Teach the Nerds to Make a Pitch: Multidimensional Skills and Selection into Entrepreneurship *

- Preliminary Version-

Damiano Argan † Leonardo Indraccolo ‡ Jacek Piosik §

October 17, 2023

Abstract

Using danish administrative data, in this paper we study how individuals' skill set composition affects self-selection into entrepreneurship. We use detailed education registry data on high school grades in math and danish language to measure analytical and communication skills. We show that highly specialized skill sets are rewarded on the labor market in terms of higher wages, but are negatively related to the probability of becoming an entrepreneur. We also find that for students with high math grades in high school the probability of starting a business is monotonically increasing in their oral grade in Danish, while it is not so for the rest of the population. Motivated by the evidence that students performing well in math run on average more profitable and bigger businesses, we propose an identification strategy to casually estimate the effect of skill multidimensionality on the probability of becoming an entrepreneur. For the population of high performing math students we use information on parents' human capital and exploit within-school, across-cohort variation in students' exposure to peers whose father has a university degree in humanities. We find that the most treated individual (90th percentile) in our sample has 1.1 percentage points higher probability of becoming an entrepreneur compared to the least treated one (10th percentile). The effect is economically significant, being equal to 20% of the overall share of entrepreneurs in the economy. We conclude by highlighting the importance of improving communication skills of individuals with high analytical abilities to incentivize the creation of high performing firms.

JEL codes: E21, L26, J24

Keywords: Entrepreneurship, Skills, Education, Human Capital

*Email: damiano.argan@eui.eu, leonardo.indraccolo@eui.eu and japr.fi@cbs.dk. Damiano and Leonardo would like to thank their advisors Andrea Ichino, Andrea Mattozzi, Árpád Ábrahám and Russell Cooper for their continuous guidance and support throughout the project. This work greatly benefitted from comments by Viola Salvestrini and Fabiano Schivardi.

†Department of Economics, European University Institute (EUI), Florence.

‡Department of Economics, European University Institute (EUI), Florence.

§Department of Finance, Copenhagen Business School.

1 Introduction

Entrepreneurial firms are responsible for a big share of economic growth and job creation in modern western economies.[1] While the contribution of entrepreneurs to the aggregate economy is well documented, less so are the determinants of self-selection into entrepreneurship. Which skills do individuals selecting into entrepreneurship have and how does the skill set composition of entrepreneurs differ from the one of workers? Other studies, starting from the seminal contribution by Lazear (2004), have investigated the role of human capital and skill set compositions as determinants of successful entrepreneurial activity. However, the question remains largely unanswered because of the unavailability of detailed datasets on entrepreneurs and their related characteristics [2].

Our contribution is to use danish education registry data, combined with other administrative data sources, to quantify and casually estimate the effect of the complementarity between analytical and communication skills on the probability of self-selecting into entrepreneurship.

We have access to a rich and detailed dataset that combines multiple administrative data sources to generate a unique dataset that contains all firm ownership in Denmark between 1996 and 2019. Our dataset allows us to follow individuals over their life-cycle and observe their characteristics before they start a business, during their entrepreneurial spell and after. We also observe the same information for paid employed workers, which enables us to study differences in skill set compositions between workers and future entrepreneurs. Motivated by the theory of Lazear (2004), we use high school grades in math and danish language in the last year of high school to measure communication and analytical skills. We start by providing observational evidence on the complementarity vs substitutability of these skills on the labor market and for selection into entrepreneurship. We find that individuals with more specialized skill sets earn higher wages, but are less likely to start a business compared to individuals with more balanced skill sets. We also show that for the group of individuals who were very good in math during high school the probability of becoming entrepreneurs is monotonically increasing in their oral grade in danish, while the same does not apply for the rest of the population of students. Additionally, we find that entrepreneurs who performed well in math during high school run on average more profitable and successful firms.

Understanding how policymakers can incentivize the creation of new successful businesses motivates us to casually estimate the effect of improving communication skills of students with

[1] For example, Decker et al. (2014) show that in the US start-ups are responsible for around 20% of gross job creation.

[2] Queiró (2022) uses Portuguese administrative data to show that more educated entrepreneurs run bigger businesses which display higher growth rates at the beginning of their life-cycle. Michelacci and Schivardi (2020) use data from the SCF to calculate the returns to education for entrepreneurs.

high mathematical grades in high school. Being able to precisely quantify the increase in the number of new businesses created by individuals with high analytical abilities when these are taught better communication skills, is crucial when designing effective training and education programs aimed at spurring the creation of high performing firms.

Our identification strategy draws from the literature on peer effects and uses information on parents' human capital. We exploit within school, across cohort variation in the share of schoolmates parents' with an academic background in humanities. Motivated by the fact that human capital and skills get transmitted across generations, we instrument communication skills of high performing math students with the human capital composition of parents' peers students. For the cohorts of students graduated between 1997 and 2004, the estimated effect of increasing the share of fathers' peers with a university diploma in humanities by 3.5% - corresponding to the difference between the most and the least treated individual in our sample- increases the probability of selecting into entrepreneurship by 1.1 percentage points. The effect is economically significant if one considers that the share of individuals who ever become entrepreneurs in the economy for the 1997-2004 cohort is 4.8%. The effect corresponds to an increase of 20% in the overall share of entrepreneurs.

Other findings by Guiso et al. (2021) have highlighted the importance of exposing aspiring entrepreneurs to entrepreneurial environments to stimulate the birth of new successful ventures. To put our results into perspective, we show that the estimated effect of improving communication skills of highly talented math students on the creation of new businesses is comparable in magnitude to increasing students exposure to a higher share of peer students whose fathers are graduated in business. The rest of the paper is organized as follows. The next section discusses our contribution in relation to other work on entrepreneurship. The third section describes our dataset in detail. The fourth and fifth section provide observational evidence on the role of gpa and high school grades on labor market outcomes and selection into entrepreneurship. The fifth section provides evidence on the complementarity between communication and analytical skills in self-selection into entrepreneurship. The sixth section discusses how high ability math students run on average more profitable businesses. Section seven introduces the identification strategy, section eigth discusses the results and the final section concludes.

2 Related Literature

This paper contributes to different strands of the literature. We contribute to the empirical literature on the role of human capital and skills for the understanding of selection into entrepreneurship (Lazear (2004), Lazear (2005), Wagner (2003), Wagner (2006), Silva (2007),Levine and Rubinstein (2017), Michelacci and Schivardi (2020)) and to empirical studies investigating whether skills relevant to entrepreneurship are learnable (Guiso et al. (2021), Liang et al. (2018)). Prior work has been concerned with bringing empirical evidence to the model of entrepreneurship developed by Lazear (2004). The model predicts that individuals selecting into entrepreneurship must have a more balanced skill set compared to workers (entrepreneurs are jack-of-all traits individuals). The result stems from the assumption that entrepreneurs are required to perform very different task using a variety of skills, while workers only need to perform very specialized tasks. Lazear (2004) tests the theory empirically exploiting the Stanford MBA alumni register. He shows that students with less specialized course tracks, namely individuals who took more courses outside of their track of specialization, are more likely to become entrepreneurs in the future. Additionally, Lazear (2005) uses the same dataset to show how MBA alumni who had a higher number of different occupations as workers are more likely to become entrepreneurs later in life. Along similar lines, Wagner (2003) uses a representative sample of the German population and shows that individuals with more professional training in life, or who changed their profession more often, are more likely to become self-employed later in life. Wagner (2006) uses German data on nascent entrepreneurs (*Regional Enrepreneurship Monitor REM Germany*) to show that individuals who reported to have been active in many different professional fields are more likely to be observed as nascent entrepreneurs. Finally Silva (2007), using the Longitudinal Survey of Italian Families (ILFI, 1997) and exploiting the panel dimension of the data, shows that while the total number of occupations an individual had in his career is positively associated with the probability of being self-employment later in life, the relationship disappears once accounting for the endogeneity by means of fixed effects .

Our research contributes to this literature along several dimensions. First, by using Danish administrative data we define an entrepreneur as the owner of an incorporated business that is economically active. Moreover, we work with the universe of Danish entrepreneurs. This represents an improvement in the quality of the data used to study the determinants of entrepreneurship. Prior work, as Lazear (2004) and Lazear (2005), used survey data not representative of the general population. On the other side, Wagner (2003) does not distinguish between entrepreneurship and self-employment, while later work by Levine and Rubinstein (2017)) has shown how this is critical, as the two groups have very different characteristics. Second, using education registries we have a clear and interpretable measure of skills, namely grades in math for analytical skills and grade in oral Danish examinations for communication skills. Previous literature used very imprecise proxies for the measurement of skills, such as the number of courses taken outside the field of specialization (Lazear (2004)) or the number of different prior occupations (Lazear (2005), or the number of professional degrees (Wagner (2003)). Finally, we address the problem of endogeneity in the relationship between multidimensional skill sets and the decision to become an entrepreneur, which so far has only been addressed by Silva (2007) using panel fixed effects.

The literature studying the human capital determinants of entrepreneurship and firm performance has found patterns of complementarity between skills affecting both self selection into entrepreneurship and firms outcomes. In particular, Levine and Rubinstein (2017) show how having high levels of cognitive ability coupled with the tendency to perform illicit activites when young is a very strong predictor for selection into entrepreneurship. The authors interpret this finding as the existence of complementarity between cognitive skills and the tendency to break established rules. Michelacci and Schivardi (2020) show that the education-labor experience complementarity is associated with higher returns from entrepreneurship and interpret this finding as the existence of complementarity between theoretical knowledge acquired during formal studies and practical skills acquired on the job. We add to these findings by showing evidence of complementarity between analytical and communication skills.

Finally, we add to the literature on how learnable entrepreneurial skills are. Liang et al. (2018) build a model where individuals learn business skills while working in high paid occupations as workers and show evidence that indirectly confirms the model predictions. Guiso et al. (2021) show how individuals that grew up in areas with a high density of firms acquire skills useful to run a business. Similarly, we show that individuals who in their last high school year, while being good in math, were exposed to an environment where they could better learn communication skills acquired abilities useful to entrepreneurship.

The paper most close in spirit to ours is Mertz et al. (2023). Mertz et al. (2023) study the effect of early exposure to entrepreneurship on reducing the gap in self-selection into entrepreneurship between men and women. While asking a different research question than the one we address here, their identification strategy and data are similar. Mertz et al. (2023) use within school across cohort variations in the share of female peers' parents that are entrepreneurs or C level managers to study the impact of early exposure to entrepreneurial environments for incentivizing entrepreneurship among women. Our identification strategy also exploits within school across cohort variation, but we use the share of peers' parents graduated in humanities as treatment. Mertz et al. (2023) also have access to danish administrative data, but while they define entrepreneurs as owners of unnicorporated businesses, we define an entrepreneur as the owner of an incorporated business which is active.

3 The Data

Our analysis is based on the full population administrative data from Denmark, covering the years from 1996 to 2019. The final dataset is composed of two underlying blocks: the entrepreneurial and the education data set. In the next two sections we describe how we construct these two datasets in detail.

3.1 The Entrepreneurial Dataset

The entrepreneurial dataset combines multiple administrative data sources to generate a unique dataset that contains all firm ownership in Denmark between 1996 and 2019. Specifically, by combining individual level characteristics from Statistic Denmark Research Database (DST) with firm level data from the Danish Central Business Register (CVR) and the commercially available KOB database (KOB) from Experian Denmark, we are able to link individual level information to entrepreneurial spells and subsequent business outcomes.

The data contained in Statistics Denmark is provided and updated regularly by relevant Danish authorities, including the Ministry of Taxation, the Ministry of Education and the Ministry of Employment. The database contains general information on individuals such as gender, age, education, wealth and income composition. In addition, detailed employment registers provide all current and previous employment relationships (employer-employee), with corresponding salaries, hours worked, and occupational codes (isco 08) that are used to characterize individual labor market histories, as well as firm-level employment. However, the DST does not contain data on incorporated firms (limited liability companies), but only data on unincorporated firms (sole proprietorship and partnership). As shown by Levine and Rubinstein (2017), when studying entrepreneurship it is key to separate between owners of sole proprietorships and owners of limited liability companies, as they display very different characteristics. To this end, we add the CVR database to the DST dataset, where the former contains information on all firms registered in Denmark since 1980. The CVR also contains detailed ownership records of sole proprietorships, partnerships and corporations and provides the timing, identity and ownership shares of all direct owners. As ownership records referring to incorporated businesses are limited to the period after 2014, we combine the CVR database with data from the commercially available KOB database, published by Experian Denmark, that contains hand-collected ownership information, which completes missing ownership in the early data years of the CVR database. The KOB database also contains detailed accounting records of corporations. All firms in the resulting dataset are identified by unique CVR-numbers, and all individuals are

identified by unique PNR-numbers, which can be matched directly to other data sources.
After combining all these datasets we obtain the entrepreneurial dataset, in which the unit of observation is an individual. For every individual and annually for every year between 1996-2019 the final dataset contains information on individuals' income, net wealth, labor market status, hours worked, occupation, whether an individual owns a sole proprietorship, a partnership or a limited liability company and if so the corresponding business outcomes for each year in which the business exits: revenues, assets, number of employees, turnover, dividends and industry in which the business operates. Concerning years before 1996, the dataset contains, in addition to a variable on accumulated labor market experience, individuals' education level (the highest educational attainment) and her type of education. In addition, for each individual the dataset reports demographic characteristics, the place of birth and the place of living. Finally, we are also able to link individuals with their parents and their siblings through the PNR number, if alive.

3.2 The Education Dataset

The second building block of the final dataset, which we use in our analysis, is the education dataset. Statistics Denmark provides education registries for all cohorts born after 1996. Specifically, for every high school graduate the registries contain information on students' grade point average in the last year of high school. In addition, for all subjects attended by a student, the registries report the grades students have achieved in every examination, as well as in every written and oral assessment and take-home project.
The Danish education system was majorly reformed in 2005. In the interest of working with a homogeneous sample in which grades are comparable, we only keep the cohorts born between 1997 to 2004. Up to 2004, the Danish high school system was characterized by two main tracks students could choose from: a mathematical and a linguistic one. A third track existed, the so called higher preparatory examination (HF), which was designed for young adults who had left the educational system.
In our analysis we focus only on students enrolled in the mathematical high school track. The reason for this is that up to 2004, students in Danish high schools could choose to take subjects at three different levels, corresponding to different difficulties and a different number of hours per subject.[3]. Clearly, grades obtained in the same subject but at different levels are not comparable as the difficulty of the classes is very different. While all students in both the math-

[3]The three different levels were level A (level I), level B (level II) and C (level III). We refer the reader to the appendix for a detailed overview of the Danish high school system.

ematical and linguistic track take Danish classes at the highest level (level A), this is not the case for math classes. Only in math orineted high schools students take math classes at the highest level[4]. In order to have skill measures that are as homogeneous as possible, we select only students enrolled in math-track high schools. In this way for every student we observe the grades he obtained both in danish and math classes, taken at the highest, and thus comparable, level. In the next section we provide an overview of the main descriptive statistics of our sample.

4 Descriptive Statistics

We define an entrepreneur as a business owner of a limited liability company, who over the sample period has hired at least one employee and whose business displays positive revenues and assets. Whenever the individual is the owner of multiple businesses we multiply firm level outcomes (revenues, employment etc.) with the equity shares he holds in the business and apply the same definition. Following the work by Levine and Rubinstein (2017) we define entrepreneurs as owners of incorporated businesses and assure that we do not define as entrepreneur an individual who owns businesses which are empty legal boxes with no economic activity.
Table 1 shows some first descriptive statistics for our sample. In the upper part we report statistics for the general Danish population of men who are older that eighteen in 2019. We see that entrepreneurship is an infrequent career choice. According to our definition of entrepreneurship, slightly less than 4% of the population become entrepreneurs in their life. Those individuals who become entrepreneurs, have on average more years of education (12.8 years against 12.2), and earned higher wages on the labor market, with a difference of around 19 percentage points.

Selection into entrepreneurship does not happen with the same frequency at different stages of the life-cycle. In Figure 1 we plot the probability of becoming an entrepreneur by age for the cohorts born between 1961 to 1975. We see that the probability of becoming an entrepreneur is hump-shaped in age, with a peak around age 38. These life-cycle patterns reveal that we ought to be able to follow individuals for long time periods over their life to study the drivers of selection into entrepreneurship.

[4] In Denmark between 1997-2004, 80% of students enrolled in math high schools took math classes at the highest level.

Table 1: Summary Statistics A

	Danish entrepreneurs	Rest of the population
Men older than 18		
Absolute number	110,356	2,745,731
Share	3.86%	96.14%
Average years of Education	12.78	12.15
	(.007)	(.001)
Log wage	5.50	5.31
	(.0006)	(.0000)
Year of birth 1979-1986		
Absolute number	11162	285267
Share	3.77 %	96.23%
Average years of Education	12.86	13.20
	(.02)	(.005)
Average log wage	5.08	5.10
	(.001)	(.0002)
Mathematical HS Students:Year of Graduation 1997-2004		
Absolute number	1885	37,385
Share	4.8%	95.2%
Average Years of Education	15.21	15.71
	(.05)	(.01)
Average log wages	5.24	5.23
	(.004)	(.000)
Average log GPA	4.18	4.20
	(.009)	(.002)
Average grade in Danish	8.52	8.60
	(.05)	(.01)
Average grade in Written Danish	8.31	8.41
	(.05)	(.01)
Average Grade in Oral Danish	8.77	8.81
	(.06)	(.01)
Average Grade in Math	7.82	8.12
	(.08)	(.019)

Notes — This table reports summary statistics of the data used in the analysis.

The second block of Table 1, with the heading year of birth 1979-1986, shows descriptive statistics for the cohorts for which we observe the high school grades, as described in the previous section. In fact, individuals that graduated from high school between 1997 and 2004 be-

Table 2: Summary Statistics B

	Gymnasiale Graduates	Non Gymnasiale Graduates
Year of birth 1979-1986		
Number	106,960	189,469
Share	36.08%	63.92%
Share of Entrepreneurs	4.15%	3.55%
Average Years of Education	14.7	12.3
	(.007)	(.006)
Average log wage	5.13	5.09
	(.000)	(.000)

Notes — This table reports summary statistics of the data used in the analysis.

Table 3: Summary Statistics C

	Mathematical HS	Linguistic HS	HF
Entire population			
Absolute number	39,270	12,827	8,939
Shares	64%	21%	15%
Gender ratio(male)	0.50	0.22	0.30
Average Years of Education	15.69	15.04	14.08
	(.01)	(.02)	(.02)
Average log wage	5.23	5.15	5.14
	(0.00)	(.001)	(.001)
Average log GPA	4.20	4.14	3.99
	(.002)	(.003)	(.005)
Share Level A Danish	100%	100%	100%
Share Level A Math	80%	0.0%	7%

Notes — This table reports summary statistics of the data used in the analysis.

long to the 1979-1986 cohorts. For this subsample, the share of individuals that become entrepreneurs is the same as for the general population, around 4%. However, we do not find the positive selection in terms of education and wages, which we observed in the general population. Part of the explanation is that entrepreneurship is a career choice undertaken late in life, as shown in Figure 1. This implies that individuals of cohorts 1979-1986 who are pursuing university degrees, have not yet had the time to actually become entrepreneurs. This explains why in the subsample 1979-1986, future entrepreneurs are slightly less educated. Similarly, the

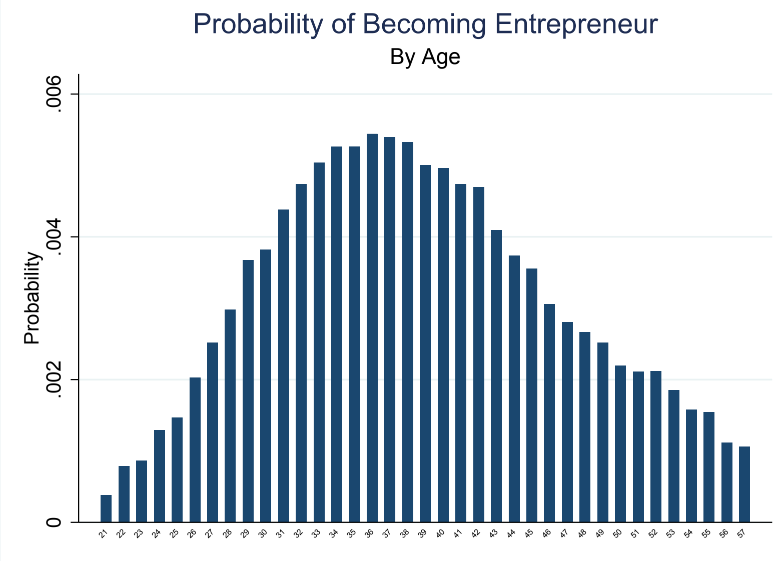

Figure 1: Probability of becoming entrepreneur by age

Notes: Share of individuals that become entrepreneurs at a given age in the Danish population of men for the cohorts 1961-1975. The data starts in 1996 and ends in 2019.

negative selection of future entrepreneurs in terms of wages likely stems from the fact that we are comparing wages of future entrepreneurs when they are relatively young, with wages of individuals who always remain workers and thus are older on average[5]. Given that wages increase on over the life-cycle, this explains why on average future entrepreneurs display lower earnings than always workers.

Of the 1979-1986 cohorts, 37% have a high school degree (gymnasiale uddanelser), while the rest do not. Individuals who have a high school diploma are likelier to become entrepreneurs (4.15% against 3.55%) and unsurprisingly have more years of education and earn more. This can be seen in Table 2.

We further provide statistics for the subsample of individuals who completed their high school degree (gymnasiale uddanelser) between the years 1997 and 2004 in Table 3. Among the 61 036 men completing high school during these years, 64% of them completed a mathematical high-school program (39 270 individuals), 21% of them a linguistic one (12 827 individuals) and15% of them the higher preparatory examination high school (8 939 individuals). The absolute number of individuals enrolled in the different high school tracks, with respect to the overall number of high school graduates, is the result of different gender ratios in the three tracks. The mathematical high school program is gender balanced with a 50% of men and women, the linguistic

[5]By definition wages are only observed for paid employed workers.

program is female dominated with only 22% of men, and similarly for the higher preparatory examination school with a ratio of men to women of 0.3. Mathematical high school students dominate the other students in terms of years of education, gpa and wages when working. A clear hierarchy emerges where mathematical high school students are more skilled than linguistic high school ones, which in turn are more skilled than higher preparatory examination graduates.

The final block of Table 1 displays descriptive statistics for the sample we work with, namely the universe of danish male individuals who graduated from a mathematical high school track between 1997 and 2004. Their average age in 2019 - the last year of the sample - is 38.3 and the median individual is born in 1981. Approximately 4.8% of individuals become entrepreneurs, 1885 out of the 39 270 individuals in total. In this subsample entrepreneurs are slightly less educated in terms of years of education and earned on average 1 percentage point higher wages compared to individuals who never start a business. When compared at the same age, future entrepreneurs earned 10 percentage points higher wages compared to always workers. Regarding educational outcomes we find that entrepreneurs have on average slightly lower gpa (2 p.p less) and also display slightly lower grades in all the different types of danish examinations. In math, future entrepreneurs report an average grade which is about 5% less than the average math grade of individuals who never start a business. In the next section, we examine the relationship between schooling, labor market outcomes and selection into entrepreneurship.

5 Balanced skills and labor market outcomes

Motivated by the theory of Lazear (2004), in this section we study how different compositions of skill sets translate into labor market outcomes for paid employed workers and how they relate to selection into entrepreneurship. We use high school grades individuals obtained in danish to capture communication skills, grades obtained in math classes to measure analytical skills and their interaction to capture skill multidimensionality. In the first subsection we provide simple correlational evidence on the association between high school performance and labor market outcomes and on the relationship between schooling and entrepreneurship. We then move to the role of complementarity in skills.

5.1 Returns to schooling on the labor market

In Table 4 we report the returns on the labor market to the average grade in math and danish for the sample of mathematical high school students graduated between 1997 and 2004. As we

Table 4: High school grades and labor market outcomes

	(2) Log-wage	(3) Log-wage	(4) Log-wage	(5) Log-wage
log_gpa	0.120***			
	(0.00199)			
bStdX	0.044			
Danish		0.00354***		
		(0.000369)		
bStdX		0.008		
Math		0.0113***	0.0117***	0.0131***
		(0.000223)	(0.000216)	(0.000639)
bStdX		0.041	0.042	0.048
Danish Oral			0.00206***	0.00368***
			(0.000287)	(0.000716)
bStdX			0.006	0.010
Danish_Oral#Math				-0.000169*
				(0.0000681)
				-0.008
Constant	4.731***	5.114***	5.123***	5.109***
	(0.00845)	(0.00298)	(0.00261)	(0.00612)
N	561674	563904	563904	563904

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. Sample of Mathematical high school students (HS) that graduate from HS between 1997 to 2004 and attended level A math courses. Log-wage is the logarithm of real-wage. bStDX stands for the variation in the expected value of the outcome variable for a standard deviation variation in the explanatory variable.

can see, a one standard deviation increase in log gpa increases expected real wages on the labor market by 4.4 percentage points, or, otherwise said, a 1% higher gpa predicts a 0.12 % higher real wage. When looking at the grades in math and danish we find that mathematical skills are more rewarded. A one standard deviation higher average grade in math predicts an expected real wage of 4.1 percentage points higher, while the same increases in the average grade in danish is only associated with 0.8 percentage points higher wages. The same patterns holds true if we look at grades in only the oral examinations of danish.

In column 5 of Table 4, we interact math and danish grades to explore whether individuals with more balanced skill sets receive a premium on the labor market as paid employed workers. We find that the coefficient for the interaction term is negative, suggesting that the labor market seems to favor individuals with specialized, rather than balanced skills. Specifically, for given grade in math and danish oral examination, we find that a one standard deviation higher product of the math and danish oral grade, predicts a 0.8 percentage points lower and statistically significant wage on the labor market. This evidence seems to suggest that more multidimensional individuals earn less than unbalanced ones, showing how the labor market rewards skill specialization.

5.2 High school grades and selection into entrepreneurship

So far we have established a positive association between average school performance, as measured by log gpa, and future wages as paid employed workers. Moreover, we have established that the labor market rewards specialized workers more than multidimensional ones.

In this section we ask how school outcomes and the complementarity between different skills relate to the probability of selection into entrepreneurship. We start by asking how average schooling ability, as measure by log gpa, is associated with the decision to start a business. We find that log gpa is negatively correlated with the probability of selecting into entrepreneurship. In the first column of Table 5 we see that a one standard deviation higher log gpa decreases the probability of becoming entrepreneur by 0.4 percentage points, which is 10% of the average share of entrepreneurs in the sample. Next, we study how communication and analytical skills and their complementarity are associated with the probability of becoming an entrepreneur. As before, we use grades obtained in danish and math to measure the two different skills. In Table 5 we report the results of running the following regression:

$$Ent_i = \alpha + \beta_1 \times Math_i + \beta_2 \times Danish_i + \beta_3 \times (Math_i \times Danish_i) + \sum_{y=1997}^{2004} \gamma_y * Year + \epsilon_i \quad (1)$$

where *Ent* is 1 if an individual ever becomes an entrepreneur over the sample period and 0 otherwise. $Math$ and $Danish$ are the average grades of all the math and danish grades received by the individual during the last year of high school and $Year$ are year fixed effects for the year in which the individual received the grades. This specification describes the association between the grades in danish and math received in high school by the student, and their interaction, given the year in which students graduate. The year fixed effects control for the fact that later cohorts have less time to become entrepreneurs in their life as the sample period ends in 2019. They also control for possible "grade inflation" that can affect the grading system across cohorts. We cluster standard errors at the year-school level.

The table shows three different specifications. In the first column of Table 5 $Danish$ is the average of every grade obtained across all types of examinations in danish. In column (2) we only use grades obtained in oral danish exams, while in column (3) we only use grades obtained in written evaluations of danish. The grade in $Math$, instead, is always the mean across all types of examination in math, both written and oral. The first specification shows that: i) there is a negative, but not significant, association between the probability of ever becoming an entrepreneur and the average grade in danish ii) a negative and significant association between ever becoming an entrepreneur and the average grade in math iii) a positive, but slightly insignificant association between the interaction term and the probability of becoming an entrepreneur. Specifications (2) and (3) help us understand whether oral or written danish skills drive the above relationship. We see that further splitting performances in danish into oral and written skills changes our results. Specifically, once we consider only oral danish evaluations the association between the probability of selecting into entrepreneurship and the interaction term becomes more significant. When, instead, we only consider written evaluations in danish, as in specification (3), the interaction term goes down to zero and becomes insignificant. This suggests that oral danish skills, interacted with math skills, are the ones predicting selection into entrepreneurship. The magnitudes of these associations are sizable. From specification (2) we see that a one standard deviation increase in math is associated with a decrease in the probability of becoming an entrepreneur of 1.3 percentage points and a one standard deviation increase in the interaction term between oral and math skills increases the probability of becoming an entrepreneur by 1.3 percentage points. These numbers are economically relevant if we consider that the overall share of entrepreneurs in the sample is 4.9%.

To further gain understanding of the magnitudes of the coefficients let us consider two different individuals who have very different skill set compositions. Let us consider a first individual with a perfectly balanced skill set, who has grade 8 both in danish and math and a second individual who has an extremely specialized set of skills with grade 15 in math and 1 in danish[6]. Using the estimated coefficients from specification (2) this would imply that the individual with a specialized skill set- grade 15 in math and 1 in danish- has a probability of selecting into entrepreneurship of 3.7%, while the individual with a balanced skill set- grade 8 both in math and danish- has almost a double probability of 6.7%.

Together, this provides observational evidence that multidimensional skill sets predict selection into entrepreneurship and that different compositions of skills have meaningful economic significance to understand the decision to start a business. Moreover, our evidence shows that the oral examinations in danish seem to best capture communication skills.

[6]The Danish grading system goes from -3 to 12. We converted it to a 1-16 scale to facilitate the interpretation. A grade of 8 is considered a fair grade, 15 is considered an excellent grade while 1 is unacceptable. The appendix contains detailed information on the danish high school grading system.

Table 5: High school grades and selection into entrepreneurship

		(1)	(2)	(3)
	Entrepreneur	Entrepreneur	Entrepreneur	Entrepreneur
log_gpa	-.0099***			
	(.003)			
bStdX	-0.004			
Danish		-0.00224	-0.00207	-0.000903
		(0.00166)	(0.00138)	(0.00149)
bStdX		-0.005	-0.006	-0.002
Math		-0.00337**	-0.00370***	-0.00202
		(0.00135)	(0.00116)	(0.00123)
bStdX		-0.012	-0.013	-0.007
Danish×Math		0.000237	0.000257**	0.0000863
		(0.000150)	(0.000126)	(0.000138)
bStdX		0.011	0.013	0.004
Graduation year f.e.		✓	✓	✓
N		31293	31293	31292

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. The table reports OLS coefficients of the regression of the probability of being an entrepreneur on log-gpa and HS school grades in math, danish, and their interaction. The sample is the universe of male Danish Mathematical high school students in their last HS school year that attended were enrolled in the last high school year between 1997 to 2004. The outcome variable *Entrepreneur* is 1 if an individual has ever been an entrepreneur in his life. The explanatory variable *Danish* is in specification (1) the average of the grades received in Danish, in (2) the average grades received in solely the oral evaluation of Danish, while in (3) the average grades received solely in the written evaluation of Danish. The explanatory variable *Math* is the average of the grades received in math courses. Grades in Danish and Math are all for level A course. The regression contains additional controls for graduation year f.e.. Standard errors are clustered at school-year-programme in parentheses: # 1113calendar year.

5.3 Teach the nerds to give a pitch: Oral skills and Very Good Math students

In this section we further deepen our understanding of the relationship between skill complementarity and selection into entrepreneurship. We do this by grouping students into three categories based on their performance in danish and math. Specifically, using official definition of grades from the Danish Ministry of Higher Education and Science we define three grade categories and assign individuals based on their mean grade in each subject, computed over all the grades in a given subject obtained in the last year of high school. [7] The categories are the same for danish and math and are: *Bad*, *Average* and *Very Good*. Category *Bad* contains all students with a mean grade that is less than 4 (8 according to our scale). These are students that are in

[7] We refer the reader to the appendix for an extended definition of grades in Denmark. The source for the Danish Grading system can be found at: https://ufm.dk/en/education/the-danish-education-system/grading-system

Table 6: Probability of becoming entrepreneur by skills

	Average Grade in Math		
	Bad	Average	Very Good
	Entrepreneur	Entrepreneur	Entrepreneur
Oral Danish Average	-0.00540	-0.00330	0.0182**
	(0.00475)	(0.00618)	(0.00821)
Oral Danish Very Good	-0.00307	0.000737	0.0259***
	(0.00956)	(0.00733)	(0.00882)
Constant	0.0679***	0.0593***	0.0447***
	(0.00752)	(0.00741)	(0.00972)
N	11790	12345	7158

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. This table reports OLS coefficients of the regression of the probability of being an entrepreneur on HS school grades in Danish Oral examination for different levels of the grades in math. The sample is the universe of male Danish Mathematical high school students in their last HS school year attended between 1997 and 2004 . Specification (1) displays the difference in the share of entrepreneurs by average grade in oral Danish for students that have a bad average grade in Math. Specification (2) displays the difference in the share of entrepreneurs by average grade in oral Danish for students that have an average average in Math. Specification (3) displays the difference in the share of entrepreneurs by average grade in oral Danish for students that have a very good average grade in Math. The baseline is the average share of entrepreneurs for students that have a Bad average grade in Danish Oral. Grades in Danish and Math are all for level A course. The regression contains additional controls for graduation year f.e. Standard errors are clustered at school-year-programme in parentheses: (1) # 1104, (2) #1090, (3) 1070

between meeting only the minimum requirement and a fair performance. The category *Average* contains all individuals with an average grade greater or equal than 4 (8 according to our scale) and less or equal than 7 (11 according to our scale). These are students whose performance is fair, between fair and good, and good. The final category *Very Good* is composed of all students with an average grade higher than 7 (11 according to our scale), who are students with a performance that is more than good, very good or excellent.

Dividing students into these categories helps us gauge more insights on the role of skill complementarity. We do this by running the following regression:

$$Ent_i = \alpha + \beta_1 \times Average_i + \beta_2 \times VeryGood_i + \sum_{y=1997}^{2004} \gamma_y * Year + \epsilon_i \quad (2)$$

Where *Average* and *Very Good* are dummies taking value 1 when an individual has a mean grade falling into one of these categories. The variable $Year$ are year fixed effects for the year in which the individual was graded. We cluster standard errors at the year-school level. With this specification *Average* and *Very Good* read as the difference in the share of entrepreneurs in those categories with respect to the base category *Bad*.

In table 6 we report the regression coefficients of equation 2. In column (1) of Table 6 we subset for the individuals that are *Bad* in math and check how the probability of becoming an entrepreneur changes as we move along the categories of the danish grades, from *Bad* to *Very Good*. We see that there are no significant differences in the share of entrepreneurs as we move from the group of *Bad* (the baseline) students to the group of *Very Good* in the oral danish exams. We observe the same qualitative pattern in column (2) where we subset for the students that belong to the *Average* category in math.

On the contrary, when we look at the effect that higher danish oral skills have on the probability of selection into entrepreneurship for students that are very good in math, we observe an increasing pattern as we move along the different categories of danish oral exams. Very good math students who score *Average* in danish oral have a 1.8 percentage points higher probability of becoming an entrepreneur compared to the baseline of *Bad* students. In turn, very good math students who belong to the *Very Good* group in oral danish exams have 2.6 percentage points higher probability of ever becoming entrepreneurs compared to the baseline category. Considering that for the baseline group of *Bad* students the probability of becoming entrepreneurs is 4.5%, it means that high ability math students scoring very well in danish oral exams have a 60% higher probability of transitioning into entrepreneurship compared to talented math students that score poorly on danish.

In Table 7 below we report the shares of entrepreneurs for the different combinations of groups (9 combinations in total), without controlling for the high school year of graduation (standard error in parenthesis). From the table we observe another couple of facts. As already found with regression 2, for students who are very good in math the share of individuals that become entrepreneurs is strongly increasing in the oral score in danish. This pattern, however, is not present for students who are bad or average in math. Second, the share of entrepreneurs is decreasing with higher grades in math (column I), but not with higher grades in the oral exam of danish. Third, the highest share of entrepreneurs is among the students that are bad both in oral danish exams and math.

This last two findings align with evidence from the previous sections. In particular, the last finding is in accordance with log gpa being negatively associated with selection into entrepreneurship, while the fact that the share of entrepreneurs is decreasing with higher grades in math can be explained considering that individuals with highly specialized mathematical skills are highly rewarded on the labor market as paid employed workers.

The most interesting correlational evidence is that the probability of selecting into entrepreneurship for high skilled math students is strongly increasing in oral danish abilities. This finding can be rationalized through models of entrepreneurship that build on the intuition initially proposed by Lazear (2004), for which individuals that start businesses need to be able to perform a variety of different tasks and must thus be multidimensional in their skill set. More recent research by Choi et al. (2019), who analyze the human capital composition of founding teams in start-ups, seem to give similar importance to the multidimensionality of human capital for the understanding of business outcomes.

Table 7: Share of entrepreneurs by group of grades

	Math grade		
	Bad	Average	Very good
Danish grade			
Bad	0.057	0.052	0.023
	(0.004)	(0.006)	(0.008)
Average	0.052	0.048	0.039
	(0.002)	(0.002)	(0.003)
Very good	0.054	0.051	0.046
	(0.009)	(0.005)	(0.004)

Notes: The table reports the share of entrepreneurs by group of grades without any additional controls.

6 Entrepreneurial outcomes of math skilled students

In this section we study the performance of businesses owned by individuals who attended a mathematical high school and took math classes at the highest level. We are motivated by our previous findings that for the group of talented math students the share of entrepreneurs increases as their communication skills improve. We now want to know whether firms owned by individuals with high mathematical skills on average also generate more revenues, more employment and are more profitable. If this is the case, then asking how we can incentivize entrepreneurship among students with high analytical skills becomes a policy relevant question to investigate.

To simply the analysis of this section, whenever in our data we have individuals who are owners of multiple businesses we only keep the outcomes of the firm in which the individual holds the highest share. This helps us create a more direct link between individuals and firm performance[8]. We start by running a simple regression in which we study the association between attending a mathematical high school - and taking level A math classes- and firm performance. We run the following regression:

$$Out_{it} = \alpha + \beta_1 * MathHS_i + \sum_{y=1997}^{2004} \gamma_y * Yearbirth + \sum_g \theta_g * year + \epsilon_{it} \quad (3)$$

where Out_{it} is firm outcome in year t, for the firm owned by individual *i*, $MathHS_i$ is a dummy taking value one if the entrepreneur attended a mathematical high school and $Yearbirth$ and $year$ are respectively the year of birth of the individual and year fixed effects. The coefficient $MathHS$ displays the average difference in firms outcome with respect to the general population of firms, for firms that are owned by an individual who attended a mathematical high school and took math courses at the highest level (level A). [9] We control for the year of birth of the owner to control for the longer period older individuals have to open and manage a firm, and we control for year fixed effects to account for aggregate economic conditions. The unit of observation is the firm-year outcome.
Table 8 reports the coefficients of regression 3. Starting from the upper left column we see that firms owned by individuals who attended mathematical high schools have: i) a higher number of employees- 1 more employee with respect to the average of the general population which is approximately 8, that is 12,5% higher number of employees compared to the baseline; ii) almost 7 percentage points higher revenue; iii) 17 percentage points higher value added; iv) 26 percentage points higher assets; v) 32 percentage points higher earnings before interests and taxes (Ebit); vi) 35 percentage points higher net income; viii) 34 percentage points higher value of equity. To sum up, these means that an entrepreneur who attended a mathematical high school on average holds bigger firms (in terms of employment, revenues and assets) and more profitable firms (ebit and net income), compared to entrepreneurs who did not attend a mathematical high school.

In Table 9 we report the coefficients of running the same regression as before, but in which we split individuals according to our three categories of *Bad*, *Average* and *Very Good* math skills.[10] Thus the coefficients *Bad Math*, *Average Math*, *Very Good Math* show the difference in average firm outcomes of individuals that attended a math high school with an average math

[8] In our data the median share is around 68%

[9] The set of observations is composed of all firm outcomes of businesses owned by individuals born between 1979 and 1985 in order to be comparable to firm outcomes of firms owned by individuals who attended a mathematical high school, as the education dataset is available for cohorts graduated between1997 and 2004

[10] The three categories are defined as in the previous section.

Table 8: Firm outcomes of mathematical high school students

	Employees	Log revenues	Log value added	Log assets	Log Ebit	Log net income	Log equity
Math HS	1.092***	0.0692***	0.175***	0.269***	0.337***	0.361***	0.351***
	(0.124)	(0.0160)	(0.0154)	(0.0179)	(0.0241)	(0.0254)	(0.0223)
Constant	7.477**	8.335***	7.827***	8.469***	6.286***	5.338***	7.397***
	(2.883)	(0.369)	(0.330)	(0.390)	(0.561)	(0.538)	(0.451)
N	45741	45274	44752	45104	36623	35004	39768

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. This table reports OLS coefficient of the regression of firm outcomes on having been in Mathematical HS. The coefficient Math HS measures the average difference in firms outcomes owned by students who attended a mathematical high school, compared to outcomes of firms owned by students who attended other high schools. An individual is considered attending a Mathematical HS if she attended Mathematical HS and took level A math classes (80% of the sample). The sample is the universe of all limited liability companies having at least one employee owned by an individual born between 1979 and 1986. The regression contains additional controls for year of birth of the owner and year fixed effect. All outcome variables are defined as in the Danish Authority accounting standards, apart from employment they all log, and they are all trimmed at the 1st and 99th percentile. In terms of entrepreneurs the sample is composed by 11162 entrepreneurs, of which 1392 are mathematical HS.

grade being either *Bad*, *Average* or *Very Good*, with respect to the general population of entrepreneurs as defined in Table IV. We find that the better an individual was in math at high school, the bigger and the more profitable his firm is. In terms of profitability, an entrepreneur who belonged to the *Very Good* math group in high school owns firms that have an ebit of 54 percentage points higher than the general population, while entrepreneurs who belong to the *Bad* math category have businesses with an ebit which is only 22 percentage points higher than the baseline. Similarly for net income, where *Very Good* math business owners have firms which display net income that is 55 percentage points higher than the general population, while for the *Bad* group it is only 24 percentage points higher. Also in terms of revenues and employment we see that i) individuals who belong to the *Very Good* category in math skills own firms that have 2 employees more than the rest the population (25% more), while individuals who belong to the *Bad* group own firms that have the same amount of employees as the general population; ii) *Very Good* math entrepreneurs also have firms that on average have 16 percentage points higher revenue, while *Bad* math entrepreneurs only 5 percentage points higher. These findings hold true also for the other firm performance measures as can be seen in Table ...

To sum up, this evidence shows that individuals who attended a mathematical high school program own firms that are bigger and more profitable compared to the rest of the population and that both firm size and profitability is increasing with analytical skills, as measured by math grades in high school. While suggestive, these findings point towards the idea that quantifying the role of complementarity between communication and analytical skills in driving self-selection into entrepreneurship for the group of high skilled math students is a policy relevant

Table 9: High school grades and firm outcomes

	Employees	Log revenues	Log value added	Log assets	Log Ebit	Log net income	Log equity
Bad_Math	0.311	0.0525*	0.0508*	0.180***	0.242***	0.252***	0.215***
	(0.190)	(0.0243)	(0.0233)	(0.0273)	(0.0369)	(0.0390)	(0.0341)
Average_Math	1.442***	0.0428	0.218***	0.255***	0.334***	0.378***	0.337***
	(0.188)	(0.0244)	(0.0234)	(0.0273)	(0.0363)	(0.0382)	(0.0338)
Very_Good_Math	1.957***	0.159***	0.348***	0.482***	0.540***	0.549***	0.649***
	(0.267)	(0.0348)	(0.0334)	(0.0386)	(0.0524)	(0.0550)	(0.0474)
_cons	7.531**	8.341***	7.836***	8.481***	6.288***	5.352***	7.415***
	(2.883)	(0.369)	(0.329)	(0.390)	(0.560)	(0.538)	(0.451)
N	45741	45274	44752	45104	36623	35004	39768

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. This table reports OLS coefficient of the regression of firm outcomes on having been in a mathematical HS and having a Bad, Average, very Good grade average in Math. An individual is considered attending Mathematical HS if she attended Mathematical HS and took level A math (80% of the sample).The sample is the universe of all limited liability companies having at least one employee owned by an individual born between 1979 and 1986. The regression contains additional controls for year of birth of the owner and year fixed effect. All outcome variables are defined as in the Danish Authority accounting standards, apart from employment they all log, and they are all trimmed at the 1st and 99th percentile. In terms of entrepreneurs the sample we have 11162 entrepreneurs, of which 1392 are mathematical HS, 471 are Bad_Math, 446 are Average_Math, and 210 Very_Good_Math.

question to explore in light of the above average quality of firms that math skilled entrepreneurs run. This leads us to the next section in which we propose an identification strategy to casually estimate and quantify the effect of increasing communication skills for mathematical high school students on the probability of starting a business.

7 Identification Strategy

So far we have provided compelling observational evidence on the relationship between individuals' skill set composition and selection into entrepreneurship. On a correlational level we observe that i) individuals who were good at math in high school run more successful and profitable businesses ii) the share of good math high school students who start a business is increasing in their communication skills. In this section we want to do an additional step and casually estimate the effect of improving communication skills of high ability math students on the probability that they will become entrepreneurs. Quantifying the effect of this treatment for the population of high skilled math students is crucial if policymakers want to design training programs or other policy interventions to incentivize the creation of high performing firms.

To causally estimate the effect of communication skills on selection into entrepreneurship for the population of high skilled mathematical high school students, we draw from the literature on peer effects. The general idea of this literature is that individuals learn from each others and that skills get transferred across students in the same school. We use information on parental education and exploit within school, across cohort variation in the share of parents peer students' with an academic background in humanities (Mertz et al. (2023)). Motivated by the fact that human capital and skills get transmitted across generations, the basic idea is to instrument individual communication skills with the human capital composition of parents peers' classmates. In particular, for every individual who attended a math high school and took level A math classes we compute the share of his parents schoolmates' who have a university degree in the field of humanities[11]. We then regress a dummy that takes value one if an individual ever becomes an entrepreneur in our sample on the share of fathers peer students' graduated in humanities, controlling for school fixed effects and school fixed effects interacted with a variables that measures the share of the parents of the peer students across the different possible education levels[12]. In other words, for the seven cohorts (1997-2004) of students graduated in mathematical high schools we compare individuals who in their last year of high school were in the same school, in two different cohorts - but exposed to the same share of fathers peers' at every education level- where for one student the fathers of his peers are more frequently graduated in humanities than other fields. We additionally control for the education level and the area of education of the father of every individual as we adopt a leave one out strategy. [13]. The specification is:

$$Ent_i = \alpha + \beta_3 * Hum_{J-i,i} + \sum_{sc=n,edccs=k}^{sc=N,edcs=K} \Gamma_{n,k} * Sc_n * edcs_k + \theta * C_i + \epsilon_i \quad (4)$$

Where Ent_i is a dummy taking value one if the individual ever becomes an entrepreneur in our sample, $Hum_{J-i,i}$ is the share of fathers peers' graduated in humanities during the last high school year of the individual,$Sc_n * edcs_k$ is the cross product between a school fixed effect and the share of maximal education level of fathers' peers, C_i contains the individual's father education level and field of graduation as well as municipality fixed effects. So for two individuals in the final year of high school, who attended the same mathematical high school in different years, whose parents peers' however have the same educational structure (same share of fathers peers' with 9,12,14,15,17,20 years of education), we estimate the difference in the share of entrepreneurs for the individual that was exposed to a higher share of peer students with parents graduated in humanities. This identification strategy has two key assumptions. First, we

[11] We consider university degrees that take 5 years to be completed (so BSc + Msc).

[12] This means that the variable measures the share of parents peers' that have 9,12,15,17 or 20 years of education, corresponding to an education level of less than high school, high school, BSc, MSc and PhD.

[13] The share is computed for each individual, leaving his father out when computing the share.

assume that individuals did not strategically self-select into school programs with schoolmates that were more frequently sons of fathers with university degrees in humanities. The second identifying assumption is that sons of fathers graduated in humanities are on average better in communication skills and that these skills transfer to peer students. This last assumption is testable. We discuss this assumption in the next section.

7.1 Variability in treatment

In the literature on peer effects, a common concern with our type of identification strategy is whether there is enough variation in peer characteristics. In our setting, this means asking if across cohorts of students that graduated from the same mathematical high school during the years 1997-2004 and who have the same number of fathers with a university degree, we observe enough variation in the share of fathers peers' with a humanistic degree.

We address this concern in Table 10 where we display the mean, standard deviation, 10th and 90th percentile of the share of fathers graduated across different fields and the residuals from the regression of the share of fathers graduated in different fields on school fixed effects and school fixed effects interacted with the shares of maximum education of the fathers of peer students (colum *residualized*). In the column *plain*, we see that on average a mathematical high school student in her last year of school has 2% of fathers peers' graduated in humanities, with a standard deviation of 0.026. To obtain a sense of the variability in the share of fathers peers' graduated in humanities across difference school-years cells, we show the 10th and 90th percentile of the share of fathers peers' humanities graduates[14]. The least *treated* students, meaning individuals exposed to the 10th percentile of the share of peers with fathers graduated in humanities, have a 0% share. That is, in their school-year cell, nobody is the son of a father graduated in humanities. The very *treated* individuals are students exposed to the 90th percentile of the share of fathers peers' graduated in humanities and have a 6.3% share of peer students with fathers graduated in humanities in their school-year cell. We additionally display the same moments for the other relevant fields of study: business, engineering, and natural sciences. In general, the share of fathers peers' graduates across the different disciplines can be thought as the different probability to meet a peer student in the school that has higher abilities in the field of his father's university subject and represents for the other students a source of accumulation of field-specific human capital through peer interactions and learning spillovers. Our identification strategy relies on exploiting different exposures of students to the accumulation

[14] For privacy policies when using the data we cannot show percentiles of distributions, but show instead averages around these percentiles.

of specific skills that arise from the human capital of fathers peers' students. In the main regression 4 we control for the overall level of fathers with a university degree. This makes sure that we do not use variation in the share of fathers peers' with humanities degrees that only come from school-years cells having a higher overall number of fathers with a university degree. For this reason, Table 10 in column *residualized* shows the moments of the share of fathers peers' graduated across different disciplines after controlling for school fixed effects and school fixed effects interacted with the shares of maximum education of the fathers of peer students. The standard deviation is around 0.016 and the difference between the least and the most treated individual is about a 3.5% to 4% difference in the share of peers with fathers graduated in humanities.

Table 10: Variability in treatment

	Plain				Residualized			
	Mean	St. Dev	mean 9-11 pc	mean 89-91 pc	Mean	St. Dev	mean 9-11 pc	mean 89-91 pc
Humanities	0.025	.030	0	0.063	0	.016	-0.017	0.017
Business	0.029	0.033	0	0.072	0	.016	-0.017	0.018
Engineering	.039	0.043	0	0.11	0	.018	-0.021	0.020
Natural sciences	0.025	0.028	0	0.065	0	0.015	-0.016	0.017

Notes: The table reports mean, St. Dev, mean 9-11 percentile, mean 89-91 percentile of the share of peers' fathers graduated in humanities and residuals of a regression of the share of peers' fathers graduated in humanities on school f.e. and school f.e. interacted with the education share of peers parents (Residualized), for Mathetimatical HS students in their last year belonging to the cohort 1997-2004.

7.2 Father field of graduation and their son performance

Given that our empirical strategy relies on using human capitaI of fathers peer students' as an instrument for the communication skills of an individual, in this section we provide evidence on the intergenerational transmission of human capital and skills between fathers and sons. Specifically, in Table 11 we show the average grade in math and danish, high school gpa and the probability of becoming an entrepreneur of high school students, for the different disciplines in which their fathers obtained a university degrees. The sample is made of all high school students graduated in a mathematical high school between 1997 and 2004, who have fathers with a university degree. The averages in the table are computed with respect to a reference made of the averages of sons with fathers graduated in other disciplines: education, social science, information, agriculture, welfare, service and unknown. In column (1) we show that the average grade in danish for students whose father is graduated in humanities is significantly higher than the average of students who have fathers graduated in other disciplines (almost 7% higher than the reference). In particular, the grade is higher than the one obtained by sons of graduates in business, natural sciences and engineering who all have lower than reference average grades. The same pattern applies when we consider oral and written examinations separately, with a greater positive difference between the grades in oral than written exams (column (2) and (3)). When we move to the grade in math (column (4)) we see that sons of humanity graduates have a higher average grade also in math, but the positive difference with respect to the reference category is smaller, while significant (2.8% higher grade). Sons of business graduates do not have a significant different average grade compared to the reference category, while the sons of graduates in natural sciences have a significant higher grade in math (6.7% higher). Somehow surprisingly, sons of engineers have lower than average grade in math, and also lower than the average math grade of sons of humanity graduates.

In Column (5) we display the average gpa in the last year of high school by father field of graduation. Students with fathers graduated in humanities are those with the highest gpa. Finally, in column (6) we show the average share of entrepreneurs by the field of graduation of the father. The son of humanity graduates have on average a 1,2 percentage point lower share of entrepreneurs than the reference category, but the difference is not significant. On the other hand, as expected, sons of business graduates have a 2.5 percentage points higher share of entrepreneurs compared to the baseline (which is more than the half of the average share of entrepreneurs in the sample).

In light of the evidence in Table 11, we showed that indeed the son of a graduate in humanities has better communication skills, as proxied by the grade in the oral danish exams, than sons of graduates in other disciplines. This evidence is suggestive of the fact that human capital is transmitted across generations and supports our empirical approach. Second, sons of humanity graduates do not become more frequently entrepreneurs, as is the case instead for sons of business graduates. This is important as it implies that there are no other factors pushing sons of humanity graduates to select into entrepreneurship. Finally, the literature on intergenerational skill transmission usually thinks of human capital being transmitted between father and sons as a bundle of skills. While it is true that sons of humanity graduates have higher gpa on average, we know from the previous sections that high school gpa is negatively associated with the probability of becoming an entrepreneur. This means that individuals graduated in humanities likely transmit a bundle of skills to their sons, but predominately so they transmit better communication skills.

Table 11: Father field of graduation and son high school grades

	(1) Danish	(2) Danish Oral	(3) Danish Written	(4) Math	(5) log-gpa	(6) Entrepreneur
Humanities	0.643***	0.641***	0.577***	0.267***	0.0437***	-0.0119
	(0.111)	(0.132)	(0.117)	(0.146)	(0.0134)	(0.00728)
Business	-0.169*	-0.116	-0.192**	-0.0971	-0.0141	0.0248**
	(0.0901)	(0.113)	(0.0890)	(0.145)	(0.0126)	(0.00931)
Natural Science	-0.246***	-0.305**	-0.185**	0.635***	0.00958	0.00484
	(0.0902)	(0.114)	(0.0900)	(0.143)	(0.0123)	(0.00851)
Engineering	-0.487***	-0.475***	-0.478***	-0.245**	-0.0717***	0.0172**
	(0.0776)	(0.0976)	(0.0792)	(0.122)	(0.0116)	(0.00741)
_cons	9.307***	9.565***	9.056***	9.450***	4.355***	0.0431***
	(0.0435)	(0.0534)	(0.0453)	(0.0692)	(0.00583)	(0.00376)
N	7050	7050	7050	7050	7050	7050

Notes: $^{*}(p < 0.1),^{**} (p < 0.05),^{***} (p < 0.01)$. This table reports OLS coefficients of the regression of the average grade in the final year of high school (for student of Mathematical HS) in (1) Danish, (2)Oral Danish, (3) Written Danish, (4) Math, (5) probability of becoming an entrepreneur and (6) log-gpa on a dummy for the father of an individual being graduated in humanities, business, Natural Science, Engineering. The coefficients display the difference in the variable of interest of the mean of every field with respect to the rest of the group of fields that are: education, social sciences, information, agriculture, welfare, service, uknown. An individual is considered attending Mathematical HS if she attended Mathematical HS and took level A math (80% of the sample of mathematical HS students). The sample is the universe of all Mathematical HS students having a father that has a university education. Standard errors are clustered at school year level. Cluster #: 1053 all specification

8 Second stage

In this section we analyze the effect of being exposed to a higher share of peers whose father is graduated in humanities on the probability to become an entrepreneur. If the two identifying assumptions hold, then equation 4 captures the casual effect of increasing communication skills of mathematical high school students on the probability to start a business and become entrepreneurs.

In Table 12, we report the estimated regression coefficients of equation 4. We display the coefficients estimated on subsamples of different cohorts: in (1) cohorts from 1997 to 2004, in (2) from 1997 to 2003, in (3) from 1997 to 2002 and in (4) from 1997 to 2001. The reason for this goes back to our initial findings that the majority of individuals open a business relatively late in life, implying that later cohorts might not have had the time yet to become entrepreneurs. When we consider the cohorts 1997 to 2004 in column (1), the effect of increasing the share of fathers peers' with a university diploma in humanities from 0% to 100% increases the probability to become an entrepreneur by 7.4 percentage points, but is not significantly different from zero. When we start to remove later cohorts- columns (2) to (4)- we see that the effect increases up to 31 percentage points (column (4)). If we assume that the estimated effect for the cohorts 1997-2001 is close to the true one if we could have followed individuals of these cohorts over their entire life-cycle, this implies that individuals being exposed to peers with all fathers graduated in humanities compared to none, increases the probability of ever becoming an entrepreneur by 31.5 percentage points. One has to consider, however, that the difference between the least and the most treated individual in our dataset -the difference between the 90th and 10th percentile - is 3.5 percentage points. Thus the difference in the probability of becoming an entrepreneur between an individual who is the least and the most treated is 1.1 percentage points. To put this into context, this corresponds to 20% of the overall share of entrepreneurs in the economy for the 1997-2001 cohort. This is an economically significant effect.

In the second row of Table 12, we report the regression coefficients of equation 4 in which we use the share of fathers peers' with a degree in business- instead of humanities- for different cohorts. We do this as an indirect test of our identification strategy, because as shown by Mertz et al. (2023) sons of business graduates are likelier to start a business. This also helps us to have a benchmark against which to compare the estimated coefficients for humanities. In column (1) we see that increasing the share of fathers peers' with a university diploma in business from 0% to 100% increases the probability to become an entrepreneur by 16 percentage points, which becomes 19 percentage points for cohorts 1997-2003 (column (2)). Comparing for the same cohorts the effect of increasing the share of fathers peers' with humanities or business degrees has similar effects in magnitude. This means that improving the communication skills of students who are good at math or providing more early exposure of students to an entrepreneurial environment has a quantitative comparable effect on spurring new business creation.

Finally, in the second and third row of Table 12 we show the same results but changing fathers' university degree to natural sciences and engineering respectively. We see that being expose to a higher share of students with fathers graduated in natural sciences has a negative effect, even if statistically not different from zero, on the probability of becoming an entrepreneur. This alines with our findings in section 7.2 that fathers graduated in natural sciences on average transmit higher mathematical skills to their sons, which alone negatively relate to self-selection into entrepreneurship. For fathers graduated in engineering the effect on the probability of ever becoming an entrepreneur is positive, but again not statistically significant.
Taken together, this evidence shows that mathematical high school students in their last year of

Table 12: Second stage regression

	1997-2004	1997-2003	1997-2002	1997-2001
	(1)	(2)	(3)	(4)
	Entrepreneur	Entrepreneur	Entrepreneur	Entrepreneur
1) Share_humanities	0.0737	0.165**	0.249**	0.316**
	(0.0649)	(0.0771)	(0.102)	(0.148)
2) Share_business	0.154**	0.190**	0.0184	0.161
	(0.0717)	(0.0872)	(0.132)	(0.157)
3) Share_natural_sciences	-0.0606	-0.0772	-0.0586	-0.0928
	(0.0727)	(0.0858)	(0.115)	(0.170)
4) Share_engineering	0.0509	0.0428	0.0728	0.223
	(0.0647)	(0.0764)	(0.107)	(0.151)
School f.e.	✓	✓	✓	✓
School f.e.×Share_HS+9	✓	✓	✓	✓
School f.e.×Share_HS+12	✓	✓	✓	✓
School f.e.×Share_HS+14	✓	✓	✓	✓
School f.e.×Share_HS+15	✓	✓	✓	✓
School f.e.×Share_HS+17	✓	✓	✓	✓
School f.e.×Share_HS+20	✓	✓	✓	✓
Father_year_of_education×education_area	✓	✓	✓	✓
High school municipality f.e.	✓	✓	✓	✓
N	29642	26291	23051	19668

Notes: $^{*}(p < 0.1),^{**}(p < 0.05),^{***}(p < 0.01)$. This table reports the OLS coefficients of the regressing the probability of ever becoming an entrepreneur on the share of peers whose father is graduated in humanities, controlling for school f.e., school f.e. interacted with share of education of peers' fathers, father year of education interacted with field of education and high school municipality f.e. In column (1) the sample contains cohort from 1997 to 2004, column (2) cohorts from 1997 to 2003, column (3) cohorts 1997 to 2003, and column (4) cohorts 1997 to 2002. In the second, third and fourth row we report the same OLS coefficients where instead of using the share of peers' fathers with a degree in humanities we use those with a degree in business, natural sciences and engineering respectively. The sample is the universe of male Danish Mathematical high school students in their last HS school year. *Entrepreneur* is 1 if an individual has ever been an entrepreneur in her life. Standard errors are clustered at year-school level and are: # (1) 1,109, (2) 969, (3) 829, (4) 691

school, who were exposed to a higher share of peers with fathers graduated in humanities - for a given school and given overall number of fathers with university diploma - have a significantly higher probability of becoming entrepreneurs compared to the rest of the population. The effect is sizable, being around 20% of the share of entrepreneurs in the economy for the 1997-2021 cohorts. The effect is also comparable, in terms of magnitudes, to exposing students to more entrepreneurial environments as measured by the share of fathers peer students' graduated in business.

9 Conclusions

In this paper we use Danish administrative data to provide new findings on the role of skill multidimensionality in explaining self-selection into entrepreneurship. We use detailed high school grades obtained by students in math and danish to measure analytical and communication skills. Observational evidence shows that individuals with specialized skills earn higher wages on the labor market, but are less likely to become entrepreneurs. For high talented math students in high school, the probability of starting a business is increasing in their communication skills. These students are also the ones owning the most profitable and successful businesses if they become entrepreneurs. To casually estimate the effect of improving communication skills of individuals with good analytical skills, we propose an identification strategy that exploits information on parents' human capital. Our findings show that teaching better communication skill to students who are very good at math in high school spurs business creation and that the effects are sizable. Our findings contribute to growing evidence on the role of human capital for entrepreneurial outcomes and inform the policy debate on the importance of education and training programs to incentivize the birth of successful entrepreneurs.

10 Appendix

The table below provides an overview of the structure of the Danish high school system up to 2004.

Figure 3: Danish Grading System 7-point grading scale

Grade	Description	ECTS	Old scale (00-13)
12	For an excellent performance displaying a high level of command of all aspects of the relevant material, with no or only a few minor weaknesses	A	13 11
10	For a very good performance displaying a high level of command of most aspects of the relevant material, with only minor weaknesses	B	10
7	For a good performance displaying good command of the relevant material but also some weaknesses	C	9 8
4	For a fair performance displaying some command of the relevant material but also some major weaknesses	D	7
02	For a performance meeting only the minimum requirements for acceptance	E	6
00	For a performance which does not meet the minimum requirements for acceptance	Fx	5 03
-3	For a performance which is unacceptable in all respects	F	00

General upper secondary education (gymnasium): distribution by subject of the total number of lessons per year (prior to the reform of 2005)

Subject	**Number of lessons per year in each form**		
	I	**II**	**III**
1. *Languages*:			
Beginner language	105	108	–
Visual arts	–	–	51
Biology	79	–	–
Danish language	79	81	102
English	105	108	–
Continuation language	105	108	–
Geography	–	81	–
History and civics	79	81	76
Physical education and sport	53	54	51
Latin	79	–	–
Music	79	–	–
Science	79	108	–
Classical studies	–	–	76
Religious studies	–	–	76
2. *Mathematics*:			
Beginner/continuation language	105	108	–
Visual arts	–	–	51
Biology	79	–	–
Danish language	79	81	102
English	79	108	–
Physics	79	81	–
Geography	–	81	–
History and civics	79	81	76
Physical education and sport	53	54	51
Chemistry	79	–	–
Mathematics	132	135	–
Music	79	–	–
Classical studies	–	–	76
Religious studies	–	–	76
3. *Optional subjects at advanced level*:			
Beginner language	–	–	127
Biology	–	135	127
English	–	–	127
Continuation language	–	–	127
Physics	–	–	127
Greek	–	135	203
Chemistry	–	135	127
Latin	–	135	127
Mathematics, mathematics stream	–	–	127
Mathematics, language stream	–	135	127
Music	–	135	127
Social studies	–	135	127

Source: Danish Eurydice Unit, 2005. Optional subjects are offered at an advanced and intermediate level. Concerning physical education and visual arts as optional subjects, the number of lessons may be combined with the lessons allocated to compulsory subjects. There must be 32 weekly lessons (each lasting 45 minutes) in the first year, and 31-32 in the second and third years. In the first year the 32 lessons are spent on compulsory subjects (27 in the second and 17 lessons in the third year).

Figure 2: Caption

References

Argan, Damiano, Leonardo Indraccolo, and Jacek Piosik, "Teach the Nerds to Make a Pitch: Multidimensional Skills and Selection into Entrepreneurship," 2023.

Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda, "Age and High-Growth Entrepreneurship," *2* (1), 65–82.

Blanchflower, David G and Andrew J Oswald, "What Makes an Entrepreneur?," *Journal of Labor Economics*, 1998.

Choi, Joonkyu, Nathan Goldschlag, John Haltiwanger, and J Daniel Kim, "Founding teams and startup performance," *Available at SSRN 3481850*, 2019.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The role of entrepreneurship in US job creation and economic dynamism," *Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi, "Learning entrepreneurship from other entrepreneurs?," *Journal of labor economics*, 2021, *39* (1), 135–191.

Lazear, Edward P, "Balanced skills and entrepreneurship," *American Economic Review*, 2004, *94* (2), 208–211.

__ , "Entrepreneurship," *Journal of Labor Economics*, 2005.

Levine, Ross and Yona Rubinstein, "Smart and illicit: who becomes an entrepreneur and do they earn more?," *The Quarterly Journal of Economics*, 2017, *132* (2), 963–1018.

Liang, James, Hui Wang, and Edward P Lazear, "Demographics and Entrepreneurship," *Journal of Political Economy*, 2018.

Lindquist, Matthew J, Joeri Sol, and Mirjam Van Praag, "Why Do Entrepreneurial Parents Have Entrepreneurial Children?," *Journal of Labor Economics*, 2015.

Lucas, Robert E, "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 1978.

Mertz, Mikkel Baggesgaard, Maddalena Ronchi, and Viola Salvestrini, "Early exposure to entrepreneurs, gender equality, and talent allocation in entrepreneurship," 2023.

Michelacci, Claudio and Fabiano Schivardi, "Are they all like Bill, Mark, and Steve? The education premium for entrepreneurs," *Labour Economics*, 2020, *67*, 101933.

Queiró, Francisco, "Entrepreneurial human capital and firm dynamics," *The Review of Economic Studies*, 2022, *89* (4), 2061–2100.

Silva, Olmo, "The Jack-of-All-Trades entrepreneur: Innate talent or acquired skill?," *Economics Letters*, 2007, *97* (2), 118–123.

Wagner, J., "Testing Lazear's jack-of-all-trades view of entrepreneurship with German micro data," *Applied Economics Letters*, 2003, *10* (11), 687–689.

Wagner, Joachim, "Are nascent entrepreneurs Jacks-of-all-trades? A test of Lazear's theory of entrepreneurship with German data," *Applied Economics*, 2006, *38* (20), 2415–2419.