CENTER FOR
ECONOMIC
BEHAVIOR &
INEQUALITY

CEBI WORKING PAPER SERIES

Working Paper 15/26

DOES HIGH SCHOOL GRADE VARIABILITY EXPLAIN LONG RUN LABOUR MARKET INCOME?

Asbjørn Juul Petersen

Jacob Richard Strabo

ISSN 2596-447X

CEBI
Department of Economics
University of Copenhagen
www.cebi.ku.dk

Does High School Grade Variability Explain Long Run Labour Market Income?*

Asbjørn Juul Petersen† Jacob Richard Strabo‡

July 10, 2026

Abstract

We investigate how performance variability is associated with long run labor market outcomes. Using administrative data on complete grade distributions for all general high-school graduates from 2001 to 2006, we measure within-student grade variability as the standard deviation of final grades and relate it to long-run labour market income. Conditional on identical GPA and a long list of background controls, a one–standard deviation increase in grade variability predicts a 4.8 percent lower income. The income differences are primarily driven by both extensive and intensive labor supply. Our results are robust in various specifications and even in a twin fixed effects analysis. In the second part of the paper, we further investigate potential mechanisms. We conjecture that performance stability is an underlying trait which impacts sorting behaviour into different higher education- and career tracks: Individuals with higher performance instability tend to sort into programmes and professions where instability is not punished or even rewarded but where the income trajectories are much more volatile. This differential sorting behaviour, especially into different fields and industries, can explain most of the income differences.

Keywords: grade variance, performance stability, education, human capital, income

JEL classifications: I26, J24, J31, I21, C23

*We are grateful for helpful discussions with Jesper Eriksen, Ida Maria Hartmann and Jeppe Søndergaard Johansen. Authors are listed in alphabetical order.

†University of Copenhagen and CEBI. The activities of CEBI are financed by the Danish National Research Foundation grant no. DNRF134. Email: `asbjornjp@econ.ku.dk`.

‡University of Copenhagen. Email: `jacob.strabo@econ.ku.dk`.

1 Introduction

High school grades, particularly the overall grade point average (GPA), are among the most powerful predictors of long-run labour market success. It measures important aspects of human capital, including general academic ability, its influence on access to higher education, and it serves as a signal of productivity to future employers (see e.g. Hansen et al. (2024), Landaud et al. (2024), Bulman (2017), and Borghans et al. (2016)). However, summarizing an entire record of grades into a single GPA hides differences in how students have achieved the same average, overlooking whether success stems from steady performances or sharp contrasts across subjects. Two students with identical GPA's can differ dramatically in the *variance* of their grades: one can be consistently average, while the other is a brilliant specialist in some subjects, but struggles in others. Whether this within-student variability in high school performances carries independent predictive power for long-run labour market success remains surprisingly unexplored.

There are strong reasons to expect that performance variability affects long-run labour market income, though the direction of this effect is ex-ante ambiguous. On the one hand, standard models of sorting and specialisation suggest that greater grade dispersion can lead to higher earnings: a student who excels in some courses but performs poorly in others can self-select into educational programmes and occupations where their comparative advantage is most highly rewarded, whereas a uniformly mediocre student cannot exploit such specialisation (see, e.g. Roy (1951), Willis and Rosen (1979), Kirkeboen et al. (2016), and Saltiel (2023)). On the other hand, many higher education programmes and jobs require competencies across a broad range of tasks, each demanding a minimum level of performance. From this perspective, performance stability may itself be a valuable trait for success in both education and the labour market (see, e.g. Kremer (1993) and Heckman and Kautz (2012)).

In this paper, we investigate how performance variability, measured as the standard deviation of high school grades, explains long-run labour market income. To examine this relationship, we combine rich longitudinal administrative data from Statistics Denmark with a special institutional feature of the Danish higher education system, namely that admission to most seats at university programmes is determined solely by GPA. By controlling for GPA-by-graduation-cohort fixed effects, we can compare individuals who not only have identical mean academic abilities but also have identical formal access to higher education, while differing in the stability of their performance across subjects. We also control for a rich set of demographic controls, being gender, ethnicity, and age, as well as socioeconomic background characteristics, being parental education and income, and also school fixed effects. This empirical specification closely resembles Sandsør (2020) who studies how grade variability impacts educational sorting in Norway.

Our sample includes all graduates from general upper-secondary (STX) schools in Denmark between 2001 and 2006. For each graduate, we observe the complete distribution of final grades, detailed background characteristics, and income measured 9 to 21 years after graduation.

We find that for a given GPA, an increase in grade variability of one standard deviation is associated with a reduction in long-run labour market income of 4.8 percent. The estimated coefficient is statistically significant and stable across different specifications once we include standard demographic controls (age, gender, and ethnicity). The estimated coefficients are surprisingly large and even larger than the association between fathers income and one's own long-run labour market income when measured on the same scale. The negative relationship between performance variability and long run income is present through the entire distribution of high school GPA's. This reveals that the variability of academic performance is an overlooked and extremely potent predictor of long-run labour market income. The results hold across a long range of alternative specifications, including the use of a different grading scale, and most notably, even in a twin fixed-effects analysis, showing that our results are robust even when holding genetic factors and parental values fixed (see, e.g. Ashenfelter and Krueger (1994) and Golsteyn et al. (2013)).

We further decompose the impact of grade variability on income into its components: (i) Time in Unemployment, (ii) Hours Worked, and (iii) Log Hourly Wage. This decomposition provides intuition about the underlying mechanisms, helping us understand which personal traits and aspects of individual behaviour drive the results. We find that higher grade variability is associated with more unemployment, fewer hours worked, and no significant effect on log hourly wage. This pattern is consistent with the idea that stable performers experience a stronger attachment to the labour market, including greater job stability and a higher labour supply, whereas students with more variable grades tend to have a weaker attachment to the labour market but not a lower hourly productivity when in work.

Finally, we investigate potential mechanisms behind our results. We show that grade variability also predict non-work related outcomes such as marital stability and number of address changes. This suggests that grade variability is not only important for labour market success but likely reveal something about underlying stability. We find that individuals to a large extent sort into different educational- and career tracks based on performance variability. We notice that differences in completed levels of higher education can only explain a small proportion of income differences, whereas sorting into different fields and industries appear to be a key driver of income differences.

To the best of our knowledge, we are the first to investigate how school performance variability is associated with long-run income. We primarily contribute to the literature examining how school grades predict later-life outcomes. Previous studies have focused on how grades from individual courses or overall GPA relate to labour market success (see, e.g. Hansen et al. (2024), Borghans et al. (2016), Landaud et al. (2024) and Bulman (2017)). A notable exception is Sandsør (2020), who uses data from Norwegian lower-secondary schools to study how grade variance relates to subsequent educational transitions. She finds that students with higher grade variation are less likely to complete high school and start a higher-education programme. Given that Norwegian students can choose specialised and vocational tracks immediately after lower-secondary school, and that admission to higher education in Norway, as in Denmark, is largely GPA-based, such sorting patterns are unsurprising as students with uneven academic performance are likely to self-select into alternative educational pathways that match their strengths. While our empirical approach builds upon Sandsør (2020), we focus on long-run labour market outcomes and not educational sorting. Nevertheless as a part of our analysis we extend and nuance the findings of Sandsør (2020) and show that grade variability negatively predict completion of higher education in general but that this relationship is reversed for the highest levels of education such as Master's- and PhD degrees. Additionally we show that sorting into different lengths of educational degrees only absorbs a negligible part of the income differences associated with grade variability.

More broadly, we contribute to the literature on how different skills are measured and rewarded in the labour market. In current days, many talks about how AI will influence labour markets in the future; however, before we can understand how shocks to the labour market, such as the automation of tasks with AI, will reshape the returns to human capital, we first need to understand which skills and traits are currently rewarded in the labour market. The idea that persistence is a trait associated with higher income links our work to the literature on non-cognitive skills. Heckman et al. (2006) shows that non-cognitive traits, such as self-esteem and locus of control, significantly predict labour market success. Furthermore, performance variability may be related to personality traits such as conscientiousness and emotional stability. While Fletcher (2013) finds that conscientiousness is positively associated with labour market outcomes, Nyhus and Pons (2005) reports no significant relationship but finds that emotional stability predicts better performance. Yet, no studies have measured school performance variability itself or assessed its predictive power for later labour market success.

Although the existing literature typically treats cognitive ability and personality traits as distinct dimensions of human capital, we combine the two by interpreting performance variability to reflect both ability and persistence. This highlights that human capital is inherently multidimensional and cannot be fully captured by a single summary measure such as GPA: studies that control only for GPA implicitly assume that doing so accounts for the relevant dimensions of human capital, yet our results show that grade variability captures something about ability and persistence that GPA alone does not. This dimension is important for several reasons. First, unlike many other non-cognitive traits, performance stability can be directly interpreted within a task-based framework. Autor et al. (2003) highlights that the general ability to perform a wide range of tasks at a consistent level is itself valued in the labour market. Second, performance variability does not rely on survey responses to abstract and low-stakes questions but is derived from (relatively) objective, high-stakes situations such as external exams and teacher evaluations, in which students have a strong incentive to perform well. It is therefore reasonable to view exam performance as a context in which individuals reveal their true abilities. Our study contributes to a deeper understanding of skills and the allocation of talent in society. Demonstrating that higher-order moments are strong predictors of future success can help admissions officers in both educational and employment settings better match applicants with suitable programmes and roles. Moreover, our findings highlight the importance of looking beyond first-order moments, such as GPA, when identifying which students may benefit most from additional support.

The remainder of the paper is structured as follows: Section 2 describes the Danish educational system. Section 3 presents the data, key variables, and descriptive statistics. Section 4 outlines the empirical strategy. Section 5 reports the main results and robustness checks. Section 6 investigates and discuss potential mechanisms. Section 7 concludes.

2 The Danish Educational System

In this section, we outline the Danish Educational System with a particular focus on the general high school, the grading scale used to evaluate students (up until 2006), and admission into higher education.[1] [2] [3]

The mandatory part of the Danish educational system consists of elementary school from grade 0 to grade 9, with the possibility of taking an optional 10th grade. After elementary school, students may choose either a vocational education track or an academic high school track, which is further divided into a general high school (STX), a business high school (HHX), a technical high school (HTX), and a two-year high school (HF). All four academic high-school programmes provide access to higher education, although with slight variations. In this paper, we focus exclusively on graduates of the general high school (STX), which is the choice of the majority of Danish high school students. This restriction is implemented to ease comparisons, as courses, structure, and grading practices differ between the four types of academic high schools.

2.1 The General High School

The General High School (STX) is a three-year academic programme combining mandatory and elective courses. Students chose between a mathematical and a linguistic track; yet, the two tracks share most courses and are both designed as broad, generalist programmes. Each track includes a mix of one-, two-, and three-year courses. At the end of each course, students receive one or more final-year grades based on the teacher's assessment of their performance. In several subjects, students receive both an oral and a written grade. These teacher-assessed grades contribute to the student's overall transcript and enter directly into the computation of the final grade point average (GPA). In addition to course grades, some subjects are selected for external examinations. Certain courses are always examined, while others are chosen randomly. Examinations can be written or oral, and each exam grade carries the same weight in the GPA as the corresponding course's teacher-assessed grades. The final GPA consists of approximately 30 grades covering various courses, including both exam grades and teacher-assessed grades.

[1] The descriptions of the general high school are based on: Uddannelsesguiden (2025)

[2] The description of the grading scale is based on Uddannelses- og Forskningsministeriet (2024).

[3] The description of the admissions system into higher education is based on Uddannelsesguiden (2024).

2.2 The Grading System

For both the exams and the teacher-assessed grades, students received a grade on the "13 scale". This scale consisted of 10 grades from 00 to 13 (00, 03, 05, 6, 7, 8, 9, 10, 11, 13). All grades of 6 and above are passing grades. At the end of high school, all the grades obtained (both the teacher-assessed grades and the exams) are used to calculate the final grade point average (GPA).[4] If the GPA is above 6, the student graduates from high school, even if they have a grade below 6 in some courses. This grading system was used until the 2006 high school graduating cohort. After 2006, the grading system was changed to the so-called "7 point scale," with grades ranging from -3 to 12.

2.3 Admission into Higher Education

Students apply for specific programme–institution combinations (for example, a BSc in Political Science at the University of Aarhus or a Professional Bachelor in Nursing at Copenhagen University College). Each student can apply for up to eight different combinations of programmes and institutions. They must prioritise them, but can only be admitted to one. If there are more applicants for a given programme-institution combination than available slots, admission is, in most cases, determined solely by high school GPA, with the exception that typically 10 percent of the places in a given programme may be allocated through so-called “Quota 2” criteria, wherein admission officers consider additional factors such as grades in specific subjects or admission tests. In most cases, places are offered exclusively to applicants with the highest GPA who have not been admitted to any, if at all, of their higher-priority choices. This implies that two applicants to the same programme–institution combination in a given year with identical GPA’s, measured to one decimal, have virtually the same probability of admission, regardless of other merits.[5]

3 Data

3.1 Data Source and Estimation Sample

We use administrative data at the individual level from Statistics Denmark. The estimation sample consists of all students graduating from general high school (STX) in the years 2001-2006 (N=78,376). 2001 is the earliest graduation cohort for which individual grades of the students are available. After 2006, the grading scale was changed, making it infeasible to compare cohorts before and after that reform, as the specific grading scale can influence the measured numeric variation in grades.

[4]Some courses might weigh more than others, depending on which academic level they are completed at.
[5]Some programmes require completion of specific high school courses.

3.2 Definition of Key Variables

In the following, we define and describe the key variables used for the analysis.

3.2.1 Performance Variability

To measure performance variability, we construct a variable $SD(Grades)$, defined as the standard deviation of all final grades obtained in high school by the individual student, including both teacher-assessed grades and final exam grades. In the regression analyses, we standardise this measure to have mean zero and unit standard deviation to ease the interpretation of the coefficient.

Using the standard deviation of grades to measure performance variability has several advantages. Each student receives a relatively large number of final grades in high school (approximately 30), which makes the measure more likely to capture actual performance variability instead of merely noise in grading. In addition, because grades are used for admission to higher education, students have a strong incentive to perform well. This means that the individual grades, and thereby our measure of performance variability, are likely to represent true student characteristics compared to, e.g. a measure based on low-stakes survey questions. Another advantage is that the general high school curriculum is broad, meaning that each student receives grades in many different subjects (e.g. English, Mathematics, History) and across a variety of assessment forms, including oral and written exams, teacher-assessed year marks, and project work. When combined, the grades cover a large variety of domains. This means that the grade standard deviation is arguably informative about students' performance variability across a range of different domains.

A weakness of using this measure of performance variability is that there is likely some stochasticity in student evaluations. This means that our measure of performance variability also contains an element of noise. This, however, is not a major concern, as it introduces an attenuation bias, which biases our estimates towards zero.

Another potential issue is that the standard deviation of the grades is closely tied to the specific grading scale used to evaluate the students. For example, a more fine-grained scale makes it easier to distinguish between students with similar performance levels. Moreover, the numerical values of the grades in the grading scale (including the numerical distance between grades) affect both GPA's and the standard deviation of grades. To ensure that our results are not driven by the specific grading scale, we conduct a robustness test by re-estimating the analysis for students graduating from high school between 2009 and 2011, as the Danish grading scale was reformed in 2007. Using these cohorts allows us to examine whether our results are driven by the specific grading scale used. These results are reported in table 7.

3.2.2 Outcome Variables

The main outcome variable is the logarithm of average labour market income in the years 2015-2022, referred to as *Log Income*. It is defined as the sum of salaries and other remuneration, as well as profits from one's own company, for self-employed individuals. We take the average over the years 2015-2022 to approximate permanent income as closely as possible, thereby avoiding influence of temporary fluctuations in income on the results. We take the log of average labour market income and winsorize income at the 99th percentile to ensure that the results are not driven by extreme observations.[6]

Since individuals in our sample graduated from high school between 2001 and 2006, income is observed on average 15.5 years after graduation (9–16 years for the 2006-cohort and 14–21 years for the 2001-cohort). Because all comparisons are made within cohorts, differences in the time since graduation are not a major concern. According to Chetty et al. (2014) and Chetty et al. (2017), measuring income even as early as age 30 does not introduce lifecycle bias. We therefore consider our income variable, measured around age 35, to be a reliable proxy for life-time labour market income.

In addition to the main outcome, we also examine its components: (i) the share of years between 2015 and 2022 spent in unemployment, with zero registered labour market income, referred to as *Zero-Income Share*; (ii) the average annual number of hours worked over the same period, referred to as *Hours Worked*; and (iii) the log of average hourly wage, defined as the log of average annual income divided by the average annual hours worked. We refer to this variable as *Log Hourly Wage*.[7] [8]

3.2.3 Other variables

We include a long list of control variables and fixed effects. We control for the exact high-school GPA used to apply for higher education, measured to one decimal. As the general level of GPA might change across different cohorts, e.g. due to grade inflation, we interact the GPA fixed effects with high-school cohort fixed effects. We also include a list of demographic controls, including birth-month-by-birth-year fixed effects, gender, and ethnicity. Moreover, we control for fathers' and mothers' highest education levels and log fathers' income as a proxy for household income in the years 1993–1997.[9] All parental controls are measured before the individuals start high school to avoid controlling for mediators.

[6] As we cannot take the log of individuals who have zero income in all of the eight years, we drop these observations. This, however, constitutes less than one percent of our sample. In section Appendix A we include observations with zero income in all of the eight years and estimate the regression equation in 4.1 for income in levels instead of logs and without winsorizing income. Our findings are robust to these alternative specifications.

[7] Log hourly wage is winsorized at the 99th percentile.

[8] To further ensure that our results are not driven by the specific functional form of our outcome variables, we further specify all four outcome variables in ranks and estimate the impact of grade variability on the labour market outcomes in ranks. The results of this analyses are shown in section Appendix A.

[9] Note that in the main analysis, we work with a balanced sample and exclude observations where one or more control variables are missing. In table A.2 we conduct a robustness test where these observations are included. The results are robust to this alternative specification.

3.3 Descriptive Statistics

In Table 1, we present descriptive statistics for the sample. The average GPA is 8.36, and the average grade standard deviation is 1.1. Our sample consists of 63 percent girls, which is consistent with the notion that girls are overrepresented in general upper-secondary schools in Denmark.[10] 19 percent of individuals in the sample have a father with a master's degree, and nine percent have a mother with a master's degree. The average age at high school graduation is 19.5 years.

Table 1: Descriptive Statistics

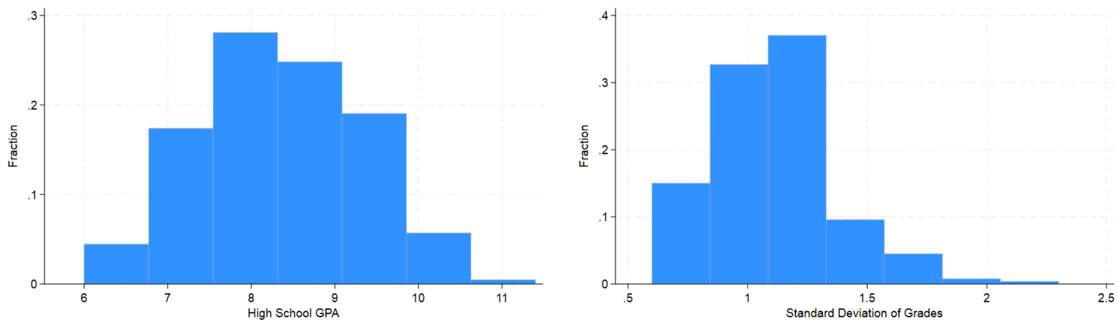
	Mean	Count (N)	SD
SD(Grade)	1.10	78,376	0.26
GPA	8.36	78,376	0.96
Income before Tax (DKK).	409,767.96	78,376	186,433.93
Zero-Income Share	0.04	78,376	0.14
Mean Hours Worked	1,539.22	78,376	505.97
Hourly Wage (DKK)	272.26	77,177	130.22
Female (dummy)	0.63	78,376	0.48
Graduation age (high school)	19.54	78,376	0.73
Age at income measurement	34.47	78,376	1.88
Father has Master's degree (dummy)	0.19	78,376	0.39
Mother has Master's degree (dummy)	0.09	78,376	0.29
Father's Yearly Income before Tax (DKK)	309,047.40	78,376	203,096.94
Observations (N)		78,376	

Notes: The table reports means, sample sizes, and standard deviations for the main variables used in the analysis. All monetary variables are in logs, unless specified with DKK. SD(Grade) refers to the within-student standard deviation in final grades. *Source:* Statistics Denmark and own calculations.

In figure 1, we report both the distribution of high school GPA's for the individual students in our sample (1a) as well as the distribution of the standard deviation of grades for the individual students in our sample (1b).

[10]See, e.g. the analysis made by Danske Gymnasier (2025).

Figure 1: Distribution of High School Grades



(a) Distribution of High School GPA's (b) Distribution of Standard Deviation of Grades

Notes: The figure shows the distribution of high school GPA's (panel a) as well as the distribution of individual standard deviation of grades ($SD(Grades)$ (panel b). The figures are based on the entire estimation sample of high school graduates from the general high school (STX) in the period 2001-2006. *Source:* Statistics Denmark and own calculations.

From figure 1a, we observe that the distribution of GPA is approximately normally distributed, with values ranging from 6 to more than 11. From figure 1b, we observe that there is substantial variation in the standard deviation of grades and that the vast majority of individuals have a standard deviation of grades between 0.5 and 1.5. The figures combined show that there is a substantial amount of variation in the first two moments of high school grades, which is needed to estimate the impact of performance variability.

4 Empirical Method

An obvious challenge in empirically estimating the relationship between performance variability and long-run income is that performance variability is likely correlated with other factors, which might, in turn, affect labour market outcomes; most importantly, it is arguably general ability and academic opportunities. The first concern of general ability arises almost mechanically from how performance stability is measured: students with very high ability are likely to obtain top grades in nearly all subjects and, due to ceiling effects in the grading system, appear as highly stable performers in our data. The second concern regarding academic opportunities is that performance stability may correlate with access to educational tracks, which, in turn, affects later earnings. To mitigate both concerns, we control for GPA-by-cohort fixed effects. (High school) GPA is typically used in the literature as a measure of general ability (see, e.g. Dale and Krueger (2002), Kirkeboen et al. (2016), and Hansen et al. (2024)). Additionally, admission into higher education in Denmark is based on high-school GPA. This means that by controlling for GPA-by-cohort fixed effects, we can keep the set of academic opportunities fixed. Two individuals graduating high school in the same year with the same GPA have almost identical admission probabilities for any given programme, regardless of their individual grade standard deviation or other performance metrics that might be related to performance variability (e.g. motivational essays, admissions tests, extracurricular activities) used to admit applicants in other countries.[11]

We estimate the following regression equation:

$$Y_{ikcj} = \alpha + \beta SD(Grade)_{ikcj} + GPA_k \times \lambda_c + \mu_j + X'_{ikcj}\Gamma + \varepsilon_{ikcj}. \tag{4.1}$$

Y_{ikcj} is the relevant labour market outcome of interest for individual i with a GPA k who graduates high school in cohort c of school j. $GPA_k \times \lambda_c$ is GPA fixed effects times cohort fixed effects. μ_j are school fixed effects. These are included as there might be differences in grading practices between high schools that might be systematically related to later income. The vector X_{ikcj} contains a rich list of standard controls for other factors that might be correlated with both performance stability and income, including demographic controls being age, gender, ethnicity, as well as and socioeconomic controls being father's and mother's levels of education and father's income.

$SD(Grades)$ is the standard deviation of the grades obtained by individual i and is our measure of performance variability. While we do not claim a strict causal interpretation of $\hat{\beta}$, it is still interesting and meaningfull. It measures the empirical relationship between performance variability and long-run income holding the other variables fixed. If β is different from zero, it means that performance variability holds predictive power over and above the included controls.

The interesting counterfactual is *not* "what happens if we exogenously change the standard deviation of grades?" it is more interesting to ask "do we see any meaningful income differences between individuals with different measures of performance variability holding fixed the standard confounders and measures of ability?" Equation 4.1 answers exactly that question. Because we flexibly control for GPA and a long list of other potential confounders, a significant estimate of β implies that variability in grades contains additional predictive power *over and above* (GPA) and the other potentially important confounders included in this study, such as standard demographics and parental backgrounds. This leaves us with two potential explanations: either performance *causally affects* long-run income, or it *proxies a trait* that does (and not the controls included in this regression). Either case is interesting and can be used to inform a discussion about which traits drive long-run labour market success. To alleviate the concern regarding potential confounders even further, we conduct a twin fixed-effects analysis in Section 5.4.1. By comparing twins, we additionally hold constant genetic factors and a shared childhood environment, making it more

[11] We note, however, that as around 10 pct. of college seats are distributed through quota 2 and hence using other metrics than purely GPA, the academic opportunities of two students with identical GPA's might differ slightly. This, however, will often be an advantage for a student with higher variation in performance. This is the case as quota 2 only considers grades from "relevant" courses; i.e., admissions to a BSc. in English consider high school grades in English and Danish to a greater degree than the GPA (see Københavns Universitet (2025).). This means that a high-variability performer, all else equal, will have a higher probability of being admitted through quota 2 into programmes in which they excel compared to a more stable performer. Additionally, admission into some programmes might require completion of specific courses, but these courses are closely related to the higher education programme, again favouring a high-variability performer specialising in their preferred field, all else equal.

plausible that differences in performance variability arise is not systematically related to background characteristics related to labour market outcomes.

5 Results

5.1 Main Results

In this section, we present the main results. We estimate regressions of log income on the standard deviation of grades and sequentially add more controls and fixed effects. In column (1), we include only the GPA with cohort fixed effects. In column (2), we include demographic controls (age, gender, and ethnicity). In column (3), we include socioeconomic controls (father's and mother's education and father's log income). In column (4), we include high school fixed effects, and this specification corresponds to the full equation in 4.1. The results are reported in Table 2 below.

Table 2: Main Results

	(1) Log Income	(2) Log Income	(3) Log Income	(4) Log Income
SD(Grade)	-0.0131***	-0.0498***	-0.0469***	-0.0476***
	(0.00335)	(0.00341)	(0.00341)	(0.00344)
Constant	12.78***	12.78***	12.06***	12.14***
	(0.0000119)	(0.0000135)	(0.0569)	(0.0582)
GPA × Cohort FE	✓	✓	✓	✓
Demographic Controls		✓	✓	✓
Socioeconomic Controls			✓	✓
School FE				✓
N	77,716	77,675	77,675	77,675
R^2	0.031	0.070	0.074	0.081

Standard errors in parentheses

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

From Table 2, we see that the association between the standard deviation of grades and income is negative and significant across all specifications at a 0.1 pct. level. After including standard demographic controls, the numerical coefficient on the standard deviation of grades increases dramatically in magnitude. The reason is likely that boys have higher grade variability (see, e.g. Oakley et al. (2024)), but also higher income. Hence, gender is an important confounder introducing a positive bias in the estimated coefficient. After this, the coefficient hardly changes after including a battery of socioeconomic controls and school fixed effects and stabilizes around –0.048. This means that an increase in the grade standard deviation of one standard deviation is associated with a decrease in long-run income of around 4.8 percent, conditional on having the exact same high-school GPA, measured down to one decimal point, and all other included controls.

While the results are statistically significant, we would like to assess whether they are also economically meaningful. To get an idea of the size of the coefficient, we can compare it to the coefficient on the log of father's income. The coefficient of log father's income on own log income is a measure of intergenerational elasticity in earnings. This coefficient has received a lot of attention in both academic work and policy discussions (see, e.g. Black and Devereux (2011) and Chetty et al. (2014)). We benchmark the coefficient on grade standard deviation against the coefficient on intergenerational persistence, using father's income as a proxy for the economic resources and opportunities transmitted across generations—factors that naturally correlate with a child's future earnings. Although this coefficient is not the primary focus of our analysis, comparing its magnitude to that of performance stability helps assess the economic relevance of grade variability.

Metaphorically, we conduct a reverse horse race: while a father's income is expected to "run forward" and positively predict a child's future labour market income, grade variability "runs backward," reflecting traits formed earlier in life. The question is: who runs the longest distance; Grade variability or your father's income?

In order to carry out this benchmarking analysis, we estimate a version of equation 4.1 in which we report both the coefficient of grade standard deviation and the logarithm of the father's income. We exclude the completed education of the father and mother from the regression, as those variables would likely capture some of the predictive power of the father's earnings since they are expected to be highly correlated. Additionally, we standardise the log of the father's income to obtain the variable in the same unit as the grade standard deviation, thereby making the comparison of the coefficient sizes meaningful. The coefficients for both grade standard deviation and the logarithm of the father's income are reported in table 3.

Table 3: Comparing Coefficient Sizes

	(1)
	Log Income
SD(Grade)	-0.0497***
	(0.00341)
Log Father's Income	0.0340***
	(0.00291)
Constant	12.78***
	(0.0000145)
GPA × cohort FE	✓
Demographic controls	✓
Socioeconomic controls	
School FE	✓
N	77,675
R^2	0.072

Standard errors in parentheses.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Notes: SD(Grade) and Log Father's Income are both standardized to have mean 0 and standard deviation 1. We control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

From Table 3, we see that the estimated coefficient on grade standard deviation is numerically larger than that of log father's income. Landersø and Heckman (2017) shows that the intergenerational elasticity is almost as large in Denmark as in the US when measured on gross income (excluding public transfers), as we do in this paper, making this comparison relevant beyond the Danish case. This comparison suggests that the association between performance variability and income is economically substantial relative to other key determinants of income. Our results in this section highlight that performance variability captures a trait that has a surprisingly large and long-run impact on labour market income.

5.2 Association Across the Grade Distribution

Next we investigate whether the estimated impact of performance variability on long run income is present throughout the entire distribution of grade point averages. If the empirical relationship is concentrated around specific regions it could suggest that our model is somehow misspecified and that the estimated coefficient to grade standard deviation is a result of the specific grading scale or some other unobserved, local factors which drive the relationship. If on the other hand, we find a strong link between grade variability and long run income for all levels of GPA it arguably speaks in favor of the conjecture that performance variability fundamentally is associated with long run income. We divide the sample into ten equally sized bins based on individuals' relative GPA rankings and re-estimate equation 4.1 separately for each decile. The results are reported in Table

4, where column (1) corresponds to the lowest GPA decile, column (2) to the second decile, and so forth.

Table 4: Association across Grade Distributions

	1 (0-10%)	2(10-20%)	3(20-30%)	4 (30-40%)	5 (40-50%)	6 (50-60%)	7 (60-70%)	8 (70-80%)	9 (80-90%)	10 (90-100%)
	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.	Log Inc.
SD(Grade)	-0.0345***	-0.0286**	-0.0371**	-0.0634***	-0.0469**	-0.0529***	-0.0503**	-0.0455***	-0.0810***	-0.0806***
	(0.00978)	(0.00811)	(0.0121)	(0.0102)	(0.0134)	(0.0104)	(0.0126)	(0.0106)	(0.0131)	(0.00933)
Constant	12.42***	11.90***	12.29***	12.14***	12.29***	12.35***	11.62***	11.92***	12.23***	12.21***
	(0.105)	(0.158)	(0.233)	(0.262)	(0.185)	(0.131)	(0.212)	(0.205)	(0.133)	(0.156)
GPA × Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Socioeconomic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
School FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	7,761	7,583	7,870	7351	8,099	8,257	7,373	7,487	8,008	7,695
R^2	0.090	0.101	0.094	0.094	0.081	0.091	0.108	0.093	0.105	0.115

Standard errors in parentheses

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: Each column reports estimates from separate regressions of log income on standardized grade variability within GPA-decile subsample. SD(Grade)is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, socioeconomic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

From table 4, we observe that the estimated coefficient for grade variability is negative and significant for all deciles of the grade distribution. This indicates that the observed link is not just a feature of the specific grading scale or driven by certain regions in the GPA-distribution, but is a phenomenon that is present everywhere in the distribution of grades and skills. We notice, however, that the association appears to be largest at the top of the distribution of GPA. In the regression on the ninth and tenth deciles, the estimated coefficient is approximately 0.08. It is, however, difficult to know whether the link is actually larger at the top of the distribution or whether it is due to the measure of both mean academic ability (GPA) and performance variability (grade standard deviation) becoming more imprecise and difficult to interpret at the top of the grade distribution due to ceiling effects. The issue of floor effects is not as severe in the lowest deciles. The reason is likely that, while all students in the sample have a GPA of at least 6 (only graduating students are included), they can still have some non-passing grades (00, 03, 05). It is, however, reassuring that the effects are present and statistically and economically significant across the entire distribution of GPA's.

5.3 Decomposition

Having established that the association between performance variability and income is negative, large and relatively stable across both different specifications and across the grade distribution we now investigate which income components drive the relationship. Mechanically, income can be decomposed into i) The share of years in unemployment, ii) average annual hours worked conditional on employment and iii) log hourly wage. We estimate equation 4.1 using each of these three outcomes as the dependent variable. This decomposition exercise is the first step towards understanding why we observe income differences between individuals who are otherwise similar but differ in their performance variability. The results are reported in table 5 below.

Table 5: Decomposition of Impact on Income

	(1) Zero-Income Share	(2) Hours Worked	(3) Log Hourly Wage
SD(Grade)	0.00977***	-30.82***	0.000564
	(0.000673)	(1.786)	(0.00149)
Constant	0.102***	1309.0***	5.259***
	(0.0122)	(32.83)	(0.0236)
GPA × Cohort FE	✓	✓	✓
Demographic Controls	✓	✓	✓
Socioeconomic Controls	✓	✓	✓
School FE	✓	✓	✓
N	78,320	77,121	77,121
R^2	0.021	0.059	0.174

Standard errors in parentheses

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, socioeconomic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

From table 5, columns (1) and (2), we see that higher grade variability is positively associated with a higher degree of unemployment (Zero-Income Share) and fewer hours worked conditional on employment. Hence, the overall impact of grade variability on income appears to be driven by a substantially lower labour market attachment for individuals with high performance variability both on the extensive and the intensive margin.

Empirically, it is difficult to distinguish between whether this weaker labour market attachment is voluntarily or involuntarily. We do not know whether high performance variability is associated with an intrinsic desire to supply less labour or whether employers are reluctant to higher workers who are more volatile in their performances. We can, however, conclude that individuals with high performance variability are less likely to obtain and sustain stable employment. We can also conclude that conditional on employment, performance variability is uncorrelated with log hourly wage. The estimated coefficient in table 5 column (3) is statistically insignificant, close to zero and estimated with a high degree of precision. This shows that conditional on employment there is no productivity penalty for high variability individuals.

5.4 Specification Checks

In this section we present the main specification checks to address potential concerns with our estimated results.

5.4.1 Twin Fixed Effects Analysis

Genetic factors or parental values beyond what is captured by parental education and income may confound the results presented in Section 5. Parents who place greater emphasis on performance stability may also be better positioned to support their children's labour market success, for instance, through professional networks. Similarly, genetic traits could be correlated with both performance stability and earnings potential. If so, the estimates in Section 5 partly reflect these underlying family-level dynamics. To address this concern, we re-estimate equation 4.1 on the subset of twins in our data and include twin fixed effects (see, e.g. Ashenfelter and Krueger (1994) and Golsteyn et al. (2013)). The underlying idea behind this analysis is that one twin may exhibit higher performance variability than the other, while both share the same genetic background and parental environment. We thus compare each individual only to their twin and estimate the following equation:

$$Y_{ikcjp} = \alpha + \beta SD(Grade)_{ikcjp} + GPA_k \times \lambda_c + \mu_j + \tau_p + X'_{ikcjp}\Gamma + \varepsilon_{ikcjp}. \quad (5.1)$$

This equation is similar to the one reported in 4.1, except for the fact that we include twin fixed effects τ_p; hence, we are only using within twin variation. The vector of controls includes only gender, as ethnicity, age, parental education, and income do not vary within twins. We estimate the equation separately using log income, unemployment, hours worked, and log hourly wages as the outcome variables. The results are reported in table 6.

Table 6: Twin Fixed Effects Analysis

	(1) Log Income	(2) Zero-income Share	(3) Hours Worked	(4) Log Hourly Wage
SD(Grade)	-0.126*	0.0365**	-51.01*	-0.0209
	(0.0496)	(0.0139)	(25.26)	(0.0240)
Constant	12.76***	0.0469***	1619.0***	5.535***
	(0.00457)	(0.00130)	(2.181)	(0.00207)
GPA × Cohort FE	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓
Socioeconomic Controls				
School FE	✓	✓	✓	✓
Twin FE	✓	✓	✓	✓
N	1,168	1,194	1144	1,144
R^2	0.776	0.755	0.829	0.796

Standard errors in parentheses.

$^+p < 0.10$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, gender, school fixed effects, and twin fixed effects. As ethnicity, birth year, birth month or parental background does not vary at the twin level these fixed effects are excluded as they are already captured by the twin fixed effects. Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

The results show that the estimated coefficients are qualitatively similar to those presented in table 2 and 5. The impact on log income and hours worked is negative and statistically significant at a five percent level, whereas the impact on the probability of having zero-income is positive and significant. The impact on the log hourly wage is insignificant and close to zero, similar to the estimate in table 5. Generally, the estimated coefficients are larger in magnitude compared to those presented in tables 2 and 5. This, however, might reflect a larger imprecision, as the sample of twins is relatively small, and the coefficients are not significantly different from those obtained in table 2 and 5 using the entire estimation sample. Overall, the twin-fixed effects analysis suggests that the results are not driven by confounders at the family level, which we have not accounted for in the main specification.

5.4.2 Utilizing a Grading Scale Reform

The results might be sensitive to the specific grading scale used to evaluate students. For instance, very high numerical values of top grades impact both the GPA and standard deviation of grades mechanically, but it also changes the incentives to study for certain subjects or exams depending on the perceived probability of obtaining the various grades. Thus the specific grading scale can distort and even drive the results in non-obvious ways. In order to make sure that our results are not driven by this specific grading scale but truly reflect an underlying relationship between performance variability and long run income we need to investigate whether our results can be replicated using a different scale. In order to do so we utilize a reform that changed the grading scale used in Denmark, implemented from the academic year 2006/2007. The new grading scale, the so-called 7-point scale, consists of the following grades: -3, 00, 02, 4, 7, 10, 12. Grades 02 and above are passing grades. This grading scale is comparable to the European Credit and Accumulation System (ECTS).[12] . We re-estimate equation 4.1 on a sample of high school graduates who graduated in the years 2009-2011 and were thus assessed on the new grading scale.[13] The results are reported in table 7.

[12]For more information about this grading scale, see Ministry of Higher Education and Science (2025).

[13]The labour market outcomes for this sample are only measured in 2021-2022, as many individuals are likely still pursuing higher education prior to these years.

Table 7: Robustness: New Grading Scale (Post-2007 Reform)

	(1) Log Income	(2) Zero-Income Share	(3) Hours Worked	(4) Log Hourly Wage
SD(Grade)	-0.0683***	0.0139***	-37.06***	-0.00501**
	(0.00533)	(0.00117)	(3.113)	(0.00182)
Constant	11.92***	0.104***	1141.7***	5.287***
	(0.112)	(0.0197)	(63.44)	(0.0319)
GPA × Cohort FE	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓
Socioeconomic Controls	✓	✓	✓	✓
School FE	✓	✓	✓	✓
N	51,833	53,329	50848	50,848
R^2	0.056	0.020	0.048	0.110

Standard errors in parentheses.

$^+p < 0.10$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. The outcome variables are only measured using the years 2021-2022, as the cohorts used in this analysis is younger (2009-2011). *Source:* Statistics Denmark and own calculations.

The results in table 7 show that the findings are robust when conducting the analyses on high school graduates who are assessed under the new grading scale. All the estimates are significant and have the same sign as those reported in table 2 and 5, except for the coefficient for log hourly wage, which is negative and significant. The coefficient size is only -0.005, so the impact on log hourly wage seems to be driving less than one-tenth of the total impact on log income.

5.4.3 Other Specification Checks

In appendix Appendix A we provide a long additional robustness tests including using different measures of both grade variability and the outcome variable, making sure that the results are not driven by specific subject compositions, and changing the estimation sample in various ways. Overall the main results are robust to all these alternative specifications.

6 Potential Mechanisms

In this section we investigate potential reasons behind this strikingly robust empirical relationship between performance variability and long run income. The purpose of this exercise is not to pinpoint exactly why income differences arise between individuals with high and low performance variability. Instead we suggest potential explanations, provide descriptive analyses to inform a general discussion and motivate directions for future research.

6.1 Performance Stability as a Personality Trait

A potential explanation behind these results is that performance variability in high school reveals an underlying personality trait of low stability in general and that this "stability-trait" is important for life trajectories but distinct from other measures of cognitive abilities which economists usually consider to be important on the labour market. This idea is motivated by the fact that performance variability does not significantly predict log hourly wage conditional on employment and the included controls in table 5.

In order to provide evidence consistent with the narrative that performance variability measures an underlying stability trait and that this is important we investigate whether our measure of performance variability can predict non-labour market outcomes where we *ex ante* would expect that stability is extremely important but (other) cognitive We estimate equation 4.1 using i) the probability of being in a stable marriage in 2022 (married for at least six years), ii) the probability of having experienced a divorce, and iii) the number of address changes from 2015-2022. These outcomes are arguably influenced by personal stability but not directly related to performance on the labour market. If performance variability is not a *deep* personal trait but only something which is important for labour market performances we wouls not expect grade variability to significantly predict these outcomes. The results are reported in table 8 below:

Table 8: Impact on Non-work Related Outcomes

	(1) Married >6 years	(2) Divorced	(3) # Address Changes from 2015-2022
SD(Grade)	-0.00818***	0.00285**	0.0176***
	(0.00197)	(0.00108)	(0.00439)
Constant	0.234***	0.0732***	1.509***
	(0.0341)	(0.0170)	(0.0797)
GPA × Cohort FE	✓	✓	✓
Demographic Controls	✓	✓	✓
Socioeconomic Controls	✓	✓	✓
School FE	✓	✓	✓
N	78,320	78,320	78,320
R^2	0.063	0.022	0.062

Standard errors in parentheses

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, socioeconomic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

The results in table 8 show that high school performance variability significantly negatively predict the probability of having been married for more than six years, and positively predict both the probability of being divorced and the number of address changes. This suggests that high school performance variability is a significant predictor also of important non-labour market outcomes and highlights that this performance variability is not something which is confined to the labour market. These findings are consistent with the notion that high school performance variability measures (the inverse of) an underlying stability trait and that this trait is important for the course of life trajectories in many domains.

6.2 Sorting into Different Life Trajectories Based on Performance Stability

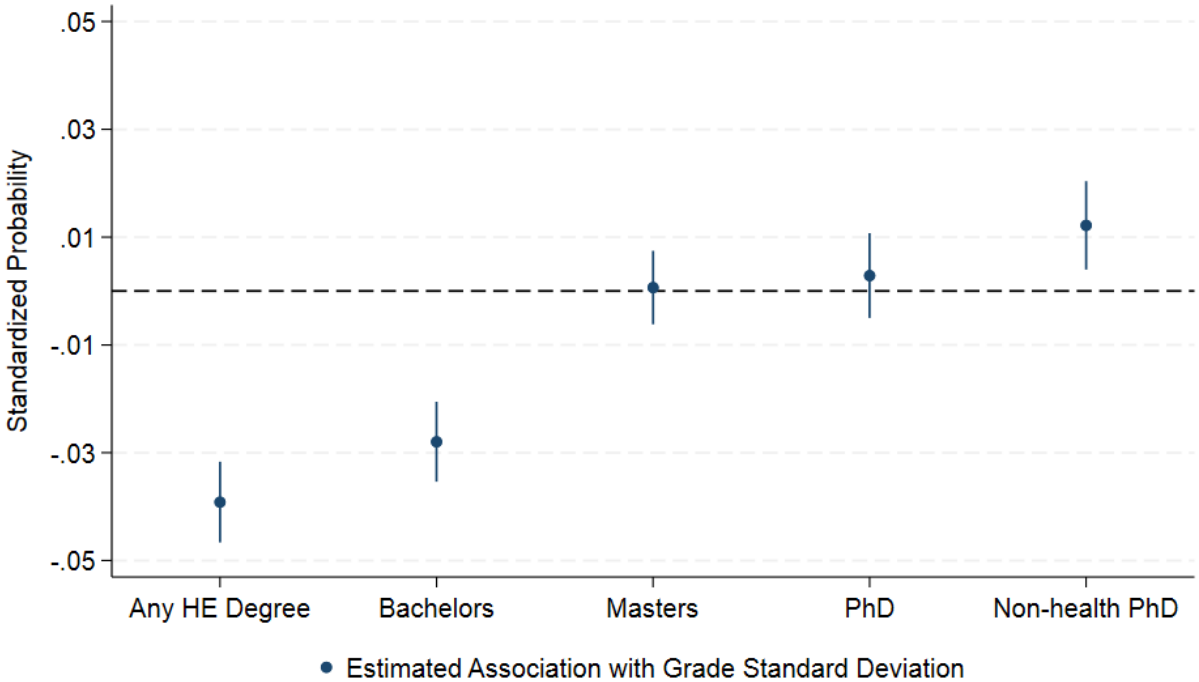
Having motivated that high school performance variability reveals an underlying personality trait related to stability in many dimensions, we now investigate how and whether students sort into different education- and career tracks based on this trait and whether these sorting decisions can explain the income differences which arise between high- and low performance variability individuals. We first document whether performance variability can explain sorting into different educational- and career tracks and finally we conduct a systematic, descriptive decomposition exercise to analyze how much of the income differences are absorbed by conditioning on sorting behavior.

6.2.1 Educational Sorting

We start our by analyzing how performance variability impact vertical educational sorting. I.e. how is grade variability related to the probability of completing any higher education and the level of highest completed education. In order to investigate this we estimate equation 4.1 using the probability of having completed *at least* the following level of higher education: i) Any higher education, ii) a Bachelor's degree, iii) a Master's degree, and iv) a PhD degree. Additionally we distinguish between a PhD degree and a PhD-degree obtained in a Non-Health related field. The reason is that amongst medical professionals, completing a PhD is very common and almost a prerequisite for obtaining a special doctor position — even if this position does not carry any research responsibilities. We therefore argue that the motives for obtaining a health PhD in a medical field is somewhat different from obtaining a PhD in other fields. [14]

[14] See e.g. Epinion (2016). The results are reported in figure 2.

Figure 2: Probability of Obtaining at least the Following Higher Education



Notes: The figure shows the standardized coefficients of regression equations of obtaining at least the following higher education degree. In all regressions we control for GPA times cohort fixed effects, demographic controls, socioeconomic controls and school fixed effects. *Source:* Statistics Denmark and own calculations.

The results show, that performance variability is negatively associated with obtaining any higher education degree. This is consistent with Sandsør (2020) who show that grade variability is negatively associated with starting a higher education. We complement her findings by showing that this also hold for *completion* of higher education. The figure, however, also reveals that when we consider higher levels of educational attainment the image is much more nuanced: While the coefficient is still significantly negative for obtaining at least a bachelor's degree, the coefficient is zero and even positive when considering the probability of obtaining Master's- and PhD-degrees. For non-Health PhD's the coefficient is statistically significant albeit small.

These results are consistent with the story that individuals sort into different educational tracjs based on stability: For lower levels of higher education we would expect stability in performances to be extremely important as shorter higher education degrees typically consist of mandatory coursework, many frequent assignments and exams, and broad curricula. For this type of work it is important to be able to deliver stable performances. Once we consider Master's- and especially PhD-degrees, however, we would expect stability to be less important. These degrees are filled with many elective courses, independent research oriented exams, the ability to come up with good ideas, and specialization into niche fields. For instance, when undertaking a PhD there is a lot of freedom of how to structure your work. You do not need to perform on a stable level in every subject everyday. Instead it is an advantage if you at some points or in some areas are able to come up with novel ideas which can advance the research frontier in that specific field. It therefore makes sense that individuals who are more volatile in their performances both have a higher probability of not completing any higher education — but also a higher probability of completing a PhD.

6.2.2 Labor Market Sorting

The results above suggest, that performance variability impact educational sorting in non-linear ways. It is thus highly likely that performance variability also impact sorting on the labour market into different professions in various ways. In this section we aim to shed light on this. Contrary to length of educational degrees, however, there exist extremely many different professions requiring different skill compositions, and sorting into all of these could potentially be impacted by grade variability. In order to select which professions to analyze we consider two seminal papers on labour market sorting and returns to different skill compositions: Lazear (2004) and Rosen (1981).

Lazear (2004) argue that individuals with a balanced skill set will sort into entrepreneurship, as founding and running a company requires a balanced skill set, whereas individuals who have an unbalanced skill set will sort into specialist occupations, where they can best utilize their comparative advantages and "hide" their comparative disadvantages. This theory would predict that performance variability is negatively associated with the probability of founding a company (and running a company as a CEO) but positively associated with the probability of entering into a specialist role.

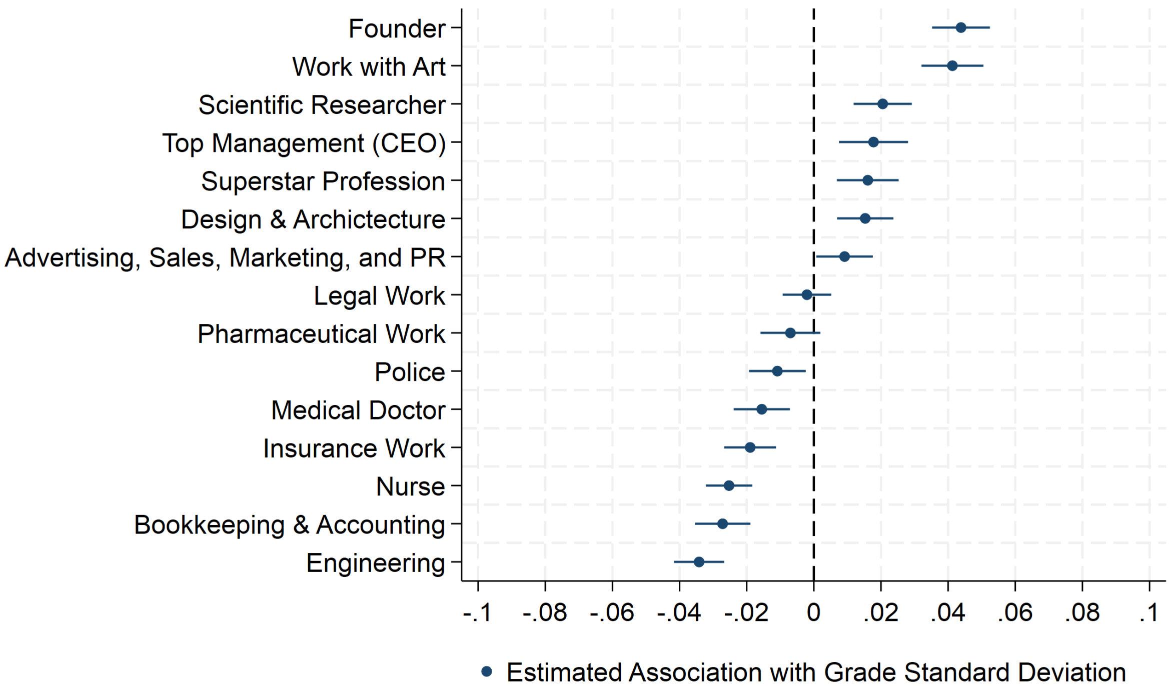
Rosen (1981) argue that in some highly specialized professions, the returns to being slightly better at specific skills than your competitors will lead to large increases in earnings. He mentions occupations such as artist, lawyer, scientific researcher, doctor, sports stars and performers. The key condition is that earnings are convex in these specific skills. We might imaigne that individuals with high performance variability are more likely to chase these niche but potentially high paying professions, as the potential earnings match their skill profiles.

In order to investigate whether performance variability can predict labour market sorting as theorized by Lazear (2004) and Rosen (1981) we investigate how performance variability is associated with the probability of sorting into i) entrepreneurial generalist roles (Founder and Top Management positions), ii) Specialist but non-Superstar roles (advertising, pharmaceutical work, policing, insurance work, nursing, accounting, engineering) iii) Specialist and superstar roles (art work, scientific researcher, design & architecture, legal work, Medical Doctor, and other superstar professions[15]) If performance variability drive labour market sorting in the manner of Lazear (2004) and Rosen (1981) we would predict performance variability to negatively predict the probability of ending up in generalist professions, more positively predict non-superstar but specialist professions and most strongly positively predict specialist and superstar professions. We estimate equation 4.1

[15]Other superstar professions include professional athletes, actors, and movie theater directors.

using the probability of employment in each of the 15 professions mentioned above as the outcome variables. The coefficient to grade variability are reported in figure 3 below:

Figure 3: Grade Standard Deviations and job



Notes: The figure shows the standardized coefficients of regression equations of obtaining at least the following higher education degree. In all regressions we control for GPA times cohort fixed effects, demographic controls, socioeconomic controls and school fixed effects. *Source:* Statistics Denmark and own calculations.

The results in figure 3 does not appear to be consistent with sorting based on generalist or specialist competences. Performance variability positively predict the probability of both becoming a fonder and hold a top management role both of which Lazear (2004) mention as generalist professions. At the same time high performance variability negatively predict the probability of ending up in roles which are typically considered highly specialized fields such as engineering and accounting. The results also appear inconsistent with sorting based on performance variability into superstar professions: While performance variability positively predict the probability of employment in some superstar professions such as working with art or scientific research, the coefficient is negative to other of such professions such as medical doctor and legal work.

Instead the results are consistent with a notion, that high performance variability individuals sort into professions in which they are evaluated based on their *best* performances such as artist or researcher. On the other hand performance variability loads negatively on professions where workers are evaluated on their ability to not make mistakes such as doctors, accountants and engineers. This is consistent with the idea that individuals sort into professions based on an underlying trait of stability. For instance, individuals who are volatile in their performances and less stable tend to sort into professions where they are not punished for their instability but instead rewarded for their occasional high performances: E.g. an artist does not need to perform well everyday but is only judged on their best work. On the other hand, individuals who are stable in their performances sort into professions where they are rewarded for their ability to not make mistakes such as accounting and engineering.

While the results above are only suggestive as it is based on a small subset of potential professions we investigate whether these patterns cary over to the rest of the labour market: The professions which load positively on performance variability tend to be "boom-or-bust professions". I.e. success in these professions depend on the ability to at some point and in some domain to come up with a brilliant idea or otherwise perform at an extraordinary level. If individuals with high performance variability generally sort into more volatile professions we would expect performance variability to positively predict *variation* in income both across individuals and within individuals over time. For instance for every successful artist there are likely many individuals who try to become professional artists but fail and instead end up in a low paying job. Even succesful artists might have struggled many years before becoming succesful. In order to investigate this, I estimate equation 4.1 using i) the squared prediction error on income, ii) the standard deviation of employment for each individual in the years 2015-2022, and iii) the standard deviation of log income conditional on employment in the years 2015-2022. The results are reported in table 9 below:

Table 9: Grade Standard Deviation and Variation in Income

	(1)	(2)	(3)
	Across Individuals Variation	$SD(Participation_{i,2015-2022})$	$SD(\text{Log Income}_{i,2015-2022})$
SD(Grade)	0.0482***	0.00842***	0.0323***
	(0.00284)	(0.000607)	(0.00225)
Constant	0.736***	0.117***	0.771***
	(0.0447)	(0.0113)	(0.0409)
GPA × Cohort FE	✓	✓	✓
Demographic Controls	✓	✓	✓
Socioeconomic Controls	✓	✓	✓
School FE	✓	✓	✓
N	77675	78320	77427
R^2	0.029	0.024	0.035

Standard errors in parentheses

$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). *Source:* Statistics Denmark and own calculations.

The results in table 9 show that grade variability positively predict variation in income both across and within individuals. This indicates that individuals with high performance variability generally tend to sort into career tracks which are more volatile, have less job security and less (paid) hours.

6.2.3 Sequential Conditioning

We have shown that performance variability impact sorting into different education- and career tracks. We now perform a descriptive decomposition exercise to investigate how much of the income differences on 4.8 pct. associated with a standard deviation increase in performance variability found in table 2 are absorbed once we condition sequentially on various educational- and career related decisions. We start out by estimating the main equation 4.1. Then we include fixed effects for vertical educational sorting (completed educational level fixed effects). Next we condition on both vertical and horizontal educational sorting by including educational level fixed effects interacted by educational field fixed effects. Next we condition on chosen industry by adding industry fixed effects.[16] Finally we condition on employment and hours worked by investigating log hourly wage. In each step we investigate how the coefficient to performance variability, $\hat{\beta}$, changes as we condition in increasingly identical choices. While this exercise obviously does not provide any causal answers it is useful to descriptively analyze how much of the baseline association each conditioning absorbs as it can motivate in which stage of choices income differences arise. The results are reported in table 10 below:

Table 10: Sequential Conditioning

Specification	$\hat{\beta}$ on SD(Grade)	t-value	Pct. of baseline remaining	Share absorbed
Baseline	-0.048	-13.8	100 %	-
+ Completed Education level FE	-0.038	-11.2	79 %	21 pp.
+ Completed Education Level×Field of Study FE	-0.020	- 6.3	42 %	37 pp.
+ Industry FE	-0.013	-4.6	27 %	15 pp.
+ Hours worked	+0.004	+3.2	- 8 %	35 pp.
N=77675				

Notes: The first column shows the estimated coefficient to SD(Grade) in a regression of log income. SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). In the second column we report the t-value associated with the coefficient. In the third column we report how much of the baseline size of the coefficient which remains, and in the fourth column we report in percentage points how much of the original coefficient which was absorbed in that step. *Source:* Statistics Denmark and own calculations.

The results in table 10 show, that conditioning on the level of obtained education only absorbs around one fifth of the baseline coefficient and still leaves the coefficient highly significant and economically large. This suggests that income the association between performance variability and long run income is not simply a story of educational attainment. This is consistent with the findings in figure 2 which reveal that high performance variability is only negatively associated with educational attainment for lower levels of higher education. For higher levels the association is zero or even positive. Once we include the field of study and indsutry fixed effects, however, the coefficient is attenuated by 37 pp. and 15 pp. respectively. These results indicate that a large proportion of the income differences between low- and high variability performers lie in their horizontal sorting decisions in terms of which field and indsutry to pursue. We note, however, that

[16]We condition on industry in 2015.

even after conditioning on both educational field and indsutry, $\hat{\beta}$ is still negative and significant — even at 0.1 pct. level. Finally, we condition on hours worked. Ths flips the sign of the coefficient and shows that conditional on having attained the same education (both length and field), working in the same indsutry, and working the same amount of hours individuals with higher performance variability hav higher hourly wages. This could reflect, that if individuals are able to signal to employers that they are stable by their educational- and working decisions, the employers actually value performance variability.

7 Concluding Remarks

We find that performance variability is strongly and negatively associated with long-run income for individuals measured in their mid-thirties, which serves as a reliable proxy for long-run labour market income (Chetty et al. (2014), Chetty et al. (2017)). The magnitude of the coefficient is comparable to that of parental income, underscoring the economic relevance of performance stability early in life. Decomposing the association, the results suggest that early-life performance variability predicts a weaker labour market attachment in the form of both a higher risk of unemployment and fewer hours worked. Overall, our findings highlight that stability in performance, rather than peaks of excellence, may be a stronger foundation for long-term economic success. Further analyses suggest that performance variability predict non-labour market outcomes as well and that individuals sort into different educational- and career tracks based on this performance variability. Our study contributes to a greater understanding of skills and the allocation of talent in society. Demonstrating that higher-order moments are strong predictors of future success can help admissions officers in both educational and employment settings better match applicants with suitable programmes and roles. Moreover, our findings highlight the importance of looking beyond first-order moments, such as GPA, when identifying which students may benefit most from additional support.

REFERENCES

Orley Ashenfelter and Alan Krueger. Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, pages 1157–1173, 1994.

David H Autor, Frank Levy, and Richard J Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333, 2003.

Sandra E Black and Paul J Devereux. Recent developments in intergenerational mobility. *Handbook of labor economics*, 4:1487–1541, 2011.

Lex Borghans, Bart HH Golsteyn, James J Heckman, and John Eric Humphries. What grades and achievement tests measure. *Proceedings of the national Academy of Sciences*, 113(47):13354–13359, 2016.

George Bulman. Weighting recent performance to improve college and labor market outcomes. *Journal of Public Economics*, 146:97–108, 2017.

Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American economic review*, 104(5):141–147, 2014.

Raj Chetty, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. The fading american dream: Trends in absolute income mobility since 1940. *Science*, 356(6336):398–406, 2017.

Stacy Berg Dale and Alan B Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117(4):1491–1527, 2002.

Danske Gymnasier. Gymnasiet i tal. `https://danskegymnasier.dk/gymnasiet-i-tal/`, 2025. Accessed 28 November 2025.

Epinion. Undersøgelse af den lægevidenskabelige ph.d.-uddannelse. Technical report, Styrelsen for Videregående Uddannelser, December 2016. `https://ufsn.dk/media/qc0fbzvg/epinion-undersogelse-af-den-laegevidenskabelige-ph-d-uddannelse`

Jason M Fletcher. The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, 89:122–135, 2013.

Bart HH Golsteyn, Hans Grönqvist, and Lena Lindahl. Time preferences and lifetime outcomes. Technical report, IZA Discussion Papers, 2013.

Anne Toft Hansen, Ulrik Hvidman, and Hans Henrik Sievertsen. Grades and employer learning. *Journal of Labor Economics*, 42(3):000–000, 2024.

James J Heckman and Tim Kautz. Hard evidence on soft skills. *Labour economics*, 19(4):451–464, 2012.

James J Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482, 2006.

Lars J Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

Michael Kremer. The o-ring theory of economic development. *The quarterly journal of economics*, 108(3):551–575, 1993.

Københavns Universitet. Kvote 2 på engelsk. `https://www.ku.dk/studier/bachelor/kvote-2/kvote-2-intervie` 2025. Accessed 28 November 2025.

Fanny Landaud, Éric Maurin, Barton Willage, and Alexander Willén. The value of a high school gpa. *Review of Economics and Statistics*, pages 1–24, 2024.

Rasmus Landersø and James J Heckman. The scandinavian fantasy: The sources of intergenerational mobility in denmark and the us. *The Scandinavian journal of economics*, 119(1):178–230, 2017.

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

Ministry of Higher Education and Science. Grading system - english. `https://ufm.dk/en/education/the-danish-education-system/grading-system`, 2025. Accessed 28 November 2025.

Ellen K Nyhus and Empar Pons. The effects of personality on earnings. *Journal of economic psychology*, 26(3):363–384, 2005.

Claire M Oakley, Reinhard Pekrun, and Gijsbert Stoet. Sex differences of school grades in childhood and adolescence: A longitudinal analysis. *Intelligence*, 107:101857, 2024.

Sherwin Rosen. The economics of superstars. *The American economic review*, 71(5):845–858, 1981.

Andrew Donald Roy. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2): 135–146, 1951.

Fernando Saltiel. Multi-dimensional skills and gender differences in stem majors. *The Economic Journal*, 133(651):1217–1247, 2023.

Astrid Marie Jorde Sandsør. Jack-of-all-subjects? the association between individual grade variance and educational attainment. *Economics of Education Review*, 75:101969, 2020.

Uddannelses- og Forskningsministeriet. The danish grading scale of 00 to 13. https://ufm.dk/uddannelse/anerkendelse-og-dokumentation/dokumentation/karakterskalaer/13-skala 2024. Accessed 28 November 2025.

Uddannelsesguiden. Sådan søger du optagelse på videregående uddannelser. https://www.ug.dk/artikel/saadan-soeger-du-optagelse-paa-videregaaende-uddannelser, 2024. Accessed 28 November 2025.

Uddannelsesguiden. Eu-ophold som arbejdstager. https://www.ug.dk/uddannelser-til-unge/gymnasiale-uddar 2025. Accessed 28 November 2025.

Robert J Willis and Sherwin Rosen. Education and self-selection. *Journal of political Economy*, 87 (5, Part 2):S7–S36, 1979.

Appendix A Additional Robustness Tests

In this section we provide additional robustness tests. We first consider whether our main results are robust to the specific way we measure the outcome variable and performance variability. In order to investigate this, we estimate equation 4.1 where we i) use income in levels measured in Danish kroner (DKK) as the outcome variable, ii) measure income as rank placement from 0 to 1 in the income distribution, iii) use the amount of public benefits paid in DKK as a proxy for unemployment, and iv) measure performance variability as the absolute difference between the highest and lowest grade obtained on the high school transcript. The results are reported in table A.1 below:

Table A.1: Other Outcomes and Alternative Measure of Grade Variability

	(1) Income (DKK)	(2) Income (Rank)	(3) Public Benefits (DKK)	(4) Mean Log Income
SD(Grade)	-11126.6***	-0.0167***	1800.2***	
	(1116.0)	(0.00122)	(133.8)	
Highest Grade Minus Lowest Grade				-0.0289***
(0.00248)				
Constant	171993.6***	0.186***	57919.7***	12.27***
	(18734.1)	(0.0211)	(2802.7)	(0.0579)
GPA × Cohort FE	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓
Socioeconomic Controls	✓	✓	✓	✓
School FE	✓	✓	✓	✓
N	78,320	78,320	78,320	77,675
R^2	0.128	0.167	0.197	0.080

Standard errors in parentheses.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Notes: Income in column (1) is defined as the mean annual income in DKK. Income in column (2) is the rank placement of individuals from 0 to 1 in the income distribution. Public benefits is defined as the annual amount of public benefits received by the government (in DKK). SD(Grade)is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). Standard errors are clustered at the GPA by cohort level. Highest Grade Minus Lowest grade is measured as the absolute difference between the highest grade on the high school transcript minus the lowest grade on the high school transcript. *Source:* Statistics Denmark and own calculations.

The results in table A.1 show that our main results are robust to this alternative definitions of both the outcome variable and measure of performance variability.

Next we investigate whether the results are robust to adjusting the sample and to controlling for subject composition in various ways. We estimate equation 4.1 where we i) include observations who have missing values on one or more of the control variables, ii) exclude individuals who have one or more years in unemployment to make sure that the results are not *entirely* driven by individuals who have low labour market attachment, iii) exclude mathematics grades from the calculation of grade standard deviation, and iv) control for the composition of high school subjects by including fields of study fixed effects.[17]

Table A.2: Adjusting the sample and controlling for composition of high school subjects

	Including Missings	Excluding individuals in part. unemployment	Excluding Mathematics	Controlling for course composition
	(1)	(2)	(3)	(4)
	Log Income	Log Income	Log Income	Log Income
SD(Grade)	-0.0556***	-0.0159***	-0.0561***	-0.0600***
	(0.00356)	(0.00199)	(0.00321)	(0.00531)
Constant	12.22***	12.51***	12.15***	11.98***
	(0.0484)	(0.0337)	(0.0582)	(0.109)
GPA × Cohort FE	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓
Socioeconomic Controls	✓	✓	✓	✓
School FE	✓	✓	✓	✓
N	90,892	67990	77675	51704
R^2	0.080	0.158	0.082	0.087

Standard errors in parentheses.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Notes: SD(Grade) is standardized to have mean 0 and standard deviation 1. In all regressions we control for high-school GPA-by-cohort fixed effects, demographic controls, and school fixed effects. Demographic controls include gender, ethnicity, and birth year by birth month fixed effects. Socioeconomic controls include mother and father's highest education, and log father's income (measured 1993–1997). We impute missing values by zero and include dummy variables for each of the control variables for whether a given observation has a missing value of each control variable. In column (2) we exclude all individuals from the sample who have at least one year between 2015-2022 in unemployment. In column (3) we exclude mathematics from the calculation of grade standard deviation. In the specification in column (4) we include dummy variables for the study line. I.e. for the three main elective courses. Note that this is based on 2009-2011 high school graduates as this information is not available for the 2001-2006 cohorts. Standard errors are clustered at the GPA by cohort level. *Source:* Statistics Denmark and own calculations.

The results in table A.2 show that the results are robust to changing the sample and controlling for the composition of subjects in various ways.

[17]Note that we can only control for the composition of high school subjects by controling for the high school fields of study for the 2009-2011 cohorts as this information is not available for the 2001-2006 cohorts.