of all other skills as well. We provide possible intuitions of our results in the paper. Furthermore, we find that resource management and social skills have higher returns when graduates are trained for middle-management programs, or are employed in service sectors and larger firms, where interaction and coordination is relatively more important.

The remainder of the paper is structured as follows. In the next section we describe the institutional context of our empirical analysis - the Dutch vocational education system. In section 3 we describe how we extract skills from curriculum data, how we link that information to labour market outcomes and show descriptive patterns of skill frequencies across various characteristics. Section 4 describes the empirical strategy while discussion of results are presented in Section 5. A number of robustness tests are discussed in Section 6. Finally, Section 7 concludes.

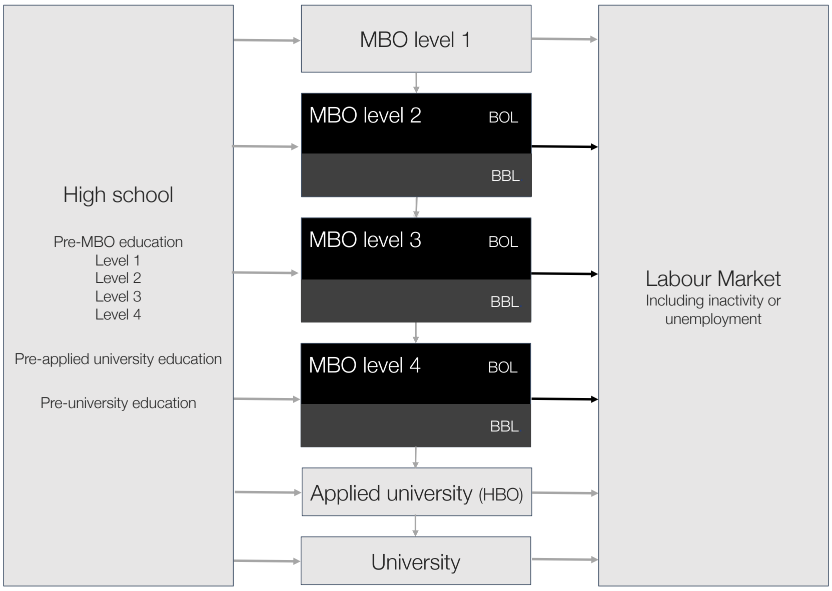
2 Institutional Setting: The Dutch vocational education system

The Dutch middle education system (MBO) is similar in nature to other education systems with distinct vocational education pathways, such as the German and Swiss VET. Typically, students enter the MBO around age 16 after having completed at least 4 years of (vocational) high school (see Figure 1). Upon entering the MBO students enrol in a training curriculum that is directly linked to a specific occupation. The training curricula are organised in domains, based on the industry in which the occupations often occur. For instance, healthcare training curricula are organised in the domain health and well-being, and engineering training curricula are in the domain of technology and process industry.[2]

[1] In terms of methods, this paper fits into a growing literature in economics that uses text data as the main source of information (Gentzkow, Kelly, and Taddy, 2019). Deming and Kahn (2017) use keywords in job postings to measure skill demand for higher-educated workers, while Hansen et al. (2021) map texts from occupational descriptions to skill clusters, in cognitive, interpersonal and operational dimensions. Webb (2019) compares the overlap in verb-noun combinations between occupational descriptions and patent data, to estimate the potential of replacement of workers by technology.

[2] These domains overlap largely with the one digit international fields of education (ISCED field), with a few exceptions. This is because the occupations are grouped into domains by similarity in tasks, whereas the fields classify programs based

Figure 1: **Institutional setting: Dutch vocational education system.**



Notes: In the Dutch education system high school graduates have three possibilities in further education: VET (MBO), Applied University (HBO) and University. Our focus is on MBO level education.

As some occupations are more challenging than others, each training curriculum is assigned a specific level, related to ISCED levels 1 to 4. Level 1 (the entry-program) is focused on acquiring basic learning and executive skills. This level does not directly lead to a starting qualification, and is not specifically oriented towards one occupation. Consequently, most students use Level 1 schooling as a stepping stone for further vocational training rather than as an entry to the labour market. Level 2 consists of basic vocational training and lasts between 2 and 3 years. Level 3 programs last for 3 to 4 years, and focus on learning to work independently. Lastly, students can enrol in level 4 programs that also last 3 to 4 years. This level covers middle management training, and prepares students for leadership positions in small teams.

The entry requirements of the levels in MBO are linked to the similar levels in the vocational high schools: a level 2 high school diploma gives access to a level 2 MBO program, but not to level 3. After finishing a level 2 MBO program, students are free to continue to level 3, and level 4 after that. More than 2/3rd of the Level 1 students and around 60% of the Level 2 students continue with their studies after graduation, in contrast to level 3 students of which around 40% continue schooling after finishing a program (Statistics Netherlands (CBS), 2016). With a diploma from the highest MBO level (4), graduates can also choose to continue to academic training, by going to a university for applied sciences (HBO), after which students can progress to university if they wish. Of those finishing level 4, around 40% of students continue to higher education, though this figure has been declining over the past decade (Statistics Netherlands (CBS), 2018). In our analysis, we focus on the highest obtained program in MBO. Students who continue studying after completing Level 4, i.e., move to (applied) university education, are excluded from the sample.

Within the MBO structure, students have to choose the form in which they prefer learning: either school-based training (BOL) or apprenticeship-based training (BBL). BOL has a focus on (theoretical)

on the similarities in skills and knowledge.

schooling, where roughly 20% of training time is spent as an apprentice. BBL is more oriented towards apprenticeships: its students are required to work at least 24 hours a week for a local firm. Both types of training lead to the same certification.

Each program has its own unique training curriculum that describes the skills that a student should have obtained at the time of graduation. These training curricula (or "qualification files") are organised by the Foundation for Cooperation on Vocational Education, Training and the Labour Market (SBB). They cooperate with the MBO schools and representatives from various industries to construct training curricula. They have a legal task (through the Dutch Act on Adult and Vocational Education) in developing and maintaining the entire qualification structure (SBB, 2022). The total set of qualification files is the main data source for this project, which will be described in more detail in the next section.

Each school chooses which training curricula to offer and students take the local offer of programs as given. VET students tend not to move for schooling - 4 in 5 students live with their parents during their entire program (Fouarge, Künn, and Punt, 2017). As a result, most students select a program from the set of training curricula that their local school offers. Empirical evidence on this selection process shows that students mostly make their choice based on what they like and what they are relatively good at in high school, plus the opinions of their friends and families. Expected labour market outcomes are generally not taken into account (Fouarge et al., 2017).

Because the Dutch setting contains a number of sorting choices, we need to add these choices to our empirical model. This means that in all our estimations we adjust for the level of education, the type of training (apprenticeship-based versus school-based), the school where they received training, and the graduation year. As we observe skills at the training curriculum level, adjusting for these variables implies that in our analysis we are comparing graduates within the same level, education track (BOL or BBL) and school, but with differences in the skills that they have learned.

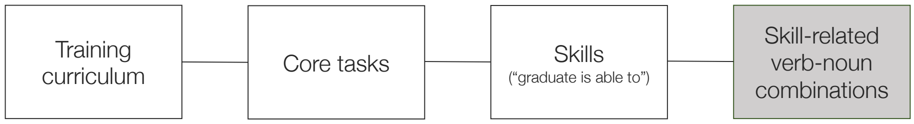
3 Data

3.1 Curriculum data

To understand which skills are part of each of the programs that we observe in the register data, we construct a new data set on curriculum-level skills. We obtain raw curriculum data from the SBB, which collects all qualification files for all training curricula in the Dutch middle education system. Each curriculum consists of a list of core tasks, associated with the occupation for which the student is trained. Each of these tasks is linked to a number of full sentences that describe what a student should be able to know or do for executing that task.

The fact that skill descriptions are written in full sentences hinders comparability across curricula. To ease comparison, we extract verb-noun combinations (as in Webb, 2019) from each sentence using a natural language processing module called Frog (Van den Bosch et al., 2007), see Figure 2. We use three features of this module. First, Frog replaces all words in each sentence by its dictionary form (lemmatizing). Second, the Part of Speech (POS) tag function categorizes each word as either a noun, adjective, verb or other grammatical form. Third, the dependency parser function determines the grammatical relation between words in a sentence, e.g. distinguishing objects from subjects. Using

Figure 2: **Set up of training curriculum, skills and extracted verb-noun combinations**



Note: Structure of qualification files from SBB. We use the program Frog (Van den Bosch, Busser, Canisius, and Daelemans, 2007) to extract verbs and nouns from skill descriptions.

the information from these features, we link verbs to their appropriate nouns within sentences.

We extract 4,045 unique verb-noun combinations. Many of these verb-noun combinations occur more than once, resulting in a full set of 12,470 such combinations.[3] To impose structure on the data, we label each of the unique verb-noun combinations to one of 30 skill descriptions based on those used by O*NET, similar to Deming and Kahn (2017); Deming and Noray (2020); Hansen et al. (2021). These skills can be grouped into five major categories: basic content, basic process, social, technical and resource management skills. Each of the verb-noun combinations is labeled to one of the skills that underlie these five categories. For instance, basic content skills contain the sub-skills active listening, mathematics, reading comprehension, speaking and writing. Basic process skills are related to procedures that contribute learning and improving on skills, and comprise the sub-skills active learning, critical thinking, learning strategies and monitoring. As an example, the verb-noun combination "read document" is labeled as reading comprehension, and thus as a basic content skill, whereas "can cooperate" is a coordination skill, falling in the category of social skills. Appendix Table A1 gives an overview of all the subskills, plus descriptions, examples of matched verb-noun combinations and relative frequencies of the sub-skills and major skill categories.[4]

The above procedure allows us to classify 50.5% of unique verb-noun combinations, which corresponds to 62.6% of all verb-noun combinations. The remaining group cannot be linked to any of the skills with a large degree of certainty, as some verb-noun combinations are either too general or too job-specific. Examples of overly general verb-noun combinations are "can work", "apply skill", or "execute technique". Too job-specific examples are "clean animal", "perform cashier operation" or "improve soil conditions". In our main analysis, we include the full sample of extracted verb-noun combinations. We include these two categories, labeled "Other: General" and "Other: Specific", in all our estimations. However, to test the sensitivity of our results to the non-classified verb-noun combinations, we perform robustness tests with the exclusion of either or both "Other" categories.

We use the labels for each verb-noun combination to construct weighted skill measures for the five main categories. Using these inputs, we construct the relative frequency of the skills within the

[3]Duplicates *within* curricula are removed, such that each entry can only appear once. For instance, if a curriculum mentions "measure weight of x" and "measure weight of y", we only count "measure weight" once, which is the unique skill. The full set of verb-noun combinations in a curriculum thus refers to the full set of unique verb-noun combinations.

[4]Manual labeling of verb-noun combinations is more prone to researcher bias. Therefore, we also introduced two independent researchers to the data. We presented them the list of verb-noun combinations, plus a list of O*NET skill descriptions (National Center for O*NET Development, 2022). See `https://www.onetonline.org/find/descriptor/browse/Skills/` for these skills and their descriptions. We asked them to label each of the verb-noun combinations to a specific skill on the list, and none if they felt no skill matched the verb-noun combination perfectly. After discussions among the co-authors and the two other researchers, differences were adjusted, either by relabeling the verb-noun combination (only once all of us agreed) or by placing them in the category "other: general" if they were too unclear to unambiguously link to one skill, or in the case of disagreement among the researchers.

curriculum of each program. The skill-frequency of skill s in program j is then calculated as

$$\text{SkillFreq}_{sj} = \frac{\text{Number of verb-noun combinations in program } j \text{ assigned to skill } s}{\text{All verb-noun combinations extracted from program} j} \tag{1}$$

Where the denominator captures the full set of unique verb-noun combinations within a program. The skill measures in the numerator add up to 100% within each curriculum. We describe the skill frequencies in subsection 3.3.

3.2 Schooling and student data

Using unique program identifiers ("crebo" codes), we are able to link each graduate to their respective program. We use the highest obtained diploma for each individual graduating between 2012 and 2018. Some of the students have been through multiple programs, of which we select the one with the highest level. We only select students who have finished a program in at least Level 2, as Level 1 programs do not lead to a starting qualification. Each training curriculum relates to a specific level of education (ISCED level 2 to 4), meaning that someone can learn reading skills in a level 2 and a level 4 program, which gives some information about the complexity of the acquired skill. The variation in our skill measurement is on the program level rather than the individual level. After cleaning, our data set consists of 296,106 employed graduates, divided over 375 curricula, 9 ISCED fields, 16 domains, and 72 schools.

Furthermore, we have individual-level characteristics of all students within programs that we use as proxies for ability: their age at graduation and their high school cumulative scores combined with the type of high school they completed.[5]

3.3 Curriculum contents

The skill frequency measure provides some insight in the skill intensities of curricula in the Dutch vocational education system. Since the skills relate to curriculum types and the people who select into them, we split the sample by demographic and program characteristics. The first column of Table 1 shows the average skill frequencies of the full sample, weighted by student population in each curriculum. Overall, basic content skills are most frequent (22%), followed by technical (13%), social (12%), basic process (8%) and resource management skills (7%). Columns 2 to 4 separate the sample by the three levels of education.[6] There is little variation in basic content skills across the levels of education, which is relevant as we will use it as the omitted category in our following

[5]The Dutch system has three types of high school: pre-university education, which lasts 6 years, a 5 year program which prepares for applied universities, and a 4 year vocational high school which prepares for the MBO. Most students will have completed the pre-MBO education, but students from the other types are also free to move to an MBO school.

[6]Descriptive statistics by gender, apprenticeship-type are provided in Online Appendix Table A2 and are shortly summarised here: men have relatively higher frequencies of technical and basic content skills, whereas women are more exposed to social and resource management skills in their curriculum. Given the practical nature of apprenticeships, students who follow an apprenticeship-based type of training are more often enrolled in technical-skill intensive curricula. In addition, we also provide brief descriptive statistics conditional on sorting into the labour market based on sector and firm size in the bottom panel of Online Appendix Table A2. Consistent with a framework where social skills are demanded in more complex environments that require coordination, they are relatively more frequent in both the high skilled services sector as well as in large firms. Technical skills, related to using equipment and machines are more present in the production of goods and in smaller firms. Resource management and basic process skills are relatively more frequent in both the low and high-skilled service sectors, and in larger firms.

Table 1: **Descriptive statistics**

		By level of education		
	Total	Level 2	Level 3	Level 4
	(1)	(2)	(3)	(4)
Skills				
Basic content	0.221	0.235	0.194	0.229
Basic process	0.077	0.063	0.059	0.092
Social	0.114	0.076	0.127	0.120
Resource management	0.075	0.028	0.086	0.086
Technical	0.132	0.221	0.139	0.097
Specific	0.068	0.071	0.102	0.049
General	0.313	0.307	0.292	0.326
Log hourly wage	2.470	2.322	2.482	2.517
sd	0.274	0.287	0.261	0.258
Covariates				
Female share	0.497	0.336	0.486	0.558
BBL share	0.275	0.393	0.445	0.146
Migrant share	0.177	0.214	0.150	0.175
Age at graduation	21.15	19.94	21.32	21.50
Observations	296,106	55,207	82,203	158,696

Note: Authors' calculations using non-public microdata from Statistics Netherlands and curriculum data from SBB. Skill frequencies are calculated using equation (1). Log hourly wage is calculated as contracted monthly income divided by contracted monthly hours in 24 months after graduation. Sample only includes graduates that found employment, students who are unemployed or continued studying are excluded. BBL stands for the share students that follow an apprentice-ship based training. Additional descriptive statistics by gender, sector of employment and firm size can be found in the Appendix Table A2.

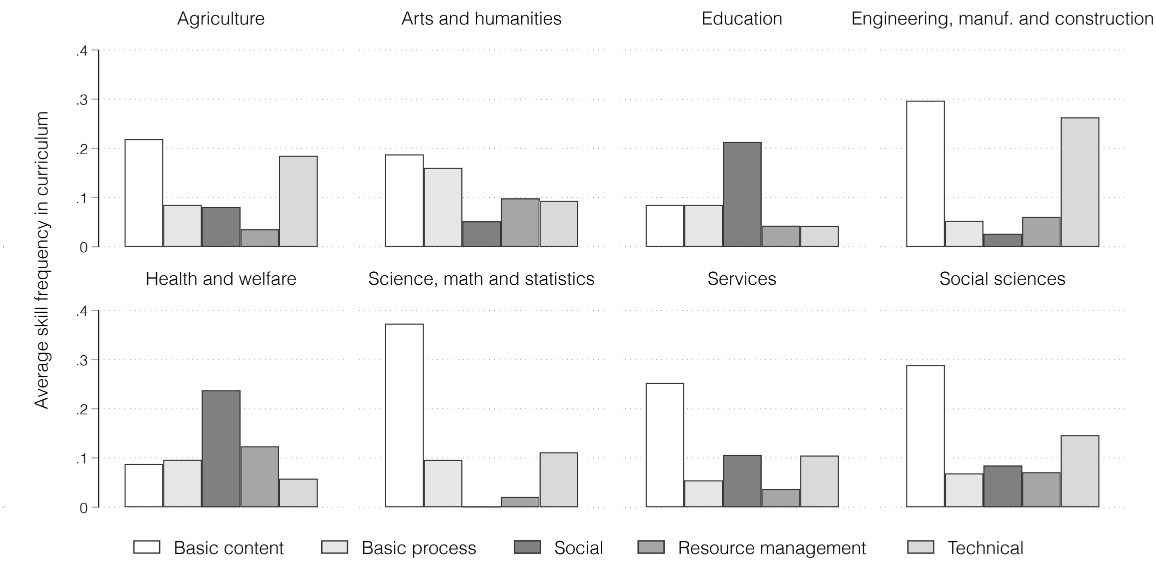
estimations. Level 2 students learn more technical skills, whereas level 3 and 4 programs are relatively more intensive in resource management and social skills. This is unsurprising, as students in level 3 and 4 programs are more often trained for (middle) management positions in organisations. Therefore, they require more social and management skills, relative to, say, technical skills.

Figure 3 describes the skill share by field of education. There is strong variation in the skill frequencies across different fields. For instance, social skills predominantly occur in education and health programs. Resource management skills are apparent across many different fields of education, and technical skills are relatively dominant in engineering and agriculture. Basic content skills appear to be relatively frequent in all programs, but especially in (social) science programs.

Table 2 shows the population-weighted[7] skill share for the more granular skill descriptions that underlie the major skill categories following National Center for O*NET Development (2022). Basic content skills (Panel A in Table 2) appear in almost all curricula, and across all fields. Reading, speaking, writing or otherwise conveying information and following instructions appear most often. These contain verb-noun combinations such as filling out forms, retrieve information, give presentation or read out data. The most common basic content skills are those related to reading, speaking, writing or otherwise conveying information and following instructions. Some of the basic content skills may be interpreted as social skills, as they are associated with either verbal or non-verbal communication. However, in O*NET they are classified as basic skills, as they facilitate the acquisition of new knowledge. Conversely, cross-functional social skills extend across several domains of activities. They are more complex, require social intuition, and are used in order to function well within an organisation. Basic social skills such as speaking are merely a means of communication, a 'background structure' necessary to acquire other skills in different domains

[7]For unweighted skill shares, please refer to Appendix Table A1.

Figure 3: **Skill frequencies, by field of education**



Source: authors' calculations using curriculum data from SBB and non-public microdata from Statistics Netherlands. The fields represent ISCED97 one-digit fields of education. Average skill frequencies are are calculated using equation (1) and weighted by student population. For descriptions of each skill category and underlying verb-noun combinations, please refer to the Online Appendix Table A1. A figure with both "Other" categories is presented in the Online Appendix, Figure A1

Table 2: **Skill frequencies (%) of curricula for the full sample, weighted by student population. See Appendix Table A1 for full descriptions of the skills.**

A. Basic content		**B. Basic process**		**C. Resource management**	
Listening	0.5	Critical think	1.9	Financial	1.3
Mathematics	2.2	Learning	3.1	Material	3.9
Reading	3.5	Learning strat.	0.2	Personnel	1.4
Speaking	4.1	Monitoring	2.5	Time	0.8
Writing	3.1				
Convey info	3.1				
Language	1.3				
Follow instruct'n	4.2				
Total	22.0	Total	7.7	Total	7.4

D. Technical		**E. Social**		**F. Other**	
Equip select	0.1	Coordination	2.3	Too general	31.5
Installation	1.4	Instructing	1.5	Too job-specific	6.7
Maintenance	1.5	Negotiation	0.2		
Operation/control	6.6	Perceptiveness	4.7		
Op. monitor	0.7	Persuasion	1.2		
Programming	0.2	Service orient.	1.7		
Quality control	1.7				
Repairing	0.9				
Total	13.1	Total	11.7	Total	38.2

Note: Authors' calculations using non-public microdata from Statistics Netherlands and curriculum data from SBB. Skill frequencies are calculated using equation (1), but for detailed sub-skill rather than broad categories. Detailed descriptions of the skills are provided in Appendix Table A1.

(National Center for O*NET Development, 2022).

Basic process skills (Panel B) help workers to rapidly acquire new knowledge and skills. Within this category, active learning and monitoring are most frequent. The latter contains all verb-noun combinations relating to evaluating one's own and others' behavior, in order to improve processes in the future. Examples of verb-noun combinations are: giving and receiving feedback and applying reflection, peer supervision or evaluation skills. Active learning concerns understanding how new information can guide future problem-solving or decision-making. It requires not only understanding the contents of information, but also interpreting it in the right way, making a connection with others types of information, and applying knowledge in the right context.

Resource management skills (panel C) are divided into four types of resources: financial, material, personnel and time. Financial resource management skills are those related to administration, determining budgets or making quotes. Managing material resources, i.e. seeing to the appropriate use of equipment, facilities and materials, is the largest group in this category. This contains, for instance, making provisions, applying safety measures, and storing products. Personnel resources are the management skills most related to social interactions. They include project management, motivating employees and demonstrating leadership abilities. Time management skills deal with all skills related to planning.

In the technical skills category (panel D) there are skills related to working with applications, machines or other systems. Most of the technical verb-noun combinations are found in operation and control, which among others captures computer-related skills. Importantly, these skills do not exclusively appear in technology-related programs, but across a wide range of fields, such as nursing, catering, electrical engineering, media studies and aviation. Programming skills, which can be perceived as a frontier skill, are only taught in very few programs, and mostly to software developers and IT experts.

Lastly, the largest collection of verb-noun combinations within social skills (panel E) are those related to social perceptiveness. These relate to applying observation techniques, reading a room, or dealing with aggression. Many of the social perceptiveness skills appear in health and well-being-related programs. For example, the skill 'dealing with group processes' is taught in the programs for educational assistant, sport coordinator and specialized pedagogical staff member. 'Recognize signal' appears exclusively in health programs - even though it could in theory also apply to recognizing signals on a machine: it appears in social work and sports curricula alone. Contrarily, there are some social skills that appear in fields that are not intuitively intensive in such skills. For instance, 'dealing with aggression' can be found in construction and infrastructure programs, and dealing with conflicts appears in service occupations (e.g. security) as well as in life-sciences occupations (e.g. environmental inspector).

3.4 Wage data

The individual's hourly wage is our main outcome of interest. Using the exact date on which each individual completed their highest program level, we are able to link the person's qualifications to the career trajectories in the early years after graduation. Wages and employment are estimated at 12-month intervals, from 12 to 84 months after graduation. We construct the hourly wage using monthly salaries and monthly hours worked. We only run estimations for wage-employed workers.[8]

Moreover, we observe each person's employer and the industry they work in. We use the first letters of the Dutch industry codes (SBI - Standaard Bedrijfsindeling), which correspond to the letter codes used in the NACE classification. For some of our analyses, we divide these industries into three sectors following Bárány and Siegel (2020): low skilled services (LSS), goods and high skilled services (HSS).[9] Descriptive statistics of log hourly wages in 24 months after graduation, by gender, level, apprenticeship-type, sector of employment and firm size are presented in Table A2.

4 Empirical strategy

In order to estimate returns to skills, we run Mincer (1974) type regressions, where we regress hourly wages in the early years after graduation on our skill measures. The goal of these analyses is to obtain a pattern of revealed skill demand for basic and cross-functional skills.

We estimate the parameters of the following equation, where we regress the log hourly wage in two years (24 months) after graduation ($t+2$) of individual i, who graduated in program j at time t:

$$\ln w_{ijt+2} = \beta_0 + \sum_{s=1}^{S} \beta_s s_j + Z_j^{'}\alpha + X_i^{'}\gamma + R_i^{'}\xi + \tau_t + \varepsilon_{ij} \tag{2}$$

where the s_j reflect the skill frequencies in program j, for $s_j \in \{$*basic content, basic progress, resource management, technical, social*$\}$ skills, Z_j is a vector containing program-related controls: level of education (2, 3 or 4) and education track (school-based or apprenticeship-based - BOL or BBL). X_i contains a vector of the demographic controls: gender, migration status, and age at graduation. R_i are school dummies and time dummies τ_t are the year of graduation.

An important point of discussion in the interpretation of the results is the potential risk of selection on skills. First, abler students might self-select into majors with high returns, which would be reflected by high and positive point estimates for certain skills. Besides ability, preferences and interests that influence one's choice of program might also correlate with labor market outcomes - diligent students are often diligent workers (Arcidiacono, 2004; Altonji, Blom, and Meghir, 2012). Positive point coefficients in Mincer equations could then wrongfully be interpreted as the return to that skill, rather than it being a return to the general ability or preference of the student selecting into this skill. This difficulty also explains why there is little empirical evidence on the effect of a curriculum on labor market outcomes (Altonji et al., 2012). In part, this is caused by a limited availability of curriculum data on a large scale. However, this gap can mostly be attributed to the fact that student selection into curricula is not a random process.

Despite the fact that returns to skills are never completely free from selection bias, we argue

[8]Those in full-time education, students working part-time, as well as unemployed and self-employed workers may earn (small) wages. Since they are not full-time available for the labour market they are excluded from the sample, as they could bias our results. We obtain the official labour market status from Statistics Netherlands.

[9]Low skilled services contain: wholesale and retail trade (NACE code G), transportation and storage (H), accommodation and food services (I), culture sports and recreation (R) and other service activities (S). The goods sector contains: agriculture, forestry and fishing (A), mining and quarrying (B), manufacturing (C) and construction (F). The high skilled service sector contains: electricity, gas, steam and air conditioning supply (D), water supply (E), information and communication (J), financial institutions (K), renting, buying and selling of real estate (L), professional, scientific and technical services (M), renting and leasing, and support services (N), public administration (O), education (P), human health and social work (Q).

that the institutional setting and our measure of skill frequency on the curriculum level resolves this bias to an extent, though we acknowledge and emphasize that it does not completely resolve the issue. First, we argue that the fine-grained nature of our text data partly circumvents the self-selection problem under some mild assumptions. Our reasoning is as follows: students will select into a higher hierarchical level (programs) than our observed data (verb-noun combinations) and though they have some expectation of the skills they will learn, they do not know the exact contents of a curriculum. At the least, they are unaware of how many times a specific verb-noun combination appears in a program's curriculum. Then, the number of its occurrences could be considered conditionally exogenous to the students. Second, since the curriculum is designed on a national level, individual schools cannot unilaterally change the curriculum based on their own preferences or local labour market demands. Lastly, to ensure that students are actually exposed to all skills described in the curriculum, the Inspectorate of the Ministry of Education tests whether each school's examination complies with the contents of the curriculum. For our strategy, this allows us to be relatively certain that whatever is described in the curriculum, will at some point have been taught to each graduate in that program.

Besides sorting into skills, school-level sorting may be an issue in choice of programs. Arcidiacono (2004) shows that high-quality schools make lucrative programs more attractive, and, since high-quality schools attract high-ability students, they contribute to the ability sorting across programs-within school. Even though, as mentioned above, there is little evidence that Dutch middle-education students choose schools specifically for its quality, it is something we want to adjust for. It might still be the case that, within schools, students with certain abilities or preferences choose programs with higher returns. Therefore, we cluster our standard errors at the curriculum level, leaving us with 375 clusters. Furthermore, in the absence of detailed program-school level information on the quality of teaching, we include school fixed effects to adjust for (teaching) quality differences between schools. Finally, we follow Altonji, Elder, and Taber (2005) by progressively adding controls that are strongly correlated with ability and preferences (unobservables) and evaluate their effects on the estimated returns to skills.

To further address the sensitivity of our results to self-selection, we perform a series of additional analyses to address ability-based sorting into skill-frequencies as well as possible - acknowledging that we cannot completely resolve a bias in our results. First, we document the pattern of sorting based on demographic and program characteristics, which will show the extent to which there is sorting into verb-noun combinations within levels, types and fields of education, as well as into schools. Second, for a sub sample of the population we have data on high school exam grades and parental income prior to enrolment into programs. We use these as proxies for ability and measure of disadvantage respectively, and show that our results are robust to their inclusion.

Lastly, even though we are interested in the average returns of our skill measures, a key empirical question relates to the underlying cause of the relationship between skills and wages. If it is the case that technological differences affect differences in skills demand, we should see different skill returns across sectors of employment. The literature shows that this might be the case, especially when observing changes in the occupational composition in sectors, which reflects a change in the demand for tasks (and thus skills) executed by workers. For instance, the goods sector has seen a strong reduction in the amount of routine workers, whereas the high skill service sector employs relatively far more abstract workers than before but sees no change in the routine-intensity of the average worker (Bárány and Siegel, 2020). Therefore, besides the standard wage equations, we also estimate the relationship between skills and earnings conditional on being employed in a certain sector. Furthermore, as social and management skills should be relatively more important in complex organizations (Hansen et al., 2021; Edin et al., 2022), we also estimate whether the effect changes for different sizes of the firms, which proxy for organisation complexity. These results are presented in the Online Appendix.[10]

5 Returns to vocational skills

5.1 Baseline results

Our baseline results are presented in Table 3. We regress the log hourly wage in 24 months after graduation on our skill measures, demographic variables and program controls. In all regressions, basic content skills are the omitted category. We regress earnings on skills by progressively adding our set of covariates, which helps determine the degree to which the observed returns could be driven by selection (Altonji et al., 2005).[11]

When all covariates are added, the results show that returns to basic process skills are not significantly different from basic content skills, which contain reading, writing and mathematics among others. However, when it comes to social and resource management skills, the returns are positive and statistically significantly different from basic content skills. Everything else being equal, a 10 percentage point increase in the share of social and resource management skills relative to basic content skills, respectively, relate to a 4% and 5.5% increase in wages in our full model (column 7).

Finally, technical skills are insignificant. There may be various reasons for this result. One of the possibilities is that the technical skills taught at vocational schools are not taught "on the frontier", or that the level of technical skills in demand in the labor market is higher than what is being taught. It may also depend on the transferability of technical skills (Eggenberger et al., 2018). The very detailed nature of our skills measure allows us to analyze skills at a granular level, which will help explore the potential mechanisms underlying our results. We do this in subsection 5.3 below.

The two "Other" skills in the table (specific and general) capture verb-noun combinations that are not classified as one of the O*NET skills. The interpretation of these coefficients is not straightforward, as both categories contain substantial noise from the data. In a robustness check (see Section 6), we test whether our results are robust to the exclusion of either or both of these "Other" categories.

Lastly, we explore the role of sorting into fields of education and industries of employment for explaining the direction and size of the skill coefficients. Online Appendix Table A5 adds ISCED field controls, domains, industry codes and firm size explicitly, to test the coefficient change. It is clear from the results that when we introduce these as controls to the model, we see an impact on the significance and size. Specifically, social skills turn insignificant when including field controls which implies that its mainly the fact that social skill-intensive fields pay well, rather than social skills being valued across all fields. Note that fields basically serve as groups of similar occupations.

[10] Tables A6 and A7.

[11] We also perform this same analysis for 12 and 48 months after graduation, but the results are qualitatively similar. These regressions are presented in Online Appendix Tables A3 and A4.

Table 3: **OLS Regression of Log Hourly Wages 24 months after Graduation on Skills**

		Demographic covariates			Programme covariates		
	(1) No ctrl	(2) Gender	(3) Migrant	(4) Age	(5) Level	(6) Type	(7) School
Basic process	-0.059	-0.003	-0.003	-0.036	-0.027	-0.011	0.079
	(0.224)	(0.206)	(0.208)	(0.184)	(0.176)	(0.176)	(0.168)
Social	0.306*	0.372**	0.373**	0.342**	0.377**	0.379**	0.396***
	(0.171)	(0.155)	(0.156)	(0.142)	(0.147)	(0.148)	(0.145)
Resource management	0.884***	0.868***	0.875***	0.677***	0.606***	0.584***	0.548***
	(0.222)	(0.211)	(0.211)	(0.194)	(0.195)	(0.194)	(0.182)
Technical	-0.013	-0.031	-0.031	-0.009	0.077	0.045	0.078
	(0.137)	(0.127)	(0.129)	(0.115)	(0.111)	(0.116)	(0.118)
Other							
Specific	-0.240**	-0.207**	-0.202**	-0.135	-0.092	-0.107	-0.097
	(0.105)	(0.097)	(0.098)	(0.088)	(0.090)	(0.092)	(0.091)
General	0.126	0.140	0.138	0.130*	0.150*	0.153*	0.180**
	(0.094)	(0.087)	(0.088)	(0.078)	(0.078)	(0.079)	(0.078)
Constant	2.277***	2.298***	2.295***	1.446***	1.448***	1.488***	1.457***
	(0.082)	(0.077)	(0.077)	(0.102)	(0.092)	(0.097)	(0.097)
Observations	296,106	296,106	296,106	296,106	296,106	296,106	296,106
R-squared	0.151	0.177	0.178	0.255	0.266	0.270	0.287

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors, clustered at the curriculum level. All skill coefficients are relative to the base category, basic content skills. Demographic controls: female (dummy), migrant (dummy), age at graduation. Program controls: ISCED level of education (2,3 and 4), apprenticeship type (BOL/BBL) and school dummies. BOL stands for school-based learning, and BBL stands for apprenticeship-based training.

Furthermore, technical skills turn significant and negative with the inclusion of field controls, again highlighting the role of occupational grouping. In terms of sorting into the labour market: the coefficient on social skills turn negative when industry codes are included, suggesting that people with social skills tend to sort into industries that reward these skills.[12] It is likely that the same occupation pays differently across different industries, based on the relative demand for skills in each industry. However, these results are only suggestive of such sorting patterns, and would require additional analyses to understand the process of sorting, conditional on skills and preferences. For this reason, our baseline model in the rest of the paper is column 7 in Table 3.

5.2 Heterogeneity in returns to skills

So far, we have estimated the average returns to skill based on the schooling background of each student. In this section we decompose the results by levels of education, to understand the returns to skill conditional on the educational level of the program. As most of the literature on skills points out, both the interaction between cognitive skills and non-cognitive skills matters, as well as one's position within a firm. Social skills are demanded especially in high-skilled, executive positions, linked to leadership roles (Hansen et al., 2021; Edin et al., 2022). In our institutional setting, level 3 and 4 programs are not only more complex, they are also oriented towards working independently or in middle management positions. In contrast, level 2 students train for assisting roles. The descriptive statistics in Table 1 highlight that different programs indeed differ in the attention given to certain skills: level 2 students learn fewer resource management and social skills than level 3 and 4 students. Instead, they have a relative stronger focus on technical and basic content skills.

[12]When we only include firm size as an employment control, the coefficient for social skills is positive. Sorting into industries is thus the main variable that changes the sign of the coefficient.

Table 4: **OLS regression, by level of education**

	(1)	(2)	(3)	(4)
			ISCED level	
	Baseline	2	3	4
Basic process	0.079	0.187	1.021**	-0.400*
	(0.168)	(0.185)	(0.452)	(0.221)
Social	0.396***	-0.069	0.504**	0.407**
	(0.145)	(0.174)	(0.240)	(0.169)
Resource management	0.548***	0.253	0.651***	0.631***
	(0.182)	(0.308)	(0.215)	(0.238)
Technical	0.078	0.014	0.452***	-0.147
	(0.118)	(0.098)	(0.156)	(0.146)
Constant	1.457***	1.217***	1.400***	1.750***
	(0.097)	(0.105)	(0.158)	(0.105)
Observations	296,106	55,207	82,203	158,696
R-squared	0.287	0.204	0.311	0.258

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors, clustered at the curriculum level. All skill coefficients are relative to the base category, basic content skills. All estimations include gender, migrant, age at graduation, type (BBL/BOL) and school dummies. BOL stands for school-based learning, and BBL stands for apprenticeship-based training.

We estimate the returns by level of education in Table 4 to show the interaction between the level of training and the skills taught. As expected, there is a much stronger positive relationship between social and resource management skills and wages for level 3 and 4 students. Furthermore, we find that especially those students that learn to work as independent professionals, rather than assistants or middle-managers, may benefit from learning technical skills. Our findings for middle-educated graduates are therefore in line with earlier findings for the higher educated population (see, for instance Deming, 2017; Deming and Noray, 2020; Hansen et al., 2021). In Online Appendix Tables A6 and A7 we additionally provide evidence on heterogeneous skill returns conditional on the sector of employment and firm size respectively. Here, sectors are defined as either the low-skilled services (LSS), goods or high-skilled service sector (HSS) following Bárány and Siegel (2020). For firm sizes, the sample is split into small firms (less than 50 employees), medium (between 50 and 250), large (between 250 and 2000), and very large firms (more than 2000 employees). Resource management and social skills have higher returns in the high-skilled services sector and in large firms, which would be consistent with a theory where interactive skills smooth coordination (Deming, 2017; Edin et al., 2022).

Our results for social skills are largely in line with those presented in Deming and Kahn (2017)), who show a positive return to such skills for university educated. However, rather than for a sample of high-skilled occupations, we show that middle-educated workers may also benefit from learning such skills. Furthermore, as we link our curriculum data to O*NET data, we also highlight the additional skill of resource management, which was also used by Hansen et al. (2021) for executive level jobs, to show that above and beyond social skills, resource management are important potential assets for middle-skilled workers. In addition, our heterogeneity analyses suggest to confirm existing evidence in the literature, in that social and management skills are especially relevant in environments that require more coordination, such as more complex middle-management occupations, high skilled service sectors and large firms.

5.3 Returns to skills at the granular level

In this section we further explore the returns to different sub-categories within each skill. We estimate equation (2) - column 7 in Table 3 - where we substitute the granular skills for the broader categories. Table 5 shows the results of this regression. We note that in the baseline model (Table 3), the positive returns for the resource management skills seem to be primarily driven by management of material resources - they are positive and significant in the granular analysis. This result seems intuitively clear since middle educated workers are likely to be tasked with planning, organizing and controlling the flow of materials, equipment and facilities necessary for the proper functioning of an organisation.

Although positive, management of personnel is insignificant. Though this sub-skill is in the resource management category, it can be seen as a complex social skill, in the sense that one is generating the social structure within a company, which requires higher level skill. Therefore, one reason for this could be that this part of management is more often performed by university-educated workers, who are likely to be higher on the occupational ladder compared to middle educated, and are therefore more likely to have overall line management responsibility within an organisation.

The exploration of social skills at a more granular level shows that most of the effect is driven by service orientation. One possible reason could be that service skills reduce cooperation costs independently of the position occupied in the company, and might be one of the key aspects for the middle educated, whereas other categories like coordination, negotiations etc are more relevant to those in upper management, who are more likely to be university educated. Just as the basic process skill coefficient is insignificant in the baseline regression, its components are also insignificant.

Within the technical skill category, most of the skills are not significant, with the exception of installation being positive and significant and quality control being marginally significant and negative. These findings suggest that for middle school graduates many technical skills are of no larger importance than basic content skills. Nevertheless, skills relating to knowing how to install machines and equipment has a positive return. As for quality control, it is likely that the low level of complexity of the task in some sectors of the economy may make middle-educated labor easily substitutable, rending the returns to quality control skills lower than basic content skills.

Table 5: **OLS Regression of Disaggregated Skills on Log Hourly Wages 24 months after Graduation**

	Basic process		Social		Resource management		Technical	
Baseline (from Table 3)		0.079 (0.168)		0.396*** (0.145)		0.548*** (0.182)		0.078 (0.118)
Disaggregated skill regression								
	Critical thinking	-0.138 (0.188)	Coordination	0.280 (0.322)	Finance	-0.858*** (0.273)	Select equip't	0.061 (0.710)
	Learning	0.208 (0.143)	Instructing	-0.373 (0.489)	Material	0.866*** (0.120)	Installation	0.369*** (0.119)
	Learning strat.	0.112 (0.999)	Negotiation	-0.920 (0.998)	Personnel	0.303 (0.366)	Maintenance	-0.077 (0.197)
	Monitoring	0.181 (0.280)	Perceptiveness	0.310 (0.255)	Time	-0.009 (0.204)	Operations	0.120 (0.113)
			Persuasion	-0.576* (0.306)			Oper. monitor	-0.597 (0.404)
			Service orient.	1.155*** (0.341)			Programming	-0.213 (0.246)
							Quality control	-0.315* (0.190)
							Repairing	-0.149 (0.148)

Observations: 296,106
R-squared: 0.345

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors, clustered at the curriculum level. The reference category for the skills is the category of basic content skills. Program controls: two-digit ISCED field, domain, ISCED level (2-4), school dummies, apprenticeship type (BOL/BBL). Firm controls: industry (NACE letters) and firm size dummies (small, medium, large, very large). See Appendix Table A1 for descriptions and (relative) frequency of the disaggregated skills. Constant is included but not reported.

Table 6: **Individual-level OLS Regression of skill shares on demographic and program characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Basic content		Basic process		Social		Resource management		Technical	
	Dem	Degree	Dem	Degree	Dem	Degree	Dem	Degree	Dem	Degree
Female	-0.110***	-0.002***	0.019***	-0.002***	0.088***	-0.000	0.032***	-0.007***	-0.091***	-0.011***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Migrant	0.017***	0.002***	0.002***	0.002***	-0.006***	-0.001***	-0.018***	-0.010***	-0.005***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Age	-0.005***	-0.001***	0.003***	0.001***	0.005***	0.001***	0.008***	0.003***	-0.005***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Level 3		-0.005***		-0.003***		0.044***		0.045***		-0.071***
		(0.001)		(0.000)		(0.000)		(0.001)		(0.001)
Level 4		0.015***		0.012***		0.034***		0.049***		-0.100***
		(0.002)		(0.000)		(0.000)		(0.001)		(0.001)
BBL		-0.046***		-0.003***		0.001***		-0.001***		0.040***
		(0.001)		(0.000)		(0.000)		(0.001)		(0.001)
N	296,106	289,800	296,106	289,800	296,106	289,800	296,106	289,800	296,106	289,800
R^2	0.113	0.615	0.024	0.396	0.178	0.711	0.095	0.482	0.113	0.563

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors. The reference category for the levels of education is ISCED level 2, and for training type the reference category is BOL (school-based training). All regressions include graduation year dummies. Demographic (dem) controls include gender, migrant status and age at graduation. Degree controls include level (ISCED level 2 to 3), type (BBL/BOL), 1-digit ISCED field of education, domains and school dummies. Constant is included but not reported.

6 Robustness tests

As discussed in the empirical strategy, estimating wage equations with skill variables poses a risk of endogeneity. In this case, our beta's might reflect sorting on ability, rather than returns to skill. We try to address this and other robustness tests in this section.

Table 6 presents estimations of simple linear models for the selection of skills. These estimations aim to quantify the extent to which the observable covariates explain a students' choice of learning certain skills. The unit of observation is a middle-educated worker. We regress the skill share in their curriculum on demographic (gender, migrant, age) and program covariates (ISCED level, apprenticeship-type, 1 digit ISCED field, domain, school dummies and graduation year). These variables correspond to the covariates used in column 7 of our baseline Table 3, apart from the field and domain covariates (see Table A5).

When degree controls are included, our observables explain at least 39% (basic process) and at most 71% (social) of the variation in our skill measures. Even though many of our observables are significantly related to skill shares, the effect sizes are small. Especially adding program-level controls change the size of the coefficients dramatically, converging them all towards zero. For instance, conditional on age and migration status, women have a 9.4% higher social skill share than men. However, this number reduces to 0.3% once we adjust for program-level variables, most notably the field controls. As Figure 3 showed, this is due to the fact that some skills are especially field-specific, such as social skills, whereas resource management skills are more evenly distributed over the several fields of education. Table 6 highlights that most of the sorting occurs in a student's decision on the level of education, the educational field, and the school where they want to study - not into specific skills.

Table 7 shows sensitivity analyses to the exclusion of certain verb-noun combinations or programs from our sample. Column 1 shows the baseline result, which is the same as column 7 in Table 3. Since some programs have sparse skill descriptions, we could only extract fewer than 10 or

Table 7: **Sensitivity tests to exclusion of i) programs with few verb-noun combinations, and ii) rare or unlabeled extracted verb-noun combinations**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Restrict # programs		Reduce denominator (# verb-noun combinations)				
		Number of vnc		Remove rare vnc		Remove (parts of) non-labeled		
	Baseline	>=10	>=20	>4 programs	>2 programs	With general	With specific	Removed
Basic process	0.079	0.055	0.272	0.091	0.212	0.030	0.043	0.034
	(0.168)	(0.219)	(0.221)	(0.130)	(0.151)	(0.140)	(0.083)	(0.074)
Social	0.396***	0.413***	0.531***	0.091	0.284**	0.302**	0.270***	0.223**
	(0.145)	(0.149)	(0.161)	(0.084)	(0.134)	(0.130)	(0.093)	(0.089)
Resource management	0.548***	0.516**	0.718***	0.502***	0.473**	0.374**	0.364***	0.277**
	(0.182)	(0.214)	(0.236)	(0.184)	(0.185)	(0.168)	(0.139)	(0.124)
Technical	0.078	0.167	0.405***	-0.043	0.049	-0.073	0.036	-0.035
	(0.118)	(0.126)	(0.120)	(0.063)	(0.094)	(0.085)	(0.091)	(0.074)
Other								
Specific	-0.097	-0.078	0.016	0.501	-0.126*		0.019	
	(0.091)	(0.097)	(0.097)	(0.376)	(0.070)		(0.028)	
General	0.180**	0.321***	0.433***	-0.002	0.123*	-0.033		
	(0.078)	(0.109)	(0.127)	(0.057)	(0.068)	(0.031)		
Constant	1.457***	1.381***	1.284***	1.514***	1.460***	1.544***	1.491***	1.514***
	(0.097)	(0.108)	(0.119)	(0.068)	(0.085)	(0.081)	(0.082)	(0.072)
Observations	296,106	272,023	226,269	292,115	292,123	296,106	291,621	291,621
R-squared	0.287	0.281	0.297	0.278	0.282	0.277	0.285	0.279

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors, clustered at the curriculum (crebo) level. The reference category is basic content skills. Program controls: ISCED level (2-4), school dummies, apprenticeship type (BOL/BBL). In col (2) and (3), programs with fewer than 10 or 20 verb-noun combinations respectively are removed from the analysis. In col (4) and (5), verb-noun combinations that appear in fewer than 5 and 3 programs are removed. In col (6) to (8) too specific, too general or all unlabeled verb-noun combinations are removed. Abbreviations: B. Process stands for basic process, Mgmt (Res) for resource management, vnc for verb-noun combination.

20 verb-nouns in them. In columns 2 and 3, therefore, we check the sensitivity of the results by removing all programs that contain fewer than 10 or 20 verb-nouns, respectively, and rerun our estimation for the sample that graduates in programs with at least 10 or 20 verb-nouns. Again, the overall results for social and resource management remain qualitatively similar.

In columns 4 to 8 we change the set of verb-nouns we used to construct our skill measure, by removing some of the verb-nouns. In columns 4 and 5, we remove all infrequent verb-noun combinations: we only use the verb-nouns that exist in more than 4 or more than 2 programs, respectively. This way, we test whether our results are driven by rare verb-noun combinations. The results remain robust in this case, though the result for social skills seems to be more driven by infrequent verb noun combinations than resource management skills. This probably has to do with the fact that especially education and health programs are social skill intensive, whereas many of these social skills are not found in other fields (e.g. assist patient). In columns 6 to 8, we remove part of the body of verb-noun combinations that was not classified as one of the O*NET skills. In column 6, we remove the verb-nouns from the category "other" that were too job specific, and in column 7 those that were too general. Column 8 shows the results when both "Other" skills are removed. The results are robust to this change in the sample of our verb-nouns and are qualitatively the same as our baseline.

Furthermore, since the data allows us to follow the wage trajectory of graduates, it is possible to test whether the skill returns change over time, see Table 8. This is done by following the 2012, 2013 and 2014 cohorts, which we can follow for 6 years after graduation, since our wage data is until

Table 8: **OLS Regressions of log hourly wages in t+x months after graduation, for cohorts 2012 to 2014**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cohort: '12-'14						'12+'13	'12
	t+12m	t+24m	t+36m	t+48m	t+60m	t+72m	t+84m	t+96m
Basic process	0.230	0.176	0.120	0.093	0.084	0.078	0.068	0.151
	(0.206)	(0.197)	(0.184)	(0.174)	(0.171)	(0.178)	(0.182)	(0.178)
Social	0.541***	0.470***	0.404**	0.373**	0.357**	0.338**	0.334**	0.352**
	(0.164)	(0.161)	(0.157)	(0.155)	(0.152)	(0.155)	(0.156)	(0.165)
Resource management	0.573***	0.563***	0.527***	0.511***	0.488**	0.468**	0.457**	0.458**
	(0.217)	(0.205)	(0.196)	(0.194)	(0.190)	(0.192)	(0.195)	(0.208)
Technical	0.115	0.084	0.069	0.084	0.092	0.108	0.100	0.137
	(0.136)	(0.129)	(0.120)	(0.117)	(0.115)	(0.120)	(0.121)	(0.118)
Observations	101,237	87,016	86,228	85,932	85,207	84,152	53,692	25,921
R-squared	0.284	0.248	0.209	0.194	0.187	0.195	0.187	0.178

Note: *** p<.01, ** p<.05, * p<.1. All models are estimated with robust standard errors, clustered at the curriculum level. All skill coefficients are relative to the base category, basic content skills. All estimations include gender, migrant, age at graduation, type (BBL/BOL) and school dummies. BOL stands for school-based learning, and BBL stands for apprenticeship-based training. Observations differ over years as workers flow in and out of employment. Constant is included but not reported.

and including 2020. In columns 7 and 8 we follow graduates for 84 and 96 months. Here, we use the cohorts of '12 and '13 in column 7, and only cohort '12 in column 8. The returns to social and resource management skills remain consistent over multiple years after graduation.

Even though some of the curricula have changed over time, we cannot tell with certainty in which version of a curriculum a graduate has completed their program.[13] Therefore, we are not able to assign skill frequencies to programs using curriculum data specific to the time each cohort was in school. This could contribute to measurement error in the skill frequencies that we document. To test whether the results are driven by one or a few cohorts, we rerun our baseline regression by graduation year. The results are shown in the Online Appendix Table A8 and are consistent across graduation years with our other findings.

Finally, to perform an extra test on whether the returns are driven by sorting on ability, the regressions are performed with two extra measures that could proxy for ability (Altonji et al., 2005, 2012): high school exam scores and parental income. For high school exam scores, we use data from Statistics Netherlands on the level of high school a person has completed and their final exam scores. For parental income, we use the log of total monthly taxable income (from either employment, self-employment or social security) in the month in which the student turned 15. At this age, students are still in high school and have not yet entered the vocational education system. Therefore, it is a relevant moment for which parental income should matter in the choice of schooling. The results are presented in Online Appendix Table A9 and are robust: both the exam as well as the parental income controls are significant, but the skill coefficients do not change.

[13] Some of the curricula in our sample will have seen updates, possibly changing the set of skills that students are required to learn. However, not all schools immediately adopt the new curricula, and due to this lag there is an unobserved difference in the skills taught to students. Combined with the fact that, in cases of small changes to the curriculum, the crebo-code (the unique curriculum identifier used to link graduates to curriculum data) remains the same, we cannot trace the version in which a student graduates. In case of substantial changes, a new crebo-code is assigned to a program.

7 Conclusions

This paper uses novel data on skills to estimate returns to specific elements in the curricula of Dutch middle-educated students. We extracted verb-noun combinations to measure basic and cross-functional skills, such as social, technical and resource management skills. We performed two main exercises with this data. First, we showed the extent to which each skill is taught in the curricula. Second, we linked the curriculum contents to wage data in the early years after graduation, to estimate returns to granulated skill levels.

We find that students graduating from programs with stronger focus on resource management and social skills have higher wages across all specifications. However, social skills are sensitive to the sorting by industry and the field of study. Additionally, we explored the sub-categories within the main category of each skill to ascertain the underlying mechanisms that drive our results. These mechanisms give us further insight in terms of what is valued more positively within each skill category. The key insights are that certain sub-skills within an overall skill drive the results.

Furthermore, we show that the returns to skill vary substantially across level of education. For instance, relative to basic content, resource management and social skills are especially valued in graduates that are trained in more complex curricula, geared towards (middle) management positions. This highlights the importance of interactive, cooperative skills for relatively high-skilled middle-educated graduates, which are in line with those found for the higher educated in the extant literature on the topic. In terms of vocational graduates learning the right skills, our results highlight that there is no 'one-skill-fits-all' solution, as different skills have different returns depending on the nature and complexity of training programs.

References

Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics* 4 (1):185–223.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113 (1):151–184.

Arcidiacono, Peter. 2004. Ability sorting and the returns to college major. *Journal of Econometrics* 121 (1):343–375.

Autor, David H. 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* 29 (3):3–30.

Autor, David H. and David Dorn. 2013. The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* 103 (5):1553–1597.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. The Polarization of the U.S. Labor Market. *American Economic Review* 96 (2):189–194.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics* 118 (4):1279–1333.

Beaudry, Paul and David A. Green. 2003. Wages and Employment in the United States and Germany: What Explains the Differences? *American Economic Review* 93 (3):573–602.

Bisschop, Paul, Jelle Zwetsloot, Bas Ter Weel, and Justus Van Kesteren. 2020. De overgang van het mbo naar de arbeidsmarkt: de positie van jongeren met een migratieachtergrond belicht. SEO-rapport 2020-05, SEO Economisch Onderzoek, Amsterdam.

Borghans, Lex, Bas Ter Weel, and Bruce A. Weinberg. 2014. People Skills and the Labor-Market Outcomes of Underrepresented Groups. *ILR Review* 67 (2):287–334.

Bárány, Zsófia L. and Christian Siegel. 2020. Biased technological change and employment reallocation. *Labour Economics* 67:101930.

Dauth, Wolfgang, Sebastian Findeisen, Jens Suedekum, and Nicole Woessner. 2021. The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association* 19 (6):3104–3153.

Deming, David. 2017. The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics* 132 (4):1593–1640.

Deming, David and Lisa B. Kahn. 2017. Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics* 36 (S1):S337–S369.

Deming, David J. 2021. The Growing Importance of Decision-Making on the Job. NBER Working Paper 28733, National Bureau of Economic Research, Cambridge, MA.

Deming, David J and Kadeem Noray. 2020. Earnings Dynamics, Changing Job Skills, and STEM Careers. *The Quarterly Journal of Economics* 135 (4):1965–2005.

Edin, Per-Anders, Peter Fredriksson, Martin Nybom, and Björn Öckert. 2022. The Rising Return to Noncognitive Skill. *American Economic Journal: Applied Economics* 14 (2):78–100.

Eggenberger, Christian, Miriam Rinawi, and Uschi Backes-Gellner. 2018. Occupational specificity: A new measurement based on training curricula and its effect on labor market outcomes. *Labour Economics* 51:97–107.

Felstead, Alan, Duncan Gallie, Francis Green, and Ying Zhou. 2007. *Skills at work, 1986 to 2006*. ESRC Centre on Skills, Knowledge and Organisational Performance (SKOPE), publisher's version ed.

Fouarge, Didier, Annemarie Künn, and Dijs Punt. 2017. De rol van arbeidsmarktinformatie in de opleidingskeuze van mbo'ers. ROA Report 009, Maastricht University, Research Centre for Education and the Labour Market (ROA), Maastricht.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. Text as Data. *Journal of Economic Literature* 57 (3):535–574.

Girsberger, Esther Mirjam, Miriam Koomen, and Matthias Krapf. 2022. Interpersonal, cognitive, and manual skills: How do they shape employment and wages? *Labour Economics* 78:102235.

Goldin, Claudia and Lawrence F. Katz. 2010. *The Race between Education and Technology*. Cambridge, Massachusetts; London, England: Belknap Press.

Golsteyn, Bart H. H. and Anders Stenberg. 2017. Earnings over the Life Course: General versus Vocational Education. *Journal of Human Capital* 11 (2):167–212.

Goos, Maarten and Alan Manning. 2007. Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics* 89 (1):118–133.

Goos, Maarten, Alan Manning, and Anna Salomons. 2014. Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review* 104 (8):2509–2526.

Hansen, Stephen, Tejas Ramdas, Raffaella Sadun, and Joe Fuller. 2021. The Demand for Executive Skills. Working Paper 28959, National Bureau of Economic Research.

Hanushek, Eric A., Guido Schwerdt, Ludger Woessmann, and Lei Zhang. 2017. General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle. *Journal of Human Resources* 52 (1):48–87.

Malamud, Ofer and Cristian Pop-Eleches. 2010. General Education versus Vocational Training: Evidence from an Economy in Transition. *The Review of Economics and Statistics* 92 (1):43–60.

Mincer, Jacob. 1974. Schooling, Experience, and Earnings. NBER Books, National Bureau of Economic Research, Inc.

National Center for O*NET Development. 2022. Skills. Www.onetonline.org/find/descriptor/browse/Skills.

Reinhold, Mario and Stephan Thomsen. 2017. The changing situation of labor market entrants in Germany. *Journal for Labour Market Research* 50 (1):161–174.

SBB. 2022. About SBB. Www.s-bb.nl/en/about-sbb.

Spitz-Oener, Alexandra. 2006. Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics* 24 (2):235–270.

Statistics Netherlands (CBS). 2016. Helft leert door na beroepsopleidende leerweg mbo. *Centraal Bureau voor de Statistiek* December 12.

———. 2018. Minder mbo'ers naar hbo in afgelopen decennium. *Centraal Bureau voor de Statistiek* January 24.

Tchuente, Guy. 2016. High School Human Capital Portfolio and College Outcomes. *Journal of Human Capital* 10 (3):267–302.

Van den Bosch, Antal, Bertjan Busser, Sander Canisius, and Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. In *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*. 99–114.

Webb, Michael. 2019. The Impact of Artificial Intelligence on the Labor Market. SSRN Scholarly Paper 3482150, Social Science Research Network, Rochester, NY.

Table A1: **O*NET Skill descriptions and examples of linked verb-noun combinations**

Skill group	O*NET skill and description	Example	Nr	Unique (%)	Total (%)
Basic content	Active Listening — Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.	Can listen Ask question Understand conversation	16	0.40	0.38
	Mathematics — Using mathematics to solve problems.	Make calculation Calculate "noun" Use measuring technique	119	2.94	2.59
	Reading Comprehension — Understanding written sentences and paragraphs in work related documents.	Read document Read communication Understand data	106	2.62	5.21
	Speaking — Talking to others to convey information effectively.	Have conversation Respond client Present idea	64	1.58	2.77
	Writing — Communicating effectively in writing as appropriate for the needs of the audience.	Make document Fill out form Use reporting skills	121	2.99	3.32
	Give information (general basic) – Giving or looking up information, that could either be in writing, speaking or else (self-added)	Use jargon Can inform Can communicate	88	2.18	3.11
	Language – Having knowledge of and efficiency in languages (self-added)	Speak English Use language skills Correct language	62	1.53	1.49
	Following instructions – Being able to follow instructions, rules and guidelines (self-added)	Follow instruction Use protocol Follow guideline	94	2.32	4.10
Basic process	Active Learning — Understanding the implications of new information for both current and future problem-solving and decision-making.	Use insight Apply knowledge Gather information	140	3.46	4.22
	Critical Thinking — Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to problems.	Analyse result Come up with solution Execute market research	96	2.37	2.03
	Learning Strategies — Selecting and using training/instructional methods and procedures appropriate for the situation when learning or teaching new things.	Apply learning process Determine educational needs Use learning strategy	6	0.15	0.10

Continued on next page

Table A1 – continued from previous page

Skill group	O*NET skill and description	Example	Nr	Unique (%)	Total (%)
Social	Monitoring — Monitoring/Assessing performance of yourself, other individuals, or organizations to make improvements or take corrective action.	Evaluate task Apply reflection skills Signal improvement	53	1.31	1.65
Social	Coordination — Adjusting actions in relation to others' actions.	Work colleague Maintain contact Can cooperate	57	1.41	2.23
Social	Instructing — Teaching others how to do something.	Can coach Use instruction skills Give directions	21	0.52	0.95
Social	Negotiation — Bringing others together and trying to reconcile differences.	Negotiate client Can negotiate Apply negotiation technique	11	0.27	0.31
Social	Persuasion — Persuading others to change their minds or behavior.	Can convince Give advice Use influencing technique	34	0.84	0.84
Social	Service Orientation — Actively looking for ways to help people.	Offer support Receive visitor Can give guidance	35	0.87	0.97
Social	Social Perceptiveness — Being aware of others' reactions and understanding why they react as they do.	Read room Dealing with aggression Apply observation technique	51	1.26	1.62
Technical	Equipment Maintenance — Performing routine maintenance on equipment and determining when and what kind of maintenance is needed.	Use protective equipment Maintain tool Apply maintenance	74	1.83	3.12
Technical	Equipment Selection — Determining the kind of tools and equipment needed to do a job.	Choose protective equipment Distinguish device Determine material	17	0.42	0.46
Technical	Installation — Installing equipment, machines, wiring, or programs to meet specifications.	Assemble part Can weld Install machine	109	2.69	2.54
Technical	Operation and Control — Controlling operations of equipment or systems.	Use tool Work system Apply ICT skill	240	5.93	6.83

Continued on next page

Table A1 – continued from previous page

Skill group	O*NET skill and description	Example	Nr	Unique (%)	Total (%)
	Operation Monitoring — Watching gauges, dials, or other indicators to ensure a machine works properly.	Read equipment Signal deviation Consult machine	30	0.74	0.82
	Programming — Writing computer programs for various purposes.	Apply programming language Create algorithm Use automation system	14	0.35	0.22
	Quality Control Analysis — Conducting tests and inspections of products, services, or processes to evaluate quality or performance.	Apply quality standard Check protective equipment Use testing equipment	103	2.55	2.88
	Repairing — Repairing machines or systems using the needed tools.	Perform repair Fixing error Repair car	45	1.11	0.77
Resource management	Management of Financial Resources — Determining how money will be spent to get the work done, and accounting for these expenditures.	Keep administration Prepare budget Make quote	60	1.48	1.62
	Management of Material Resources — Obtaining and seeing to the appropriate use of equipment, facilities, and materials needed to do certain work.	Apply safety Store material Make provisions	97	2.40	3.52
	Management of Personnel Resources — Motivating, developing, and directing people as they work, identifying the best people for the job.	Handle project management Apply leadership Motivate employee	45	1.11	1.06
	Time Management — Managing one's own time and the time of others.	Plan project Apply time management Make work schedule	36	0.89	0.85
Other	General (too general or too vague to be labeled to any of the above)	Can work Apply skill Execute technique	1378	34.07	27.72
	Specific (too specifically related to an occupation to be labeled to any of the above)	Clean animal Perform cashier operation Improve soil conditions	623	15.40	9.69
			4045		

Note: Nr. stands for number of unique verb-noun combinations retrieved. Unique (%) is the share of unique verb-noun combinations, and Total (%) represents the share of the full set of verb-noun combinations.