Learning the Right Skill: The Returns to Social, Technical and Basic Skills for Middle-Educated Graduates*

Femke Cnossen†
University of Groningen and GLO

Matloob Piracha
University of Kent and GLO

Guy Tchuente
University of Kent and GLO

October 2021

Abstract

Technological change and globalization have sparked debates on the changing demand for skills in western labour markets, especially for middle skilled workers who have seen their tasks replaced. This paper provides a new data set, which is based on text data from curricula of the entire Dutch vocational education system. We extract verbs and nouns to measure social, technical and basic skills in a novel way. This method allows us to uncover the skills middle-skilled students learn in school. Using this data, we show that skill returns vary across students specialized in STEM, economics or health, as well as across sectors of employment.

Keywords: Curriculum skills, vocational education and training, skill returns

JEL: J24, I26, I21

*The authors are indebted to Robert Inklaar, Milena Nikolova, Noemi Peter, Juliette de Wit, Olena Nizalova, Klaus Zimmermann, Christian Siegel, Steven Brakman, and participants from the GLO Virtual Young Scholar program, ESPE (2021), and SOM Conference (2021) for helpful comments and advice. We would specifically like to thank Max Boiten, Elwin Buisman and Welmoed Cnossen for support related to the text analysis, and the Stichting Beroepsonderwijs Bedrijfsleven for access to the curriculum data. Any errors are our own.

†Corresponding author. Contact via: `f.d.cnossen@rug.nl`

1 Introduction

Technology, globalization and structural change have worsened labour market outcomes of middle educated workers, and especially for the relatively recent entrants to the labour market. Even though this has been related to changing demand in skills and tasks, we know very little about the skills that middle educated students are currently learning in school, and the returns to those skills. 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 social, technical and basic cognitive skills. We develop a new measure of the relative weight of these three skills in curricula, by extracting text from the full set of training curricula in the Dutch middle education system. By linking the curriculum data to register data on wages for the graduates, we estimate skill returns on a fine-grained, curriculum-content level. To the best of our knowledge this is the first paper that uses the granulated skills level, using curriculum data, to highlight the importance of these skills for graduates entering the labour market.

The central notion behind this paper originates from the literature describing how technology has created poorer labour market trajectories of middle skilled workers. New technologies and offshoring have 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: 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.

Even though automation increases high skill job creation in the service sector, it reduces middle skill job creation in the manufacturing (goods) sector (Dauth, Findeisen, Suedekum, and Woessner, 2021). Students can adjust by either attending college to increase their skill level, or to take jobs requiring more abstract and less routine intensive tasks. However, this is not an option for all students. In the Netherlands, this is reflected by a decreasing probability of employment for middle-educated students, and especially if they are trained for routine professions (Bisschop, Zwetsloot, Ter Weel, and Van Kesteren, 2020; Ter Weel, Zwetsloot, and Bisschop, 2021).[1] Yet, the middle education sector in the Netherlands still accounts for over 40% of the student population after high school.[2]

[1]See also Reinhold and Thomsen (2017) who show that German students graduating from middle education have seen decreasing starting wages and slower wage growth than cohorts before the turn of the century.

[2]Source: CBS Statline, accessible via `https://opendata.cbs.nl/statline/#/CBS/nl/dataset/03753/table?`

Despite the decline in middle-skilled 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. Rather than a focus on one-dimensional skill, it makes a case for shifting towards an analysis of multi-dimensional skills. In this paper, we allow students from the same level of education to differ in the type of skills they have been taught in schools, in order to estimate skill-based wage inequalities in a novel way.

The extant literature shows a number of specific skills that are increasing in importance. The most prominent are the so-called people skills; there is growing evidence of relative employment and wage growth for occupations that require social skills (Borghans, Ter Weel, and Weinberg, 2014; Deming, 2017). These are jobs that require high levels of coordination, persuasion, negotiation, social perceptiveness, influencing, and decision-making (Felstead, Gallie, Green, and Zhou, 2007; Borghans et al., 2014; Deming and Kahn, 2017; Deming, 2021). Relevant to this paper, Deming (2017) highlights the importance of the combination of skills. More specifically, jobs where social skills are combined with high levels of cognitive skills have fared well, which is shown both using occupational task data (Deming, 2017) as well as using data from job postings (Deming and Kahn, 2017). In contrast, the opposite happened to high-math, low-social skills jobs (including many Science, Technology, Engineering and Math – or STEM – occupations).

However, most of the current literature on the technology-induced changes in demand for different types of skill focus on the higher-earning end of the labour market. For instance, the findings of Deming and Kahn (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, Ramdas, Sadun, and Fuller (2021) use only a sample of executive occupations to show the growing importance of social skills. Following Deming (2017), the growing importance of social skills in high-paying occupations might be explained by the fact that there is specific complementarity between cognitive and social skills. Therefore, solely improving middle-educated students' social skills cannot be seen as a panacea for potential substitution if it is not accompanied by other skills, or strong cognitive abilities.

Besides social skills, there is also evidence for a need of skills related to working with specific technologies. STEM graduates earn more than other graduates in the first years after graduation (Deming and Noray, 2018), which can be explained by the fact that STEM graduates have technology-specific skills that complement technologies in a unique way. It is still unclear whether this is caused by the actual skills students are learning, as there is little insight into the curricula of STEM students. Nevertheless, it is very plausible that the technological orientation of these programs increase complementarity between their taught skills and a labour market characterized

fromstatweb.

by fast technological change. And again, the main sample of these analyses are college-educated workers, whose patterns need not be the same for middle-educated graduates.

There is a large stream of literature on the changing demand for task-specific skills, with a number of methods of measuring this change in demand. Such measurements might involve: i. increases in the employment or wages of occupations with certain task-intensities (e.g. Autor and Dorn, 2013), ii. the growth or decline of certain activities within occupations (e.g. Spitz-Oener, 2006), iii. the change in return to specific tasks within occupations (e.g. Cnossen, 2021) or iv. the change in skills described in vacancies (Deming and Kahn, 2017). Yet, there is only a small number of papers that describe how skill acquisition at school affects labour market outcomes, and how different sectors vary in the relative labour demand for skills in recent 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. We merge the curriculum data to non-public microdata from Netherlands Statistics to obtain labour market outcomes of its graduates based on the skills taught. More precisely, we first measure skills from the curriculum text by extracting verb noun combinations from skill descriptions. We retrieve the underlying structure of the skills data using both exploratory factor analysis and a labeling of verb noun combinations to three classes of generic skills: social, technical and basic skills. Second, we link the skills data to the labour market outcomes of the 322,205 students that graduated in the period 2010 to 2018 in the programs included in our curriculum analysis.[3] We estimate Mincer (1974) equations of wages in the first years after graduation on our skill measures. We specifically differentiate between the three main fields of education (STEM, economics and healthcare), three middle-educated levels (ISCED level 2, 3 and 4), and apprenticeship-based tracks versus class-based training track.

We present three main findings. First, we find that graduating from relatively social-skill intensive degrees is negatively associated with wages in the first year after graduation while technical skills are associated with positive returns. Both relations persist until at least 10 years after graduation. Second, we show that demand for technical, social and basic skills differ strongly across fields, levels and tracks of education. For instance, students that graduate in the health-related field of education have higher returns to technical skills, as compared to STEM and economics graduates. Third, we show that wage returns to skill are conditional on the sector of employment: social skills are more strongly negatively associated with wages in the high skill service sector than in the low skill service sector. Our results imply that, within the same field of education, degrees focusing relatively more on social skills have lower wage returns. This does not necessarily mean that the demand for social skills is lower for middle-educated students. As our analysis concerns a study of curriculum texts, we show that a relative focus in the curriculum on social skills, rather than on other types of skills, does not positively affect wags. This could have potential implications for the construction of curricula.

[3]This is the sample of graduates for which we have data for all (control) variables. The actual number of graduates in the middle education system is slightly higher.

Even though education data tends to suffer from endogeneity due to students selecting in degrees based on ability and schools changing curriculum based on local labour market preferences, the nature of our data can partly circumvent these issues. We argue that under three, not highly restrictive assumptions, our estimates can be viewed as consistent. First, students select into degrees, and not in verb noun combinations. It is likely that they choose a field (e.g. STEM) or sub-field (Craft, Laboratory and Health Technology) based on their own interests and knowledge about abilities. However, upon entering a specific program, students do not know the exact contents of that curriculum, and do not know the differences between degrees within these (sub)fields. This assumption can be substantiated by empirical studies on how Dutch middle-educated students choose a degree: the contents of a curriculum are rarely the reason (Fouarge, Künn, and Punt, 2017). Since this is the source of variation in our data, we deem it safe to assume that conditional on choice of field, the skills acquired by students are exogenous. Second, we are aided by the fact that curricula are constructed on a national level: schools have little freedom in changing curricula based on local preferences. Combined with the fact that over 80% of students live with their parents, the set of local skills available are exogenous to the student, and due to the nationally oriented curricula, also to the local labour market.

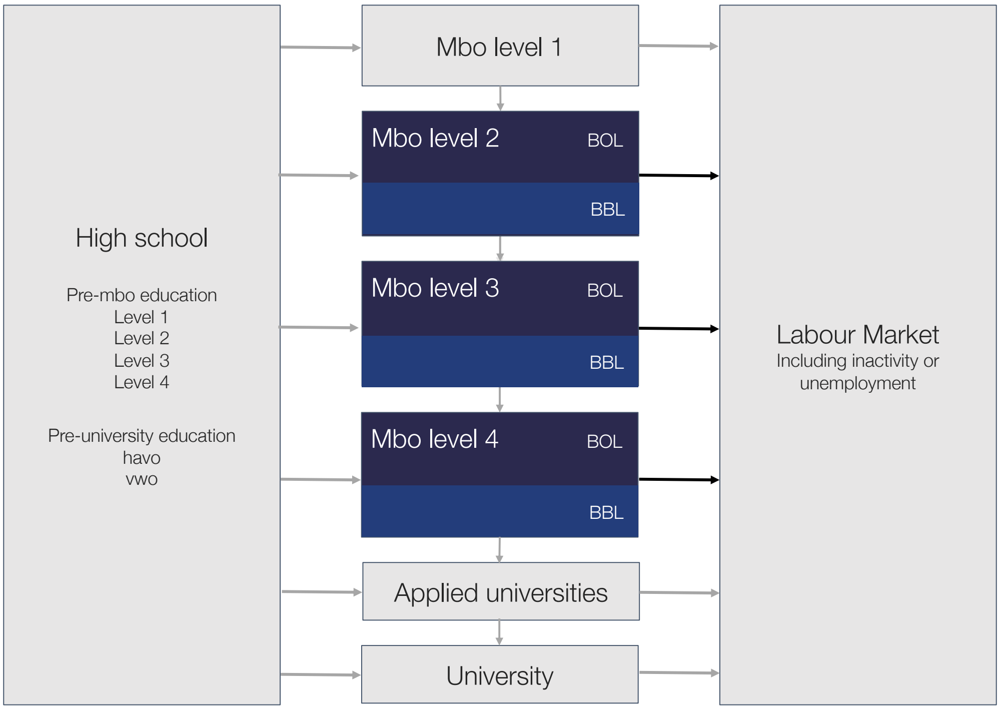
Unlike the existing literature on vocational and middle-educated schooling systems (e.g. Malamud and Pop-Eleches, 2010; Golsteyn and Stenberg, 2017; Hanushek, Schwerdt, Woessmann, and Zhang, 2017; Eggenberger, Rinawi, and Backes-Gellner, 2018), our focus is not on the specific versus general skills but rather the type of skills that students learn. More precisely, we distinguish between social, technical and basic cognitive skills (such as reading and mathematics) and estimate the returns to these skills.

In terms of methods, this paper fits into a growing literature in economics that uses text data as main source of information (Gentzkow, Kelly, and Taddy, 2019). For instance, Eggenberger et al. (2018) analyse the contents of training curricula of Swiss middle-educated graduates to measure skill specificity. Deming and Kahn (2017) use keywords in job postings to measure skill demand, while Hansen et al. (2021) map texts from occupational description 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. In this paper, we perform a content analysis of training curricula, similar to Eggenberger et al. (2018). We use verb noun combinations, similar to (Webb, 2019), and label these combinations in social, technical and basic skill categories, similar to Deming and Kahn (2017) and Hansen et al. (2021). Furthermore, we extend the analysis by estimating a factor model that retrieves the underlying structure of the extracted verb nouns.

2 Institutional background: the Dutch middle education system

The Dutch middle education (mbo) system is similar in nature to other education systems with distinct vocational education pathway (e.g., German, Swiss). It is oriented to provide vocational

Figure 1: Visual representation of the Dutch education system.



education training, where each major (or degree, or program) is linked to a profession. It is similar in nature to junior college education. Figure 1 shows the flow-chart of the Dutch education system. The students have a number of choices after they finish their primary education at the age of 12. Our focus is on the blue boxes, which are part of the Senior Secondary Vocational Education (mbo), in the middle of Figure 1.[4] After completing high school, students select into one of the four different levels, each more complex than the other, each having a broader and deeper bundle of skills than the other. Level 1 (the entry-program) is focused on acquiring basic learning and executive skills. This level does not lead to a starting qualification. 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 enroll in level 4 programs, that also last 3 to 4 years. This level covers middle management training, and prepare students on having leadership positions in sub-teams. Besides a higher difficulty in cognitive skills, these programs also focus more on responsibility, whereas level 2 and 3 are relatively more focused on foundational skills.

The entry requirements of the levels in the mbo are directly linked to the levels in the preparatory secondary education (pre-mbo levels 1 to 4; see Figure 1) or on previously obtained levels within the mbo. Once a degree is obtained, students are free to continue learning, or start earning. If they decide to continue learning, they have two options: i) stay in the same field, but move one level up (skill deepening), or ii) switch fields (skill broadening). The student receiving training

[4]Mbo is an abbreviation of the Dutch name, middelbaar beroepsonderwijs, for the middle vocational education system.

for employee fast service can thus choose to either start working, or train to become an assistant supervisor or switch fields. More than two third of the Level 1 students (69%) and around 60% of the Level 2 students continue with their studies while around 40% of the levels 3 and 4 students stay in education (Centraal Bureau voor de Statistiek, 2016). The percentage of students continuing studying to higher education has been declining over the past decade, from more than 40% to 35% for the most recent cohort (Centraal Bureau voor de Statistiek, 2018). In our analysis, we focus on the highest obtained degree in the mbo. Students who continue studying, for instance by going to an applied university, are excluded from the sample.

When selecting a degree, students can choose between two different pathways: either they opt for class-based training (BOL) or apprenticeship-based training (BBL). BOL has a focus on (theoretical) 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 (roughly 60% of training time). Both orientations lead to the same certification. In our analyses, we distinguish between these two groups of students, to see whether certain types of skills are more valued when taught in apprenticeships or in a school setting. For instance, it is likely that technical skills have higher returns when they are included in an apprenticeship training, as the extra practice with using technologies at work should increase the skill-level of these students.

Each level- and field-specific program has its own unique training curriculum: a qualification file. The Dutch Organisation for Vocational Training and Labour Market (S-BB) cooperates 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, 2021).[5] The total set of qualification files is the main data source for this project.

A last element of the Dutch secondary vocational education system, is that its schools are specifically oriented at their local labour market, which is why they are often referred to as regional education centers. Each school works in close cooperation with local companies that provide apprenticeships. Schools may choose which training curricula to offer, based on their own analysis of what their local labour market needs and what local students want to study. Students have a large degree of freedom in selecting their preferred training curriculum.[6] However, characteristic of the student population in the middle education system is that they tend not to move for schooling: 80% of students live with their parents during their entire degree (Fouarge et al., 2017). As a result, most students select a degree from the set of training curricula that their local school offers. Furthermore, there is empirical evidence that students hardly take into account the labour market prospects of each degree, but rather choose degrees based on personal interests and abilities, plus the opinions of their friends and families (Fouarge et al., 2017).

[5]The Dutch vocational education system is not characterised by having central examination, but rather a centrally described curriculum for each degree. Schools may decide for themselves how they examine the students in the prescribed learning goals and competencies from the curriculum. To ensure that the quality of examination does not create significant differences between the skill acquisition of graduates across schools, the Dutch Inspectorate of the Education closely monitors the teaching syllabuses and examination (Ministerie van Onderwijs, 2013).

[6]Note that the actual contents of each degree are decided at a national level.

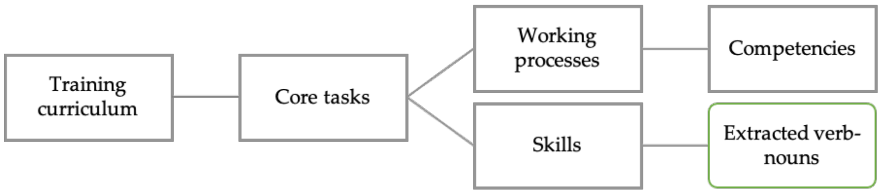
3 Empirical Strategy

The goal of this paper is twofold. We propose a new measure of granulated skills, and we estimate returns to social, technical and basic cognitive skills in middle educated graduates. In this section, we explain how we develop our measures of skill and how we use these measures for estimating returns to skills.

3.1 Measurement of Skill in Curricula

We obtain curriculum data from the SBB, which collects all qualification files for all degrees in the Dutch middle education system. Each curriculum consists of a list of core tasks, associated with the profession for which the student is trained. Each task is linked to a number of skill descriptions, which are deemed necessary for the execution of each task. We construct our skill measure based on these skill descriptions, see Figure 2.

Figure 2: **Composition of training curricula.**



Each curriculum consists of a list of skills, each a sentence long. The average number of sentences per degree is 18.74, with a standard deviation of 13.29. From these sentences, we extract verb noun combinations using a language programming module for Dutch, called Frog (Van den Bosch, Busser, Canisius, and Daelemans, 2007). We use two features of this module: first, Frog replaces all words in each sentence by its dictionary form (lemmatizing) and second, the Part of Speech (POS) tag function categorizes each word as either a noun, adjective, verb or other grammatical form. We keep all the verbs and nouns in each sentence, and use these two Parts of Speech to create a verb noun combination.

In terms of cleaning and preparing text data, we make a few selections. First, in order to make sure that the extracted verb noun combination relates to the main skill in each sentence, we only use main sentences, and delete all prepositional sentences. For example, any part of a sentence following prepositions like "such as", "among which", "for example", or "with respect to" are not part of the analysis. Second, we do not take into account sentences that start with "has knowledge of", which only retrieves 'has knowledge', and also captures knowledge-abilities rather than skills. Lastly, the word "is able to" ("can") is deleted, because it is the first verb in each sentence, e.g., "the student is able to ...". This cleaning process ensures that the verb noun combinations that are extracted from the skill sentences capture the central element of skill in each sentence.

Table 1 shows an example for all types of sentences used for the verb noun extraction. We let

Table 1: **Example of verb noun extraction, type of sentence, and frequencies of the type of sentences in entire data set**

Translated raw text data	Extracted verb noun combinations	Translation	Type	Freq.
Can apply safety requirements	(toepassen, veiligheidsvoorschrift)	(apply, safety requirement)	1 verb, 1 noun	40%
Can use (digital) relevant registration systems and ICT applications	(gebruiken, registratiesysteem) (gebruiken, ICT-toepassing)	(use, registration system) (use, ICT-application)	1 verb, 2 nouns	19%
Can interpret work order cards, drawings or models	(interpreteren, werkorderkaart) (interpreteren, tekening) (interpreteren, model)	(interpret, work order card) (interpret, drawing) (interpret, model)	1 verb, 3 nouns	9%
Can read and understand product information	(lezen,productinformatie) (begrijpen, productinformatie)	(read, productinformation) (understand, productinformation)	2 nouns, 1 verb	8%
Can cooperate and consult with colleagues and supervisors when upholstering furniture	(samenwerken, collega) (samenwerken, leidinggevende) (overleggen, collega) (overleggen, leidinggevende)	(cooperate, colleague) (consult, colleague) (cooperate, supervisor) (consult, supervisor)	2 verbs, 2 nouns	5%
Can read and interpret specifications, drawings and contract documents	(lezen, bestek) (interpreteren, bestek) (lezen, tekening) (interpreteren, tekening) (lezen, contractdocument) (interpreteren, contractdocument)	(read, specification) (interpret, specification) (read, drawing) (interpret, drawing) (read, contract document) (interpret, contract document)	2 verbs, 3 nouns	4%
			Total	85%

a verb noun combination exist conditional on the fact that there are at most two verbs and three nouns in the sentence. In this way 15% of sentences are deleted, because there is no clear match between the verbs and the nouns in that sentence. In all other cases, all verbs and nouns in the sentence are matched. The full list of extracted verb noun combinations contains 4450 unique sets. To refrain from overly profession-specific verb noun combinations, we restrict our sample to the verbs that exist in at least 5 study programs - which results in a sample of 482 verb noun combinations.[7]

One of the benefits of the Dutch language in this study is the fact that adjectives are often compounded with a noun, which creates more context for our analysis. An example is presented in the first row of Table 1: the extracted noun is "veiligheidsvoorschrift", which translates to "safety requirement". "Voorschrift" (requirement) is the noun in this case, but because of the added information from the compound, it becomes clear that this skill deals with a specific requirement related to safety ("veiligheid"). Similar cases are e.g. ICT-application ("ICT-toepassing"), contract document ("contractdocument") or work order card ("werkorderkaart"). These compounds add contextual information that allows us to improve the understanding of the skill described in each verb noun combination.

There are two options to impose structure on this type of skill data: manual labeling of verb noun combinations and exploratory factor analysis to uncover the underlying structure of the verb noun data. For the former, the researcher chooses to pre-impose a structure, by taking a list of skills currently used in the literature (Deming and Kahn, 2017; Deming and Noray, 2018; Hansen et al., 2021). The latter takes the entire list of skills, and through a factor analysis on the entire sample,

[7]We perform robustness analyses using a sample of 152 verb noun combinations that appear in at least 10 curricula, to test sensitivity of our results for sampling verb noun combinations that are more commonly used.

matches certain verb noun combinations to each other in a single factor. We adopt both measures here. Though true to the actual structure of the data set, the disadvantage of exploratory factor analysis is that the combination of elements within factors do not always immediately lead to an intuitive overarching theme, which manual labeling does.[8]

We use the labels for each verb noun combination to construct weighted skill measures for the three main categories. For each skill category, we create a dummy variable that we assign the value 1 if a verb noun combination can be labeled to that skill category (e.g. "cooperate colleague" is a social skill), and a 0 if not. Table 3 presents a list of examples, and how these verb noun combinations are matched to categories. We remove duplicates, such that a verb noun combination that might be mentioned multiple times in a curriculum, still counts as one verb noun combination. Using these inputs, we construct the relative frequency of these skills within the curriculum of each degree. The skill-frequency of skill s in degree j is then calculated as:[9]

$$\text{SkillFreq}_{sj} = \frac{\text{Number of verb noun combinations in degree } j \text{ assigned to skill } s}{\text{All verb noun combinations matched to degree} j} \quad (1)$$

For the factor analysis, we use a sample of 152 verb noun combinations. We have also performed the factor analysis using the list of 482 verbs, like above. However, for this sample we end up with 75 factors with an eigenvalue higher than 1, of which many factors have empty or too small (<0.4) loadings on verb noun combinations. Especially many of the later factors in this model are simply picking up degrees, rather than underlying structures of skills. Selecting a model with fewer factors did not solve this issue, as the first few factors contain the bulk of verb noun combinations and all other factors load on one or two verb noun combinations. This proved to be a poorer factor model than using the list of 152 verb noun that exist in at least 10 programs, rather than 5. A scree plot of a factor analysis for this sample is presented in Figure A1a and Figure A1b.

We choose a model of 36 factors, which is at the cut-off point in the screeplot where the eigenvalue is 1 (see Figure 5a). We rotate using oblique rotation, which assumes that underlying factors may correlate. Given that after a standard varimax rotation, there existed many correlations between the factors higher than 30%, the choice for oblique rotation is more appropriate.

To simplify terminology, for the labeling of each of these factors we again try to closely resemble O*NET occupational skill descriptions (National Center for O*NET Development, 2021). The main factors, their descriptions, and underlying verb noun combinations are presented in Table A1.

[8]However, manual labeling of verb noun combinations is more prone to researcher bias. Therefore, we also introduced four independent researchers to the data. We presented them the list verb noun combinations, plus a list of O*NET skill descriptions (National Center for O*NET Development, 2021). 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. As the O*NET skills fall into broader categories skills, we checked whether all connected O*NET skills from these researchers matched to the larger category (i.e. social, technical or basic skills).

[9]Where the maximum value for the denominator is 482: the sample size that contains the most common verb noun combinations extracted and labeled. In the robustness analyses, we also estimate regressions based on a larger sample of 152 verb noun combinations, that appear in at least 10 programs, to show our results are not sensitive to the inclusion of less common verb noun combinations.

There are four main categories of skills that emerge from our data, which do not necessarily capture all skills that O*NET describes. Most skills are technical in nature, such as “management of material resources" (factor 2), which captures elements such as maintaining materials, products and tools. Other skills are social, such as “coordination" (factor 1) or “persuasion" (factor 28). In the category of basic skills, factors emerge such as "reading and following instructions" (factor 8).

The skill-frequency measures from (1) and the 36 factors from the factor model are the main independent variables in our estimations on skill returns later on. The first measures the relative importance of certain types of skill in comparison to the total set of described verb noun combinations. The second tries to uncover underlying skills, by retrieving factors. The former gives more intuitive results in our wage estimations, whereas the latter is mostly useful for providing insight in the types of skills that are taught in different fields and levels of education, and how e.g. manufacturing degrees differ from health or economics degrees.

3.2 Estimation Strategy: Returns to Skills

In order to estimate the returns to skill, we run Mincer (1974) type regressions, where we regress hourly wages in the year after graduation on our skill measures. The goal of these analyses is to obtain a pattern of revealed skill demand for certain generic skills or competencies, and whether these skill returns differ across fields, levels or type of education. Furthermore, we aim to understand how these patterns might be explained by sector-sorting after graduation, by measuring whether students who have learned certain skills are more likely to be employed in certain industries.

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 major-choice might also correlate with labour market outcomes: diligent students are often diligent workers (Arcidiacono, 2004; Altonji, Blom, and Meghir, 2011). 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 hard evidence on the effect of a curriculum on labour market outcomes (Altonji et al., 2011). 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.

Even though returns to skill are never completely free from selection bias, we argue that the fine-grained nature of our text data circumvents part of the self-selection problem. Our reasoning is as follows: students will select into a higher hierarchical level (majors) than our observed data (verb noun combinations). It is also one level more disaggregated than e.g. courses, which would be more likely to be part of the student choice (Altonji et al., 2011). In our data, neither students nor schools have any influence over the curriculum requirements, as these are decided on a national level. Majors are thus essentially fixed bundles of skills.

The structure of the data and the institutional setting therefore allows us to partly circumvent selection issues. First, it is likely that students have a preference for learning either social or technical skills. Therefore, we should expect sorting into *fields* of education based on ability and comparative advantage in skills. However, given that students are unaware of the exact contents of the specific degree that they have chosen, our level of observation (verb noun combinations) is exogenous to the students. They cannot control the curriculum, as it is formed nationally, and they also cannot know the relative importance of social, technical and basic cognitive skills beforehand, as they do not access this type of information in their schooling choice. Combined with the fact that 80% of students do not move out of their parents' hometown, also makes sorting on ability less likely: students take the degrees presented to them at their local school as the only given options (Fouarge et al., 2017). Second, the nationally-oriented curriculum also makes it difficult for individual schools to change the curriculum based on their own preferences or local labour market demands. Furthermore, we can also be assured that the skills described in the curriculum will be tested at school, as schools are inspected on compliance with the national curriculum by the Ministry of Education. We can thus assume that each student will acquire the skills mentioned in the curriculum, and we can assume that students do not know in advance which skills they will learn precisely. As these verb noun combinations are the level of observation in our data, this should improve the reliability of our coefficients in light of potential self-selection issues.

Nevertheless, it might be the case that some skills are overrepresented in difficult programs, i.e. programs that cost more effort for students with low abilities (Deming and Noray, 2018). The correlation between the presence of a certain skill and the average ability within a degree might then be high, which would imply that the skill returns still reflect ability, and thus self-selection. This is something we cannot directly solve with the data at hand. However, this would only be problematic for our results if either all social, technical or basic verb noun combinations that are part of the skill-frequency equation (1) would be more difficult than all verb noun combinations in another category. In other words, it would be a concern if,for example, all social verb noun combinations are more difficult than all basic combinations. It is a reasonable assumption to make that this is not the case in our data, and therefore the skill frequency measure would not be a proxy for ability..

Still, the fact that there should be non-random selection of students into majors that might influence the returns to some verb nouns requires us to introduce a series of controls. First, we highlight the importance of including gender and immigration background in our estimations, as they are significant predictors of study field choice. Women and students with an immigration background are underrepresented in STEM related degrees, both in the Netherlands (de Koning, Gelderblom, Den Hartog, and Berretty, 2010) as well as in other countries (MacPhee, Farro, and Canetto, 2013). Reasons for this can be related to academic self-efficacy, where women and students with an immigration background tend to be less confident in finishing a STEM degree - irrespective of their actual academic performances (MacPhee et al., 2013). It has also been associated to group-related preferences or poorer labour market information (de Koning et al., 2010). Whichever the reason of sorting may be, we indeed observe large differences in the student population across fields: students with a migration background are overrepresented in economics degrees, women are overrepresented in health degrees, and both are underrepresented in STEM.

Furthermore, school-level sorting may be an issue in choice of majors. Arcidiacono (2004) shows that high-quality schools make lucrative majors more attractive, and, since high-quality schools attract high-ability students, they contribute to the ability sorting across majors - within school. However, there is little empirical evidence to believe that Dutch middle education students choose majors based on the quality of schools, since they mostly sort into their local school (Fouarge et al., 2017). However, it might still be the case that, within schools, students with certain abilities or preferences choose higher returning degrees. We would want to account for the fact that students within schools are therefore not independently and identically distributed (i.i.d). This is why we cluster all our estimations at the school level, which contains 73 clusters: one for each regional education center. Furthermore, we include school fixed effects, by adding school dummies in all our regressions.

Besides selection, another empirical question relates to the underlying cause of the relationship between skills and wages. If it is the case that technological differences have an effect on differences in skill demand, we should see different skill returns across sectors of employment. So far, 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).

The complementarity between skills and the tasks workers perform in certain sectors might then relate to wage differences between sectors, conditional on skills. To test whether skill demand is equal across the labour market, or whether different production structures require different skill inputs, we perform a few sector-specific analyses. Some skills might have higher returns in more in high-skilled service industries, manufacturing, or low skilled services. Therefore, besides the standard wage equations, we also estimate the relationship between skills and earnings conditional on being employed in a certain sector.

To summarize, 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 argue that our estimates can be viewed as consistent, under the following assumptions. First, students select into a field or domain of their preference, but the actual skill acquired in their training is exogenous to that decision. Students are not aware of the relative skill intensity in their programs upon entering a degree. Therefore, our measures of social, technical and basic skills, which are based on verb noun extractions, are exogenous to the sorting decision of the student. Second, we assume that the ability to acquire social, technical and basic cognitive skills is the same. In other words, technical skills are not necessarily less costly for more able students. Third, as each curriculum is constructed at a national level, we have exogenous variation in skill supply. Schools cannot control the curriculum, and thus cannot match the skill supply to skill demand in local labour markets through changes in the curriculum. This is an extra level of exogeneity that improves the consistency of our estimates.

4 Data

This paper relies on two main data sources: curriculum text data from the qualification files of the Dutch VET system and a linked employer-employee data set on earnings and employment. We obtain labour market data of graduates from non-public microdata from Netherlands Statistics. In this section, we present a brief description of the our constructed skills data, and we explain how we link the curriculum-level data to register data on wages and employment.

4.1 Curriculum Contents: what are students learning?

We can describe the contents of the Dutch curricula in vocational education, based on the skill frequencies and the factor analysis discussed above. Figure 3 presents the skill frequencies for the three categories over domains, which is one hierarchy lower than fields. As could be expected, programs in the STEM domains focus more on technical skills, whereas those programs in the field of economics and health are more oriented towards social skills. However, even within the fields we see variation: the STEM domain of Craft, Lab and Health technology is oriented more towards social skills than other STEM programs.

We are also interested in how curricula differ in the relative prevalence of the retrieved factors. These 36 factors are presented in Appendix Figure A2 by field of education, and a more detailed description over domain for a few selected factors in Figure A3. In terms of skill factors by sector, the first main distinctive feature of STEM programs is a stronger focus on technical skills. This is of course in line with the nature of these degrees, as they are mostly related to manufacturing and thus are taught more technical skills relative to the other sectors. The social skill that is disproportionally part of STEM curricula is 'coordination', which is a factor concerning verb noun combinations related to functioning in a firm, distinguishing one's specific tasks and responsibilities and being able to deliver reports in meetings.

Economics and healthcare graduates have a more pronounced focus on social skills. However, economics degree have high factor scores for 'conversation techniques', 'sales' and 'discussing calculations'. This last factor combines mathematics with cooperation skills. Furthermore, economics students are more likely to learn 'foreign languages' and 'complex problem solving'. Healthcare students are relatively focused on social skills related to managing teams. 'Management of Personnel Resources' captures elements such as feedback, reflection and observation skills, combined with instructing others, acting as a contact and recognizing problems. 'Management in an Office Setting' is more oriented towards meeting and presenting skills, and applying ICT. Apparently, both are dominantly present in such programs.

Figure 3: **Standardized skill frequencies, by domain**

Source: Authors' calculations using non-public microdata from Netherlands statistics. *Note:*Skill frequencies calculated following Equation (1). The domain 'Safety and Sports' has two bars, as it contains both economics-programs as well as health-programs.

Figure A3 outlines how factors differ across domains. The figure presents 4 factors: equipment maintenance (technical), conversation techniques (social), complex problem solving and interpreting manuals (both basic). From Figure A2 we could already observe that the technical skill of equipment maintenance is specifically dominant in STEM degrees, but here we see that this is mainly because the skill emerges in two domains: construction and process industry. On the other hand, the social skill of conversation techniques and both basic skills have more variation across domains, as well as fields of education. Some STEM domains use conversation techniques intensively, such as mobility and vehicles, but the skill is more often found in the economics degrees, such as administration, catering, and ICT.

4.2 Wage data

To estimate returns to skills we link the curriculum data to non-public microdata from Netherlands Statistics on graduation, earnings and employment of students enrolled in these programs. We are able to link each graduate to their respective degree, through unique degree-codes.[10]

We construct a linked employer-employee data set using various data sets from the Dutch microdata: wage data in the years after graduation, demographic characteristics (gender, age, migration background), firm-level data to determine the industry of employment, and data on enrollment and graduation from middle and higher education. We use students that graduated between 2010 and 2018 in one of the 333 training curricula.[11] Some of these students have been through multiple programs, of which we select the most recent degree. Furthermore, we check whether students did not enroll in a new study program in either middle education or higher education. Those that continued studying and are either still in education in 2018 or have graduated in higher education afterwards are removed from the sample.

The wage data consists of monthly information on hours worked, gross and net wages, type of contract[12], plus industry and employer. We construct an hourly wage indicator, which is the average hourly wage across multiple jobs in case a worker has more than one job.[13] We take the yearly average of this hourly wage in the year after graduation as our main dependent variable. Besides wages, the microdata also provides information on the industry of employment.[14] We only

[10] Here we use data on graduates (mbo gediplomeerden). This contains data on crebo codes (the Central Register on Professional Education), which we use to link curriculum text information to graduates.

[11] The full set of curricula contains 500 programs, but we remove programs from the agricultural sector (as these contain a small number students per program) and programs from level 1, as they do not lead to a basic qualification. Furthermore, we only keep full time degrees.

[12] This can be either tenured or temporary, where a tenured contract applies to workers with a contract for an indefinite period, plus interns, directors/major shareholders ('directeur-grootaandeelhouder or dga in Dutch), and people employed under the Sheltered Employment Act (wsw in Dutch). Temporary contracts apply to temporary employees, sub-contracted or on-call employees (uitzendkracht and oproepkracht in Dutch, respectively)

[13] This variable is computed using data on number of paid hours and total wage in monetary value, which excludes parts of the wage bill that are transferred in tangible ways, such as lease cars or company lunch.

[14] The industry of employment is obtained by matching the employer-identifier we obtain from the POLISBUS data to firm-level register data (ABR - Algemeen Bedrijven Register), which contains industry data. The industry classification used by Statistics Netherlands is the SBI, which has the same first digits as the NACE classification used by the European Union.

use the first digit industry code, which we use for our analysis on the differential demand for skills across sectors. We divide industries into three sectors following (Bárány and Siegel, 2020): low skilled services (LSS), goods and high skilled services (HSS). Table A2 describes the classification of industries into sectors, as well as the relative employment and standardized skill frequencies across industries.

Table 2 provides descriptive statistics of our main sample, and for subsamples by gender and general field of education. In total, we have 322,205 students in our sample that have information on all control variables, of which most graduated in economics (40%), followed by health (36%) and STEM (24%). Hourly wages are highest for STEM graduates, and lowest for economics graduates. Furthermore, STEM students tend to be somewhat older upon graduation. In terms of gender distribution, STEM programs are highly skewed towards male students, with only 19% female students, whereas 81% of Health graduates are female. Economics degrees have the highest share of students of non-Dutch descent. Lastly, females - and thus health degrees - are over-represented in level 4 programs, whereas STEM has the highest share of level 2 students. Given that STEM is a male-dominated field, this is also reflected in the relative share of men in level 2 programs.

Table 2: **Descriptive Statistics**

	By gender		By field of study			By sector of emp.			
	(1) Total	(2) Male	(3) Female	(4) STEM	(5) Econ	(6) Health	(7) LSS	(8) Goods	(9) HSS
Wage									
Log hourly, $t+1$	2.23	2.24	2.22	2.29	2.16	2.27	2.17	2.28	2.29
sd	(.42)	(.40)	(.44)	(.37)	(.41)	(.46)	(.36)	(.42)	(.47)
Demographics									
Age	2.63	2.71	2.55	2.88	2.43	2.68	2.33	2.70	2.89
sd	(2.09)	(2.11)	(2.07)	(2.07)	(2.01)	(2.17)	(1.82)	(2.19)	(2.27)
Female share	.51	0	1	.19	.44	.81	.51	.17	.58
Dutch share	.75	.76	.73	.83	.67	.77	.75	.89	.72
Level									
Level 2	.26	.31	.22	.32	.26	.21	.27	.35	.24
Level 3	.20	.21	.19	.15	.24	.19	.19	.24	.20
Level 4	.54	.49	.60	.53	.50	.60	.54	.41	.57
Field									
STEM	.24	.40	.09				.21	.68	.18
Economics	.40	.46	.35				.50	.23	.35
Health	.36	.14	.57				.29	.10	.47
Track									
Class-based (BOL)	.81	.75	.87	.67	.82	.90	.82	.58	.85
Apprentice (BBL)	.19	.25	.13	.33	.18	.10	.18	.42	.15
Observations	322,205	156,988	165,217	76,619	129,972	115,614	143,404	28,430	150,371
Share	1	.49	.51	.24	.40	.36	.45	.09	.47

Source: Authors' calculations using non-public microdata from CBS. LSS/HSS stands for low/high skill services.

5 Returns to Curriculum Skills

As a first descriptive step in the analysis of the data we perform individual regressions on the entire set of verb noun combinations on wage in the first year following graduation. Wage is the one-year average of log hourly wage, for the first full year after graduation. In Table 3 we report the ten verb noun combinations associated with the highest (positive and) significant coefficient for different subsamples.[15] The ranking of verbs is based on regressions with the inclusion of one verb noun combination at a time, with a full set of controls and standard errors clustered at the school level.

For the entire sample, the top 10 verb noun combinations contain three social skills, related to having conversations, either in Dutch, English or a different language. There are three skills related to working with certain (ICT) applications or machines, the remaining four verb noun combinations relate to interpreting specifications, making analyses, writing in English and applying principles.

We see variation in the returns to verb noun combinations between the two tracks in the mbo: class-based (BOL) versus apprenticeship-based (BBL). In the latter, students learn most of their skills on the job, whereas students in the class-based track spend more time learning skills in the classroom. Interestingly, the apprenticeship-based track sees higher returns for learning social skills, such as discussing with colleagues, sales techniques and giving feedback, whereas the top 10 verb noun combinations in the class-based track are less socially oriented.

The pattern for these returns also differs strongly across STEM, economics and health programs. The most obvious differences are in the importance of social and technical skills. Especially in health programs, students see higher returns on social skills (highlighted in the darkest shade). Interestingly, these are not necessarily social skills directly related to typical healthcare activities, but more in social skills relating to managing and cooperating in teams: presenting, meeting, sales and negotiation activities. For STEM programs, technical skills (i.e. using certain tools, equipment or applications) are naturally more strongly correlated with wages, but also specific social skills regarding conveying information and taking up a role as a contact in professional settings. Only in STEM programs, math skills are part of the top verbs. For economics programs, the top correlating verb noun combinations are more related to reading skills or processing information. Furthermore, whereas the highest-returning social skills in health curriculums are related to teamwork, we see that the type of social skills that correlate most strongly with economics students are related to having conversations. Even though some skills might be social or technical in nature, different types of each of the three skills are relevant for different fields of education, e.g., social skills relevant for the STEM degree may not be the same as those in, say, health.

Next, columns 5 to 7 show the same results for subsamples of the three levels of middle education. Level 2 students have higher returns to quite general verb noun combinations, related to applying skills, tasks and rules. This would be in line with the fact that level 2 degrees are more related to acquiring basic (learning) skills, rather than specialized technical or social skills. For

[15] This type of analysis is inspired by a similar analysis of measuring the importance of certain tasks in metropolitan areas over time Michaels, Rauch, and Redding (2019).

Table 3: Verb Noun Combinations that Positively Correlate Most Strongly with Hourly Wages in one Year after Graduation

Rk	(1) Full sample	(2) Class-based (BOL)	(3) Apprenticeship-based (BBL)
1	Converse English	Contribute	Use ICT application
2	Write English	Apply equipment	Execute calculation
3	Have conversation	Apply machine	Discuss colleague
4	Use equipment	Read English	Apply skill
5	Apply principle	Read assignment	Apply guideline
6	Use language	Apply quality requirement	Apply sales techniques
7	Use ICT application	Apply safety requirement	Understand information
8	Interpret specifications	Converse English	Make calculation
9	Use application	Work system	Give feedback
10	Make analysis	Write English	Apply firm

Rk	(4) STEM	(5) Economics	(6) Health
1	Read information	Read manual	Apply sales techniques
2	Use tool	Apply task	Apply ICT application
3	Act as contact	Apply meeting technq	Apply meeting technq
4	Follow guideline	Apply skill	Apply ICT skill
5	Use equipment	Interpret drawing	Have insight
6	Make calculation	Apply ICT skill	Contribute
7	Keep records	Gather information	Apply presenting technq
8	Make analysis	Have conversation	Apply skill
9	Use application	Write English	Give feedback
10	Convey information	Converse English	Use skill

Rk	(7) Level 2	(8) Level 3	(9) Level 4
1	Read text	Work equipment	Apply equipment
2	Apply rules	Write English	Apply machine
3	Apply task	Apply principle	Contribute
4	Apply skill	Use language	Use ICT application
5	Write English	Have conversation	Have conversation
6	Contribute	Keep records	Converse English
7	Apply information	Use material	Use tool
8	Work equipment	Use application	Interpret drawing
9	Converse English	Converse English	Write English
10	Act as contact	Use equipment	Read information

(dark grey)	Technical
(light grey)	Social
(very light grey)	Basic

Note: Ranking of coefficients, estimated from a regression on the average hourly wage in the year after graduation on a dummy for whether a verb noun combination is part of the curriculum. A separate regression is estimated for each verb noun combination. Controls are: level, field and track of education, age, gender, and school dummies. The verb noun combinations included in the analysis appear in at least 10 programs, such that positive coefficients are not implicit degree-returns.

levels 3 and 4, we see that more specific technical and social skills seem relevant, where especially level 4 students have relatively higher returns to verb noun combinations related to using certain machines, tools or ICT applications. In both cases, both levels 3 and 4 students' have highest returns to social skills related to having conversations, either in Dutch or in another language.

However, using each individual verb noun combination in itself we cannot conclude whether social, technical or basis skills have different returns across subgroups of students. Therefore, in the next section we use the relative skill frequency measures of social, technical, and basic skills to indicate whether some fields or levels of education can benefit more from certain generic skills.

5.1 Returns to Skill Frequencies

To further dive into the types of skills that are rewarded upon entry in the labour market, we estimate a Mincer (1974) equation of wages on the relative frequency of skills in the curriculum. The outcomes show whether curricula that are relatively more focused on certain skills, e.g. social skills, result in better starting wages in the first year after graduation.

Given the considerations outlined before, our empirical strategy is as follows. Using the skill frequencies as constructed following equation (1), we estimate the following regression:

$$\ln w_{ijt+1} = \beta_0 + \beta_1 social_j + \beta_2 tech_j + \beta_3 basic_j + X_i\gamma + Z_j\alpha + \pi_t + \varepsilon_{ij} \tag{2}$$

where X_i contains a vector of the demographic controls gender, migration background (first or second generation migrant, and place of origin[16]), and age at graduation. Z_j is a vector containing degree-related controls: field of education (STEM, economics and health), level of education (2, 3 or 4), and track (class-based or apprenticeship-based). We include time dummies π_t as the year of graduation.

Besides a baseline regression, we also add interactions between the three skills and i) the levels of education, ii) fields of education and iii) class-based (BOL) versus apprenticeship-based (BBL) tracks. The results in Table 4 show how skill-demand differs for students from varying schooling backgrounds. Note again that the sample of students is restricted to those that decided to move to the labour market after graduation from middle education. Students that continued education or have obtained a degree in higher-education following their middle education degree are excluded from the analysis. As such, the results show the returns to skills, conditional on the decision of students to not continue studying.

The results show the following pattern. First, the returns to social-skill intensive curricula are negative across all estimations - even when including interactions. Technical skills are not significantly positive in the baseline estimation, but become positive and significant in all three estimations that include interaction terms. This highlights that especially technical skills are not

[16]This is a variable taking seven options for the most common (migration) backgrounds. These countries, and there approximate share in the total population of mbo-students in 2015, are: Netherlands (73%), Turkey (4%), Morocco (4%), Suriname (4%), Dutch Caribbean (2%), Western (6%), and other non-Western (6%). Source: CBS Statline.

Table 4: **OLS Regressions of Log Hourly Wage in First Year after Graduation on Skill Frequencies in Curriculum and Interactions with Degree Characteristics**

	(1)		(2)		(3)		(4)
Social	-0.019***		-0.024***		-0.017***		-0.020***
	(0.002)		(0.003)		(0.005)		(0.002)
Tech	0.016***		0.008*		-0.012**		0.014***
	(0.004)		(0.004)		(0.005)		(0.004)
Basic	-0.002		0.004		-0.013**		-0.000
	(0.003)		(0.004)		(0.006)		(0.003)
Ref. cat.: Level 2							
Level 3	0.108***		0.106***		0.076***		0.108***
	(0.006)		(0.006)		(0.006)		(0.006)
Level 4	0.238***		0.238***		0.189***		0.237***
	(0.006)		(0.005)		(0.006)		(0.006)
Ref. cat.: STEM							
Econ	-0.045***		-0.043***		-0.069***		-0.048***
	(0.005)		(0.005)		(0.005)		(0.004)
Health	0.075***		0.075***		0.113***		0.072***
	(0.008)		(0.008)		(0.012)		(0.007)
Female	-0.035***		-0.035***		-0.042***		-0.036***
	(0.003)		(0.003)		(0.003)		(0.003)
BBL	0.182***		0.181***		0.167***		0.176***
	(0.008)		(0.008)		(0.008)		(0.008)
		Skill frequencies interacted with					
		Level		Field		Track	
		Ref.: Lvl 2		*Ref.: STEM*		*Ref.: BOL*	
		Social × Level 3	0.031***	Social × Econ	0.020***	Social × BBL	0.015***
			(0.005)		(0.005)		(0.006)
		Social × Level 4	-0.009*	Social × Health	-0.000		
			(0.005)		(0.006)		
		Tech × Level 3	0.013*	Tech × Econ	0.005	Tech × BBL	0.012**
			(0.007)		(0.006)		(0.006)
		Tech × Level 4	-0.000	Tech × Health	0.226***		
			(0.006)		(0.012)		
		Basic × Level 3	0.021**	Basic × Econ	0.028***	Basic × BBL	-0.016**
			(0.009)		(0.007)		(0.007)
		Basic × Level 4	-0.021***	Basic × Health	-0.056***		
			(0.006)		(0.012)		
Constant	0.770***		0.773***		0.848***		0.781***
	(0.027)		(0.027)		(0.025)		(0.028)
Obs	322,205		322,205		322,205		322,205
R-squared	0.233		0.234		0.243		0.234

Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level. Each regression includes the full set of demographic, degree, year, and school controls. BOL stands for class-based track, BBL for the apprenticeship-based track.

necessarily valued in each and every curriculum, but that there are level- and field-specific effects at work. Basic cognitive skills are only significant once interactions with levels of education are included in the model.

Next, the bottom panel of the table shows the interactions between levels (column 2), fields (column 3) and tracks of education (column 4). We find little evidence for strict linearity in the relation between skills and levels of education. Social skills are relatively valued more in level 3 programs than the reference category level 2, but are not significantly different from level 4 students. Technical and basic skills are valued most in level 2 students, as indicated by the negative coefficients for level 3 and 4 for both skills. Basic skills have lower returns in level 4, than level 3. Each of these three types of skill thus interact differently with varying levels of education. These results can either hint towards the existence of ability-biased skill returns: some types of skills are mostly valued in combination with certain levels of education. Alternatively, returns to some skills might be higher if they are relatively scarce in degrees of a certain level. This might explain why technical skills have the highest returns for level 2 graduates, whose curricula are relatively more focused towards foundational skills.

Column (3) shows the results with the inclusion of field-skill interactions. Social skills have no field-specific returns. Technical skills are valued more in STEM than in economics programs, and technical skills seem to be especially valued more for health care students, indicated by the coefficient of 0.14. In this case, increasing technical skills in a curriculum thus works best for programs in the health care sector, whereas economics students are more benefited by increasing basic cognitive skills as part of their curriculum.

Lastly, column (4) shows the interactions with tracks: students who follow the apprenticeship track (BBL) have higher returns to social and technical skills than those in the class-based track (BOL). The increase in wages for BBL-students in social skills even offsets the negative general coefficient of -0.018, implying that social and technical skills both significantly relate to higher wages for students in this track. On the other hand, basic skills have lower returns in students of the apprenticeship track. Therefore, the manner in which skills are taught apparently influences the size and sign of the returns to certain types of skills. Social and technical skills relate to higher returns if they are learned by spending relatively more time practicing these skills on the job, whereas basic cognitive skills result in higher returns when learned in class.

5.2 Decomposing by Sector of Employment

So far, we have estimated the returns to skill based on the schooling background of each student: their field, the skills required and the level of education. In this section we decompose the results by sector of employment. We re-estimate equation (2) conditional on being employed in either the low skilled services, goods or high skilled service sector. We control for the field and level of education, such that the results obtained here should not capture matches between the sector of education and the sector of employment. The results are presented in Table 5.

Table 5: **OLS Regressions on Skill Frequencies in Curriculum, by Sector of Employment**

	(1) Low-skilled services	(2) Goods	(3) High-skilled services
Social	-0.000	-0.010*	-0.035***
	(0.001)	(0.005)	(0.002)
Tech	0.003	-0.025***	0.031***
	(0.002)	(0.005)	(0.006)
Basic	0.001	-0.018**	0.000
	(0.002)	(0.008)	(0.003)
Ref. cat.: Level 2			
Level 3	0.077***	0.119***	0.107***
	(0.005)	(0.007)	(0.009)
Level 4	0.198***	0.249***	0.246***
	(0.005)	(0.010)	(0.008)
Ref. cat.: STEM			
Services	-0.047***	-0.130***	-0.001
	(0.004)	(0.015)	(0.007)
Health	-0.008	-0.061***	0.163***
	(0.005)	(0.019)	(0.013)
Ref. cat.: BOL			
BBL	0.116***	0.273***	0.227***
	(0.006)	(0.009)	(0.012)
Female	-0.024***	-0.066***	-0.048***
	(0.002)	(0.008)	(0.004)
Constant	0.502***	0.755***	1.060***
	(0.034)	(0.023)	(0.031)
Observations	143,404	28,430	150,371
R-squared	0.260	0.341	0.212

Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level. Each regression includes demographic, degree (track, level and field), year, and school controls, and a constant. BOL stands for class-based track, BBL for the apprenticeship-based track.

Intuitively, social-skill intensive curricula should result in high returns in service sectors - that generally use social skills more intensively. However, this seems not to be the case for middle-educated graduates. The results in Table 5 show that socially-oriented curricula are associated with lower returns in the high skill services sectors, and insignificant in the low-skilled services sector. In the goods sector, the return to social skills in curricula is also negative, though it is slightly smaller and less significant.

On the other hand, the returns to technical skills are positive in the high-skilled services sector, yet negative in the goods sector. This is an interesting result, and could possibly be explained by constrained supplies of technical oriented students in the high-skilled services sector, whereas these skills are complementary to the tasks executed in this sector. Regardless, there seem to be sector-specific factors that influence skill-based pay differentials.

5.3 Robustness

We perform a few robustness tests to our baseline results from Table 4. First, we test the sensitivity of our results to the in- and exclusion of degree-specific controls. The estimations are presented in Table 6. First, we would like to test whether our results remain consistent over the inclusion and exclusion of level, field and sub-field controls. Column (1) shows the baseline estimation, as also presented in Table 4. In Column (2), field controls are removed, and in column (3) level controls are removed. The results on social skills are highly robust to the exclusion of these variables. The estimates of technical skills also remain positive and significant, but appear to be more sensitive to the exclusion of field and level controls. Next, in column (4) we include sub-field controls, beyond field controls. These sub-fields (or domains) are also presented in Figure 3. We see that the effect sizes of both technical and social skill increase in size. Our results are thus robust to the inclusion of more fine-grained field controls. Social (technical) skills are thus robustly and positively (negatively) associated with wages in the first year after graduation, even within disaggregated fields of education.

Lastly, we test whether our results are robust to the sample of verb noun combinations used to construct the skill frequency measure. In the baseline estimations, we use the constructed measures based on 482 verb noun combinations. This sample was selected on the condition that they should appear in at least 5 programs. We recreate this measure based on 152 verb noun combinations that appear in at least 10 programs. To check whether constructing the skill measure based on more common words result in different estimates, we rerun the baseline equation. The results are presented in Column (5). We can see that the results for social skills are robust for the sample size. Technical skills are insignificant, which is most likely due to the fact that specific technologies are mentioned in fewer programs. Therefore, the positive estimate for technical skills in our baseline estimates are likely to be lower-bound estimates, given that they could include even more technology-specific skills if we would include more verb noun combinations in our sample. Basic cognitive skills remain insignificant.

Table 6: **OLS Regressions on Skill Frequencies in Curriculum, by Sector of Employment**

	(1) Base-line	(2) No field controls	(3) No level controls	(4) Add domain	(5) Skill sample
Social	-0.019***	-0.013***	-0.011***	-0.024***	-0.016***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Tech	0.016***	0.008***	0.040***	0.034***	-0.001
	(0.004)	(0.003)	(0.005)	(0.005)	(0.002)
Basic	-0.002	-0.017***	0.048***	0.005	-0.002
	(0.003)	(0.002)	(0.003)	(0.004)	(0.002)
Level 3	0.108***	0.105***		0.123***	0.110***
	(0.006)	(0.006)		(0.005)	(0.006)
Level 4	0.238***	0.246***		0.247***	0.235***
	(0.006)	(0.006)		(0.006)	(0.006)
Services	-0.045***		-0.009	0.186***	-0.058***
	(0.005)		(0.008)	(0.021)	(0.005)
Health	0.075***		0.140***	0.090***	0.054***
	(0.008)		(0.015)	(0.021)	(0.005)
Female	-0.035***	-0.007**	-0.029***	-0.030***	-0.033***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Constant	0.770***	0.704***	0.682***	0.811***	0.775***
	(0.027)	(0.029)	(0.034)	(0.034)	(0.027)
Observations	322,205	322,205	322,205	322,205	322,205
R-squared	0.233	0.222	0.194	0.244	0.232

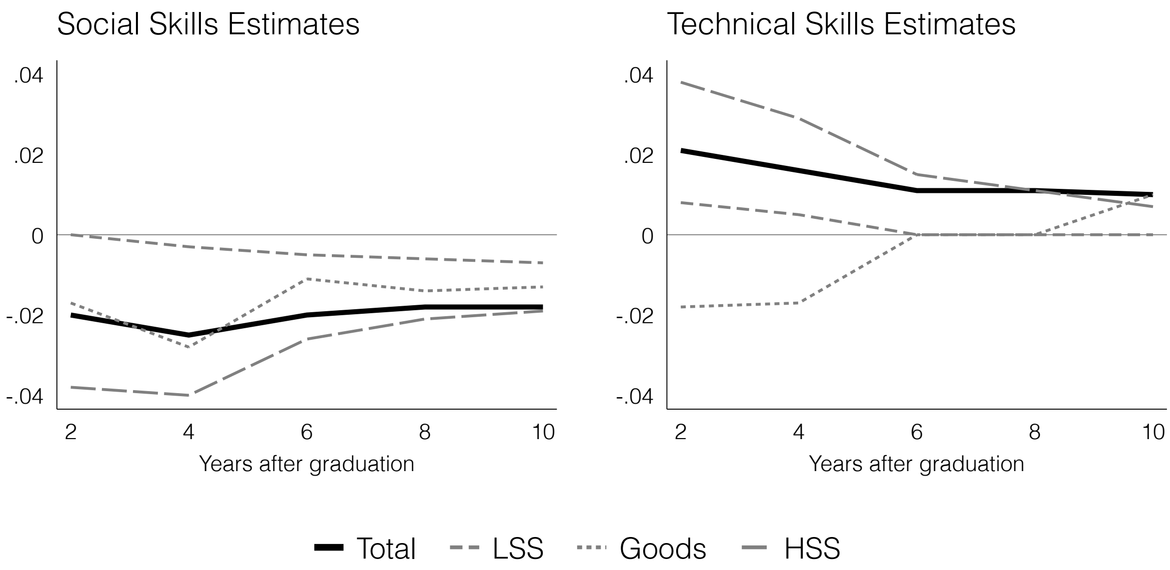
Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level. Each regression includes demographic, degree (track, level and field), year, and school controls, and a constant.

5.3.1 Skill returns in several years after graduation

Since the results in the baseline estimations are all for only one year after graduation, we also run analyses for wages later in the careers of the students: 2, 4, 6, 8 and 10 years after graduation. The results are presented in Appendix Table A4. The coefficients of the full sample, as well as split out by sector of employment, are graphically presented in Figure 4.

We find that our results are not only an artifact of the first year after graduation, but remain persistent in the following years as well. The coefficient of social skills remains significantly negative over time. It is most strongly negative in the high skill services sector, but this effect decreases over time. The same holds for the positive effect of technical skills, this also decreases in importance over time. This is sensible, as it is likely that the impact of skills taught in school reduce over the period after school, where learning on the job becomes increasingly relevant.

Figure 4: **Coefficients of Returns to Social and Technical Skills in Multiple Years after Graduation**



Note: Coefficients of regressions presented in Table A4 with full set of controls. Insignificant estimates are valued at 0. LSS/HSS stand for low skill service sector and high skill service sector respectively.

5.3.2 Skill combinations

Existing literaure (see, for instance, Deming (2017) and Deming and Kahn (2017)) highlights the fact that combinations between skills explain pay differentials between workers. To test whether we can find similar evidence, we rerun equation (2) with the inclusion of interactions between the three skills. We perform this exercise for the full sample, but also for subsamples by sector of employment. The results are presented in Appendix Table A3.

We find little evidence for skill complementarities in curricula: there are no positive interactions between any of the three skills. This does not need to imply that these skill complementarities do not exist for middle educated workers. However, we find no evidence that including both in a curriculum positively affects starting wages.

5.3.3 Returns to skill factors

The results on social, technical and basic skills highlight that different fields, levels and tracks result in different skill returns. However, the data allow us to go one level deeper, to find out which types of social, technical of basic skills are valued more in these different fields. To do this, we describe the results that use the 36 factors from our factor model.

We estimate the following regression:

$$\ln w_{ijt+1} = \beta_0 + \sum_{f=1}^{F} \beta_f f_j + X_i \gamma Z_j \alpha + \pi_t + \varepsilon_{ij} \tag{3}$$

For F containing 36 factors f on the degree level. The results are presented in Table 7.

Column (1) shows the baseline estimation of (3), columns (2) to (4) are estimations for subsamples of graduates employed in the low-skilled services, goods or high-skilled services sector. The factors are split up by the type of skill: technical, social and basic skills. Below, we only discuss positive correlations. As each skill factor might have an intrinsic value in and of itself, we refrain from the discussion of negative factors in this section. Note that all factors are standardized across the entire sample. For a list of the descriptions of these factors, see Table A1.

Technical factors that positively correlate with wages are Equipment Maintenance (f5), Management of Maintenance Equipment (f6), Operation and Control (f7), Installation (f10), Following Technical Regulations (f17) and Using Applications (f34). Social factors with positive returns are Coordination in Firm (f1), Management of Human Resources (f4), Conversation Techniques (f26). For basic skills, only Interpreting Manuals (f21) and Following Regulations (f25) are significant and positive.

The results presented here hint towards variation in sector-skill complementarities, and call for further research into understanding sector-specific skill demand. A broad perspective on social skills would overlook the fact that some social skills are more associated with high wages in the high and low skill sector (e.g. Conversation techniques (f26)), whereas others have positive wage returns in the goods sector (e.g. Coordination in Firm (f1)).

6 Conclusions

This paper uses novel data on skills to estimate returns to specific elements in the curricula of Dutch middle-educated students. We created skill measures based on the relative frequency of social, technical and basic cognitive skills mentioned in the curriculum. We performed two main exercises with this data. First, we showed how skill supply differs between sectors and levels of education, in terms of what is being taught in the curricula. Second, we linked the curriculum contents to wage data in the first years after graduation, to estimated returns to granulated skill levels.

We found that students graduating from degrees with a stronger focus on social skills have lower wage outcomes in the first years after graduation, even after controlling for specific fields of education. A decomposition analysis by sector shows that this seems to be mainly driven by returns in the high skill services sector. Given the nature of our data, we believe our estimates have little bias, since students select into degrees, not into verb noun combinations, and since the curricula are constructed at a national level. This makes our skill measure of observation relatively exogenous to the student choice.

Table 7: **OLS Regression of Log Hourly Wage in First Year after Graduation on 36 Factors of Verb Noun Combinations**

		f	(1) Full	(5) LSS	(6) Goods	(7) HSS
Technical	Mgmt of Material Resources	2	**-0.020***	0.011	-0.002	**-0.052*****
	Quality Control Analysis	3	**-0.003***	-0.004	**-0.007****	**0.016*****
	Equipment Maintenance	5	**0.034*****	**0.028*****	**0.016*****	**0.041*****
	Mgmt of Maintenance Equip	6	**0.050*****	**0.040*****	**0.021*****	**0.078*****
	Operation and Control in Manuf	7	**0.022*****	0.002	**0.011***	0.013
	Installation	10	**0.009****	**0.012*****	**0.017*****	0.011
	System Quality Analysis	11	-0.001	-0.013	-0.013	**0.047****
	Installation in Construction	15	**-0.014****	0.004	**0.016****	**-0.050*****
	Following technical regulation	17	**0.008*****	0.002	-0.005	**0.011*****
	Apply Quality Procedures	22	**-0.010****	-0.004	**-0.015****	**-0.014*****
	Coordination in Operations	23	**0.015***	-0.013	-0.001	0.012
	Using Tools	27	**-0.017*****	-0.006	0.004	**-0.037*****
	Tool Selection	29	**-0.009*****	**-0.010*****	**-0.017***	**-0.010****
	Reading Equipment Instructions	31	**-0.010*****	-0.003	0.003	**-0.013*****
	Using Applications	34	**0.006****	0.004	-0.008	-0.002
	Using Systems	35	**-0.010*****	-0.001	**-0.017****	**-0.018*****
Social	Coordination in Firm	1	**0.015*****	-0.003	**0.017*****	-0.006
	Mgmt of Human Resources	4	**0.030*****	**0.008*****	**0.006***	**0.033*****
	Mgmt in Office setting	13	**-0.006*****	**-0.005*****	**-0.013*****	**-0.004***
	Convseration techniques	26	**0.010*****	**0.006*****	0.002	**0.016*****
	Persuasion	28	-0.006	**-0.010*****	-0.001	-0.006
	Sales	30	**-0.007*****	**-0.004*****	**-0.010*****	**-0.005*****
	Administrative Work	32	-0.002	-0.007	-0.001	**0.014***
	Discussing Calculations	33	0.002	0.002	0.003	**0.007*****
Basic	Reading/Following Instructions	8	0.003	**0.005***	-0.003	0.007
	Discussing Size Calculations	9	**-0.007*****	-0.002	**-0.011*****	**-0.010*****
	Reading in ICT setting	12	**-0.021****	**-0.011***	0.013	**-0.038*****
	Apply skill	14	**-0.016*****	0.004	**-0.022*****	**-0.039*****
	Field-specific Reading	16	0.001	**0.006*****	0.001	-0.002
	Complex problem solving	18	**-0.013*****	**-0.010*****	**-0.026*****	**-0.021*****
	Foreign language	19	**-0.005*****	**-0.005****	**-0.015*****	**-0.005****
	Reading Specifications	20	-0.002	**-0.010*****	-0.000	0.010
	Interpreting Manuals	21	**0.010*****	**0.009*****	**-0.005***	**0.017*****
	Reading Comprehension	24	**-0.009*****	**-0.011*****	-0.005	-0.008
	Following regulation	25	-0.001	**0.004****	0.008	-0.005
	Following guidelines	36	**-0.003*****	**0.004*****	0.003	**-0.009*****
Controls	Level 3		0.085***	0.087***	0.115***	0.063***
	Level 4		0.240***	0.221***	0.282***	0.243***
	Economics		-0.064***	-0.043***	-0.114***	-0.057***
	Health		-0.023**	0.000	-0.015	-0.022
	Female		-0.042***	-0.020***	-0.076***	-0.057***
	Constant		0.846***	0.484***	0.727***	1.139***
	N		322,205	143,404	28,430	150,371
	R^2		0.254	0.267	0.359	0.233

Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level (not reported here). Each regression includes demographic, degree (track, level and field), year, and school controls, and a constant. LSS/HSS stands for low/high skill service sector.

Furthermore, we would like to stress that there is no 'one-skill-fits-all' for Dutch middle educated students. The returns to social, technical and basic skills strongly differ across fields of education, sector of employment, level of education, and class-based or apprenticeship based tracks. We also showed that learning social skills in school may not be beneficial for middle-educated students upon entering the labour market. Nevertheless, it might be the case that social skills will pay off later in their careers - if they are learned on the job. The results from the factor analysis and the regression on factors also highlight that social, technical and basic skills are multifaceted terms and different aspects of each of those skills are needed in different sectors.

Importantly, our findings imply that those who enrol in a degree with a relatively stronger focus on obtaining social skills have lower wages than degrees focusing on other types of skills, e.g., technical skills. The methodology employed and the results provided could be used for further research on the topic. For instance, it could be further explored that rather than encouraging students to enter a STEM degree, it might be a better policy to increase the relative importance of technical skills in all degrees - and especially for those students who are educated in the field of healthcare, or working in the high skill services sector. Overall, our results address some of the key aspects of middle education in the Netherlands, with implications for a number of other countries using similar education systems, and its labour market implications.

References

ALTONJI, J. G., E. BLOM, AND C. MEGHIR (2011): "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 4, 185–223, publisher: Annual Reviews.

ARCIDIACONO, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

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

AUTOR, D. H. AND D. DORN (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–1597.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2006): "The Polarization of the U.S. Labor Market," *American Economic Review*, 96, 189–194.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118, 1279–1333.

BEAUDRY, P. AND D. A. GREEN (2003): "Wages and Employment in the United States and Germany: What Explains the Differences?" *American Economic Review*, 93, 573–602.

BISSCHOP, P., J. ZWETSLOOT, B. TER WEEL, AND J. VAN KESTEREN (2020): "De overgang van het mbo naar de arbeidsmarkt de positie van jongeren met een migratieachtergrond belicht," SEO-rapport 2020-05.

BORGHANS, L., B. TER WEEL, AND B. A. WEINBERG (2014): "People Skills and the Labor-Market Outcomes of Underrepresented Groups," *ILR Review*, 67, 287–334.

VAN DEN BOSCH, A., B. BUSSER, S. CANISIUS, AND W. DAELEMANS (2007): "An efficient memory-based morphosyntactic tagger and parser for Dutch," *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*, 99–114.

BÁRÁNY, Z. L. AND C. SIEGEL (2020): "Biased technological change and employment reallocation," *Labour Economics*, 67, 101930.

CENTRAAL BUREAU VOOR DE STATISTIEK (2016): "Helft leert door na beroepsopleidende leerweg mbo," .

——— (2018): "Minder mbo'ers naar hbo in afgelopen decennium," .

CNOSSEN, F. (2021): "Technology, Tasks and Careers," (mimeo).

DAUTH, W., S. FINDEISEN, J. SUEDEKUM, AND N. WOESSNER (2021): "The Adjustment of Labor Markets to Robots," *Journal of the European Economic Association*.

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

DEMING, D. AND L. B. KAHN (2017): "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, 36, S337–S369.

DEMING, D. AND K. L. NORAY (2018): "STEM Careers and the Changing Skill Requirements of Work," Working Paper 25065, National Bureau of Economic Research.

DEMING, D. J. (2021): "The Growing Importance of Decision-Making on the Job," Working Paper 28733, National Bureau of Economic Research.

EGGENBERGER, C., M. RINAWI, AND U. 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, A., D. GALLIE, F. GREEN, AND Y. ZHOU (2007): "Skills At Work, 1986 to 2006," Tech. rep.

FOUARGE, D., A. KÜNN, AND D. PUNT (2017): "De rol van arbeidsmarktinformatie in de opleidingskeuze van mbo'ers," ROA Reports No. 009.

GENTZKOW, M., B. KELLY, AND M. TADDY (2019): "Text as Data," *Journal of Economic Literature*, 57, 535–574.

GOLDIN, C. AND L. F. KATZ (2010): *The Race between Education and Technology*, Cambridge, Massachusetts; London, England: Belknap Press.

GOLSTEYN, B. H. H. AND A. STENBERG (2017): "Earnings over the Life Course: General versus Vocational Education," *Journal of Human Capital*, 11, 167–212.

GOOS, M. AND A. MANNING (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 89, 118–133.

GOOS, M., A. MANNING, AND A. SALOMONS (2014): "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, 104, 2509–2526.

HANSEN, S., T. RAMDAS, R. SADUN, AND J. FULLER (2021): "The Demand for Executive Skills," Working Paper 28959, National Bureau of Economic Research.

HANUSHEK, E. A., G. SCHWERDT, L. WOESSMANN, AND L. ZHANG (2017): "General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle," *Journal of Human Resources*, 52, 48–87.

DE KONING, J., A. GELDERBLOM, L. DEN HARTOG, AND T. BERRETTY (2010): "Techniek: Exact goed? Het keuzeproces van allochtone en autochtone leerlingen in het (v)mbo verklaard." Tech. rep., SEOR, Erasmus University Rotterdam, Rotterdam.

MACPHEE, D., S. FARRO, AND S. S. CANETTO (2013): "Academic Self-Efficacy and Performance of Underrepresented STEM Majors: Gender, Ethnic, and Social Class Patterns," *Analyses of Social Issues and Public Policy*, 13, 347–369.

MALAMUD, O. AND C. POP-ELECHES (2010): "General Education versus Vocational Training: Evidence from an Economy in Transition," *The Review of Economics and Statistics*, 92, 43–60.

MICHAELS, G., F. RAUCH, AND S. J. REDDING (2019): "Task Specialization in U.S. Cities from 1880 to 2000," *Journal of the European Economic Association*, 17, 754–798.

MINCER, J. (1974): "Schooling, Experience, and Earnings," NBER Books, National Bureau of Economic Research, Inc.

NATIONAL CENTER FOR O*NET DEVELOPMENT (2021): "Skills," .

MINISTERIE VAN ONDERWIJS, C. E. W. (2013): "Home - Inspectorate of Education," .

REINHOLD, M. AND S. THOMSEN (2017): "The changing situation of labor market entrants in Germany," *Journal for Labour Market Research*, 50, 161–174.

SBB (2021): "About SBB," .

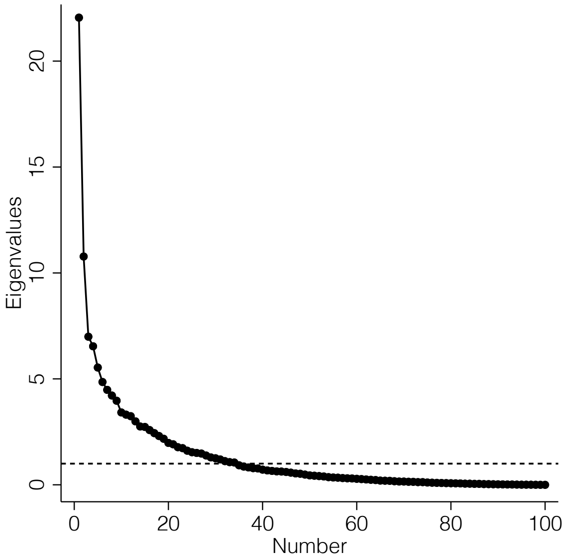
SPITZ-OENER, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 24, 235–270.

TER WEEL, B., J. ZWETSLOOT, AND P. BISSCHOP (2021): "Technologie verslechtert arbeidsmarktkansen van mbo'ers," *Economisch Statistische Berichten*.

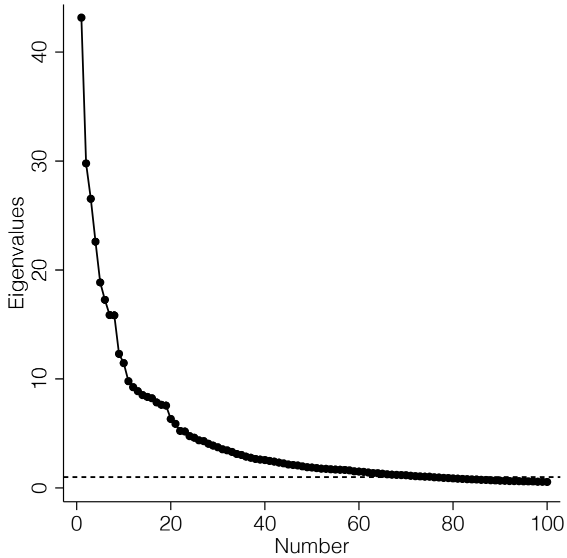
WEBB, M. (2019): "The Impact of Artificial Intelligence on the Labor Market," *SSRN Electronic Journal*.

Appendix

Figure A1: **Screeplots of Eigenvalues for factors, on more and less restrictive sample**



(a) Sample: 152 verbs, each verb in at least 10 programs. 36 factors with Eigenvalue>1.



(b) Sample: 482 verbs, each verb in at least 5 programs. 80 factors with Eigenvalue>1.

Table A1: **Factor Descriptions, sorted by type of Skill following the O*NET classification of Job Skills**

f	Skill type	Factor description	Factor elements
2	Technical skills	Management of Material Resources	Meet, save and maintain materials, products, provisions, tools
3		Quality Control Analysis	Work quality demands, using testing equipment/tools, looking up instructions/systems
5		Equipment Maintenance	Maintain machines/tools,, recognize/process materials, apply quality norm
6		Management of Maintenance Equipment	Maintain/use protective equipment, maintain equipment
7		Operation and Control in Manufacturing	Operate means of transport, shield waste container
10		Installation	Edit component/part/material
11		System Quality Analysis	Apply systems, check quality, apply machines
15		Installation in Construction	Working with blueprint, slinging loads
17		Following technical regulation	Apply environmental rule/company regulation
22		Apply Quality Procedures	Apply quality demand, apply procedure
23		Coordination in Operations	Operate machine, use protective equipment, cooperate with colleague
27		Using Tools	Apply tool/protective equipment/materials
29		Tool Selection	Selecting materia/tooll, work with equipment
31		Reading Equipment Instructions	Interpret scheme, execute task, use equipment
34		Using Applications	Apply part, use application
35		Using Systems	Work system
1	Social skills	Coordination	Distinguish task/function/responsibility, report in and contribute to meetings
4		Management of Personnel Resources	Recognize problem, act as contact, feedback/reflection/observation skills, instructing others
13		Management in Office setting	Apply ICT skills/meeting skils/feedback skills/presenting skills
26		Conversation techniques	Apply knowledge/tools/conversation techniques
28		Persuasion	Convey information, apply negotiation technique, make decision
30		Sales	Apply presenting skills/sales techniques/techniques
32		Administrative Work	Keep up administration, apply observation techniques, take action, use registration system
33		Discussing Calculations	Perform calculation, give feedback, cooperate colleague
8	Basic skills	Reading and Following Instructions	Read/use checklist/assignment, consult source of information
9		Discussing Size Calculations	Calculate dimensions, discuss with colleague
12		Reading in ICT setting	Understand/read assignment, use ICT applications/systems
14		Apply skill	Apply skill
16		Field-specific Reading	Reading field, reading text
18		Complex problem solving	Prepare SWOT analysis, make analysis, interpret data, make calculation
19		Foreign language	Converse/write in English, converse in language
20		Reading Specifications	Reading/interpreting specifications and blueprints
21		Interpreting Manuals	Reading manuals, interpret schemes, have insight
24		Reading Comprehension	Read documentation/information/instruction/English, make drawing
25		Following regulation	Apply regulation
36		Following guidelines	Work guideline

Table A2: **Classification of industries into sectors, employment shares and skill frequencies**

			Employment				Shares (in %)			Skill freq (std)		
Sector	SBI	Industry	Total	STEM	Econ	Health	STEM	Econ	Health	Social	Tech	Basic
Low-skilled services	G	Wholesale and retail trade	83,745	18,541	42,546	22,658	22	51	27	0.65	-0.23	-0.30
	H	Transportation and storage	8,139	3,875	3,183	1,081	48	39	13	0.41	-0.19	-0.31
	I	Accommodation and food services	37,369	5,277	22,194	9,898	14	59	26	0.76	-0.31	-0.51
	R	Culture, sports and recreation	7,517	1,150	2,436	3,931	15	32	52	1.14	-0.40	-0.58
	S	Other service activities	6,634	1,477	1,125	4,032	22	17	61	-0.15	-0.56	-0.62
		Subtotal	**143,404**	**30,320**	**71,484**	**41,600**	**21**	**50**	**29**			
Goods	A	Agriculture, forestry and fishing	3,330	1,309	1,170	851	39	35	26	0.54	-0.20	-0.24
	B	Mining and quarrying	55	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	-0.40	-0.19	0.55
	C	Manufacturing	14,053	8,545	3,936	1,572	61	28	11	-0.07	0.18	-0.01
	F	Construction	10,992	9,377	1,284	331	85	12	3	-0.42	0.09	0.18
		Subtotal	**28,430**	**19,231**	**6,390**	**2,754**	**68**	**22**	**10**			
High-skilled services	D	Electricity, gas, steam and air cond supply	287	212	60	15	74	21	5	-0.24	0.25	-0.33
	E	Water supply	314	198	86	30	63	27	10	0.14	0.06	-0.33
	J	Information and communication	5,559	1,182	3,748	629	21	67	11	0.76	-0.40	-0.51
	K	Financial institutions	1,981	344	1,158	479	17	58	24	0.63	-0.35	-0.24
	L	Renting, buying and selling of real estate	1,084	211	594	279	19	55	26	0.63	-0.34	-0.20
	M	Professional, scientific and technical services	11,267	3,480	5,973	1,814	31	53	16	0.60	-0.18	-0.18
	N	Renting and leasing, and support services	62653	15975	31031	15,647	25	50	25	0.59	-0.28	-0.33
	O	Public administration	4,767	1,140	2,827	800	24	59	17	0.13	-0.40	-0.28
	P	Education	6,822	2,631	1,441	2,750	39	21	40	0.39	-0.18	-0.41
	Q	Human health and social work	55,637	1,655	5,166	48,816	3	9	88	1.23	-0.36	-0.85
		Subtotal	**150,371**	**27,028**	**52,084**	**71,259**	**18**	**35**	**47**			
		Total	322,205	76,579	129,958	115,613	24	40	36			

Note: Classification of industries into sectors based on Bárány and Siegel (2020). Last three columns represent standardized values of skill frequencies.

Figure A2: **Standardized factor scores by major sector**

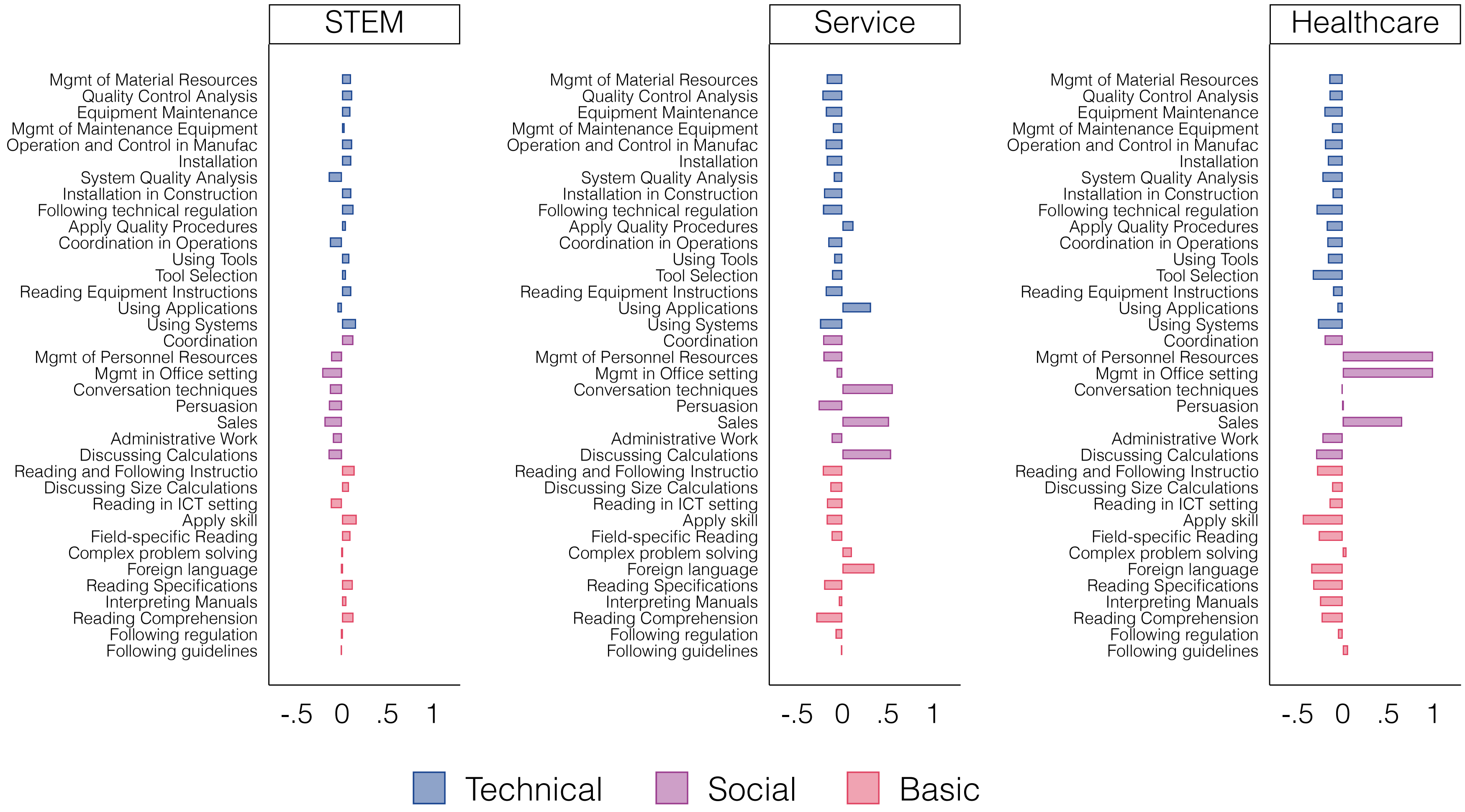


Figure A3: **Standardized factor scores of selected factors, by domain**

Source: Authors' calculations using non-public microdata from Netherlands statistics. *Note:* The value of factor 26 for 'Catering and Bakery' is capped at from 2.8 to 1.5 to accomodate common scales. The domain 'Safety and Sports' contains some economics programs as well as health programs, which is why it has two bars.

Table A3: **OLS Regressions of Log Hourly Wage in First Year after Graduation on Skill Frequencies in Curriculum and Interactions with other Skills, by sector and track**

	(1) Full	(2) LSS	(3) Goods	(4) HSS
Social	-0.031***	-0.014***	-0.030***	-0.039***
	(0.003)	(0.003)	(0.009)	(0.004)
Tech	0.007**	0.005	-0.043***	0.022***
	(0.004)	(0.003)	(0.006)	(0.006)
Basic	-0.018***	-0.003	-0.046***	-0.012*
	(0.005)	(0.003)	(0.011)	(0.007)
Social × Tech	-0.007**	-0.007***	-0.031***	-0.005
	(0.003)	(0.002)	(0.006)	(0.004)
Social × Basic	-0.009***	-0.012***	-0.014*	-0.001
	(0.002)	(0.002)	(0.008)	(0.003)
Tech × Basic	-0.016**	0.005	-0.034***	-0.017*
	(0.006)	(0.004)	(0.009)	(0.009)
Ref. cat.: Level 2				
Level 3	0.113***	0.083***	0.131***	0.107***
	(0.006)	(0.005)	(0.007)	(0.011)
Level 4	0.241***	0.206***	0.255***	0.245***
	(0.006)	(0.005)	(0.012)	(0.007)
Ref. cat.: STEM				
Service	-0.043***	-0.038***	-0.122***	-0.002
	(0.004)	(0.004)	(0.014)	(0.007)
Health	0.073***	-0.009	-0.077***	0.161***
	(0.008)	(0.005)	(0.022)	(0.012)
Ref. cat.: BBL				
BBL	0.184***	0.116***	0.275***	0.227***
	(0.008)	(0.005)	(0.009)	(0.012)
Female	-0.035***	-0.024***	-0.067***	-0.049***
	(0.003)	(0.002)	(0.008)	(0.004)
Constant	0.761***	0.493***	0.734***	1.056***
	(0.027)	(0.033)	(0.024)	(0.032)
Observations	322,205	143,404	28,430	150,371
R-squared	0.234	0.261	0.343	0.213

Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level. Each regression includes demographic, degree (track, level and field), year, and school controls, and a constant. LSS/HSS stands for low/high skill service sector. BOL stands for class-based training, BBL for apprenticeship-based training.

Table A4: **OLS Regressions of Log Hourly Wage in 2 to 10 Years after Graduation on Skills**

A. Full Sample	(1) t+2	(2) t+4	(3) t+6	(4) t+8	(5) t+10
Social	-0.020***	-0.025***	-0.020***	-0.018***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Tech	0.021***	0.016***	0.011***	0.011***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Basic	-0.006***	-0.010***	-0.010***	-0.011***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Observations	320,912	312,708	270,711	199,730	137,528
R-squared	0.142	0.107	0.122	0.150	0.160

B. Low-skilled services	(1) t+2	(2) t+4	(3) t+6	(4) t+8	(5) t+10
Social	0.000	-0.003**	-0.005***	-0.006***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Tech	0.008***	0.005*	0.003	0.003	0.005
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Basic	-0.000	-0.001	-0.006**	-0.013***	-0.017***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
Observations	128,317	104,128	79,914	55,521	36,639
R-squared	0.157	0.076	0.093	0.141	0.170

C. Goods	(1) t+2	(2) t+4	(3) t+6	(4) t+8	(5) t+10
Social	-0.017***	-0.028***	-0.011***	-0.014***	-0.013***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Tech	-0.018***	-0.017***	-0.006	0.005	0.010***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)
Basic	-0.028***	-0.021***	-0.012**	-0.004	-0.002
	(0.008)	(0.006)	(0.005)	(0.003)	(0.004)
Observations	31,109	34,819	33,380	26,569	19,576
R-squared	0.223	0.155	0.142	0.172	0.203

D. High-skilled services	(1) t+2	(2) t+4	(3) t+6	(4) t+8	(5) t+10
Social	-0.038***	-0.040***	-0.026***	-0.021***	-0.019***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Tech	0.038***	0.029***	0.015***	0.011***	0.007**
	(0.005)	(0.005)	(0.004)	(0.003)	(0.004)
Basic	-0.006**	-0.015***	-0.011***	-0.011***	-0.009***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Observations	161,486	173,761	157,417	117,640	81,313
R-squared	0.138	0.124	0.133	0.152	0.153

Note: *** p<.01, ** p<.05, * p<.1. All models are weighted with robust standard errors, clustered at the school level. Each regression includes the full set of demographic, degree (track, level and field), year, and school controls, and a constant.