Grade variance*

Astrid Marie Jorde Sandsør†

Feburary, 2016

Abstract

This paper investigates the importance of the second moment of individual grade distribution; grade variance. Transcript data from the U.S. National Longitudinal Survey of Youth, 1979, along with detailed register information for students in Norway are used to investigate the association between grade variance and educational attainment. For both the United States and Norway, grade variance is negatively associated with educational attainment across the grade distribution. Estimates are robust to controlling for socioeconomic characteristics and school fixed effects and remain negative for both genders and when including measures of cognitive and non-cognitive skills. My results suggest that institutions should consider more than just grade point average in admission decisions.

Keywords: grades, cognitive skills, non-cognitive skills, human capital
JEL codes: I21, J24

*Thanks to Angela Duckworth and the Character Development in Adolescence Project as well as Torberg Falch and the project Governance, Management and Performance in the Norwegian Educational System financed by the Norwegian Research Council (grant no. 197760) for allowing me to use their data. Thank you to Torberg Falch, Kalle Moene and Alexander Koch as well seminar participants at the Norwegian University of Science and Technology and the Workshop on Economics of Education, Mainz, September 2015 for helpful comments and suggestions. The paper is part of the research activities at the ESOP center at the Department of Economics, University of Oslo. ESOP is supported by The Research Council of Norway (S/179552).

†Department of Economics, University of Oslo

1 Introduction

What are the effects of the individual distribution of skills on school attainment and school performance? We know that cognitive skills are an important predictor for future outcomes for the individual, including education and labor market outcomes (Murnane, Willett, and Levy, 1995; Herrnstein and Murray, 2010; Heckman, 1995), and aggregate measures of cognitive skills are important for economic growth and development (Hanushek and Woessmann, 2008; Hanushek and Kimko, 2000). However, for a given average level of skills, is it better that skills are evenly divided between subject areas or is it better to be particularly good at some subject area?

One measure of cognitive skills is student grades received in school, commonly measured as the grade point average. Grades are highly correlated with short-term and long-term outcomes such as educational attainment and income. Additionally, grades have direct consequences for students, by for instance forming part of the college admission decision and determining their post-education job qualifications. Grade point average captures the first moment of the individual grade distribution, the mean. The second moment of the distribution, the variance, is a measure of grade dispersion; how far the grades are from the individual's mean. For a given grade point average, which student might be expected to have higher educational attainment; the student with high or low grade variance?

On the one hand, grades might reflect non-cognitive skills, such as motivation, perseverance and conscientiousness which have been shown to be meaningful predictors of educational, labor market and behavioral outcomes. If high grade variance is associated with low non-cognitive skills and vice versa, then a negative relationship between grade variance and educational attainment is expected. On the other hand, grades might mainly reflect knowledge in the subject, i.e., cognitive skills. As higher education allows students to specialize in their preferred field, high variance students, who are particularly good in some subjects, might be expected to have a higher educational attainment.

As there are reasons to believe that grade variance could be either positively or negatively associated with educational attainment, this makes grade variance particularly interesting to study empirically. Finding a negative association between grade variance and educational attainment, especially at the lower end of the grading distribution, supports the non-cognitive skills hypothesis while finding a positive association, especially at the upper end of the grading distribution, supports the generalist/specialist hypothesis.

In order to investigate the importance of grade variance empirically, I use three different data sources; The U.S. National Longitudinal Survey of Youth, 1979 (NLSY79),

Norwegian register data (NRD) and data from the Character Development in Adolescence Project (CDAP). The NLSY79 is a longitudinal survey with a nationally representative sample of young Americans first interviewed in 1979 and includes high school transcript data, educational attainment and socioeconomic characteristics. The NRD contains the entire population of students graduating from lower secondary education in Norway from 2002-2004 and includes transcript data, educational attainment and socioeconomic characteristics. The CDAP is a longitudinal survey of middle school students and their teachers from 8 different schools and includes transcript data along with various self-reported and teacher-reported measures of non-cognitive skills.

The NLSY79 and NRD are both used to investigate the association between grade variance and educational attainment and whether the association differs across the grading distribution or by gender. The NLSY79 includes long-run educational outcomes while the NRD only includes short-run educational outcomes. In Norway, grades are the main determinant of acceptance into upper secondary and higher education, and grading practices are monitored by central authorities, reducing potential measurement error. Along with the richness of register data, this allows for a more detailed analysis in the NRD than in NLSY79. By investigating data from two different countries, I am able to investigate whether the association between grade variance is context specific or more general.

Next, the paper investigates how grade variance is associated with cognitive and non-cognitive skills. The NLSY79 includes measures of cognitive and non-cognitive skills previously used by Heckman, Stixrud, and Urzua (2006) while a subset of grades is used as measures of cognitive and non-cognitive skills in the NRD. However, in both data sets the measures of cognitive and non-cognitive are simple and may not be capturing the skills that could be expected to be associated with grade variance. The CDAP includes grades together with a rich set of non-cognitive skills measures allowing for a more robust analysis of non-cognitive skills and grade variance.

For both the United States and Norway, grade variance is found to be negatively associated with educational outcomes. In the NLSY79, grade variance is negatively associated with educational attainment. In the NRD, grade variance is negatively associated with (1) starting the academic track in upper secondary, (2) upper secondary grade point average, (3) graduating from the academic track in upper secondary and (4) continuing on to higher education. Estimates are robust to controlling for socioeconomic characteristics and school fixed effects in the NLSY79 and school by cohort fixed effects in the NRD. The estimate for grade variance is negative across the grading distribution for both countries and no significant differences

are found between boys and girls.

The association between grade variance and educational outcomes remains negative when including measures of cognitive and non-cognitive skills. In the NLSY79, the estimate for grade variance is reduced when adding cognitive skills but remains unchanged when adding non-cognitive skills. In the NRD, adding cognitive and non-cognitive measures do not change results in a systematic way. The CDAP data confirm that grade variance does not seem to be related to non-cognitive skills. While the negative association between grade variance and educational attainment supports the non-cognitive skills hypothesis, all results are robust to adding measures of non-cognitive skills which does not support this hypothesis. My results support the alternative hypothesis that being a generalist rather than a specialist is beneficial for educational attainment.

The paper proceeds as follows. Section 2 discusses why one might expect grade variance to matter. Section 3 presents the main analysis for the NLSY79 data while section 4 presents the main analysis for the Norwegian register data. Section 5 investigates whether the importance of grade variance depends on the grading distribution, gender and cognitive and non-cognitive skills using all data sources. Section 6 presents the conclusion.

2 Grade variance

Standardized tests, such as the PISA test and the SAT,[1] are designed to be able to determine a student's skills in the specific subject relative to all other students. Grades, however, are a much more subjective measure. Grades are usually decided by the teacher of the subject, are not standardized across classes and schools and can be absolute measures or measured relative to classmates. They are often a combination of knowledge in the subject (cognitive skills) and other skills such as showing up to and participating in class (non-cognitive skills) (Borghans, Duckworth, Heckman, and Ter Weel, 2008; Segal, 2012; Kautz, Heckman, Diris, ter Weel, and Borghans, 2014). In addition, the degree to which cognitive or non-cognitive abilities matter will depend on the subject. Falch, Nyhus, and Strøm (2014), for instance, use math and science grades in school as a proxy for cognitive skills while they use grades in physical education, food and health, arts and crafts and music as a proxy for

[1] The Programme for International Student Assessment (PISA) is a standardized test carried out every three years among a representative sample of 15 year olds, and measures their competency in mathematics, reading and science. Around 510,000 students in a total of 65 countries participated in PISA in 2012 (OECD, 2015). The SAT is a standardized test developed to test students' academic readiness for college. The SAT, along with the ACT, form a large part of the admission decision for many colleges (ACT, 2015; SAT, 2015).

non-cognitive skills.

On the one hand, grades might reflect non-cognitive skills, such as motivation, perseverance and conscientiousness [2]. Non-cognitive skills have been shown to be meaningful predictors of educational, labor market and behavioral outcomes (Kautz, Heckman, Diris, ter Weel, and Borghans, 2014; Heckman, Stixrud, and Urzua, 2006; Borghans, Duckworth, Heckman, and Ter Weel, 2008; Carneiro, Crawford, and Goodman, 2007; Falch, Nyhus, and Strøm, 2014). Also, non-cognitive abilities have been shown to be more important for the lower part of the skill distribution (Lindqvist and Vestman, 2011). If high grade variance is associated with low non-cognitive skills while low grade variance is associated with high non-cognitive skills, then high grade variance is expected to be associated with low educational attainment, especially at the lower end of the grading distribution. This is the non-cognitive skills hypothesis.

On the other hand, grades might reflect knowledge in the subject, i.e. cognitive skills. High grade variance students have both good and bad skills (specialists) while low grade variance students have more similar skills across subjects (generalists). As higher education allows students to specialize in their preferred field, high variance students might be expected to have a higher educational attainment. This might be especially true for students at the upper end of the grade distribution as these students are more likely to go on to higher education. This is the generalist/specialist hypothesis.

However, it is not clear that being a specialist is always most beneficial. It might be beneficial to be a generalist for some studies or occupations (Lazear, 2004) or it might be beneficial to be a generalist in the long run due to greater adaptability (Hanushek, Woessmann, and Zhang, 2011). Lazear (2004) finds that individuals with balanced skills (jacks-of-all-trades) are more likely to become entrepreneurs. The idea is that rather than having a comparative advantage in a specific skill, entrepreneurs have a comparative advantage in having a span of skills, which is necessary to be successful as an entrepreneur. Being a jack-of-all-trades might be beneficial for the educational outcomes studied in this paper. Higher education is often based on general knowledge suggesting that generalists might be better at higher education. This could particularly be true in the United States where there is

[2]Non-cognitive skills are referred to as soft skills, personality traits, non-cognitive skills, non-cognitive abilities or character and socio-emotional skills, among others. Heckman and Kautz (2013) refer to them as character skills, rather than traits, as they are constant at any age but may change over time. Character skills include "conscientiousness, perseverance (grit), self-control, trust, attentiveness, self-esteem, self-efficacy, resilience to adversity, openness to experience, empathy, humility, tolerance of diverse opinions and the ability to engage productively in society" (Heckman and Kautz, 2013, p. 6).

a long tradition for a liberal arts education in four-year colleges. The specialist might therefore see the benefit of a short specialized education rather than a long general one. Hanushek, Woessmann, and Zhang (2011) study the impact of vocational versus general education, and find that although individuals with vocational education have an early labor-market advantage due to for instance higher employability, these gains are often offset by reduced adaptability later in life. Being a generalist could be more beneficial for long-run outcomes due to greater adaptability. This is the reversed generalist/specialist hypothesis.

It might also be the case that the association between grade variance and educational attainment differs by gender. A common finding is that while average skill differences between boys and girls tend to be small, the variance of skills is higher for boys than for girls.[3] Although variance across individuals is higher among boys than girls, there is no reason to believe that individual variance is higher for boys than for girls. Even if individual grade variance is higher for boys, it does not necessarily mean that the association between grade variance and education attainment, conditional on grade point average, varies by gender. However, if grade variance to a greater degree reflects being a generalist or specialist for one gender, while it reflects high or low non-cognitive skills for the other gender, results may differ for boys and girls.

Finding a negative association between grade variance and educational attainment, especially at the lower end of the grading distribution, supports the non-cognitive skills hypothesis while finding a positive association, especially at the upper end of the grading distribution, supports the generalist/specialist hypothesis. Also, results could differ by gender if grade variance reflects being a generalist or specialist for one gender, while it reflects high or low non-cognitive skills for the other gender.

Finally, measures of cognitive and non-cognitive skills are added to the analysis. Grade point average might not be the best measure of cognitive skills. Roth, Becker, Romeyke, Schäfer, Domnick, and Spinath (2015) investigate the relationship between standardized intelligence tests and school grades employing a psychometric meta-analysis and find a population correlation of $\rho = .54$, suggesting that grade point average only proxies as a measure for cognitive skills. Adding improved measures of cognitive skills might therefore strengthen the analysis. Non-cognitive skills are added to the analysis to see whether they explain part of the association between grade variance and educational outcomes. If they do, this suggests that grade vari-

[3]Hedges and Nowell (1995) study six representative large scale surveys with data on mental abilities and find that although average sex difference generally are small, males consistently have larger variance in test scores.

ance is capturing a measure of non-cognitive skills and supports the non-cognitive skills hypothesis. If a negative association is found between grade variance and educational outcomes, but results remain unchanged when adding non-cognitive skills, we are left with the reversed generalist/specialist hypothesis that being a generalist rather than a specialist is beneficial for educational attainment. These potential mechanisms are investigated in Section 5.

One concern is that even if we find an association between grade variance and educational attainment, the coefficient for grade standard deviation is picking up a mechanical correlation between grade standard deviation and grade point average due to for instance ceiling effects. By controlling for grade point average, the analysis compares students with the same grade point average, but with different grade variance. However, ceiling effects could affect the association at the lower or upper end of the grading distribution. To investigate whether we are picking up such mechanical effects, the samples are separated into medians and quartiles and separate regressions are run. Finding similar results across all samples removes much of the concern for ceiling effects. Also, in the Norwegian sample, students are bunched at certain values of grade point average where they have exactly the same grade point average but different grade variance. Running a regression for each of these values isolates grade variance from grade point average. Again, finding similar results for all subsamples removes much of the concern for ceiling effects. For more details and results, see Section 5.

3 Grade variance in the United States

In the following, the main results from the National Longitudinal Survey of Youth, 1979 (NLSY79) are presented. The NLSY79 is a longitudinal survey with a nationally representative sample of young Americans and includes high school transcript data, educational attainment and socioeconomic characteristics.

3.1 Institutional setting in the United States

Each state is divided into several school districts, which have jurisdiction over school curricula, budgets and policies for the public schools. State governments set the overall educational standards and funding for education is a combination of funding from the federal, state and local government. About 10% of students attend private schools (National Center for Education Statistics, 2015) which are free to determine their own curriculum. Compulsory education varies by state, starting between ages five and eight and ending between ages 16 and 18, and may be completed in

public schools, private schools or though approved home school programs. Most schools divide their schooling into three levels: elementary school, middle school and high school. "There is no uniform configuration throughout the country in the organization of primary and secondary education. Elementary school begins with kindergarten, but may continue through grades 5, 6, or 8 ... High school typically begins at grade 9 or 10, with middle or junior high schools usually covering the intervening years between elementary school and high school. Students graduate from high school following grade 12". (Stevenson and Nerison-Low, 2002, pp. 15-16) Usually, children are divided into grades by age groups, starting with kindergarten, and then continuing from grades 1 (age 6) to 12 (age 17), where grade 12 is the final year of high school.

A student completing high school will receive a high school diploma, while those students who have not completed high school, or do not meet the requirements for the diploma, have the option of passing a General Education Development (GED) test, a high school equivalency credential. After high school, students may continue on to post-secondary education at colleges or universities. When applying to higher education, the major determinants for admission are grades in college preparatory courses, test scores from the ACT or SAT, and overall grades. Class rank, an application essay or writing samples and letters of recommendation may also be admission criteria (Clinedinst and Hawkins, 2011). Colleges are usually either two-year colleges (community college or junior college) or four-year colleges. Two-year colleges provide academic, vocational and professional education rewarding associate degrees and some students will transfer on to a four-year college. Four-year colleges usually reward a bachelor degree qualifying students for graduate schools where master and doctoral degrees are rewarded.

With this as the institutional background, the analysis uses data on grades received in high school and data on educational attainment, measured as years of completed schooling. A high school degree is equivalent to 12 years of completed schooling while completing a four year college is equivalent to 16 years of completed schooling.

3.2 Data from the National Longitudinal Survey of Youth, 1979

The National Longitudinal Survey of Youth, 1979 (NLSY79) is a survey with a nationally representative sample of 12,686 young Americans between ages 14 and 22 who were first interviewed in 1979. The survey collects information on parental background, schooling decisions, labor market experiences, cognitive and non-cognitive

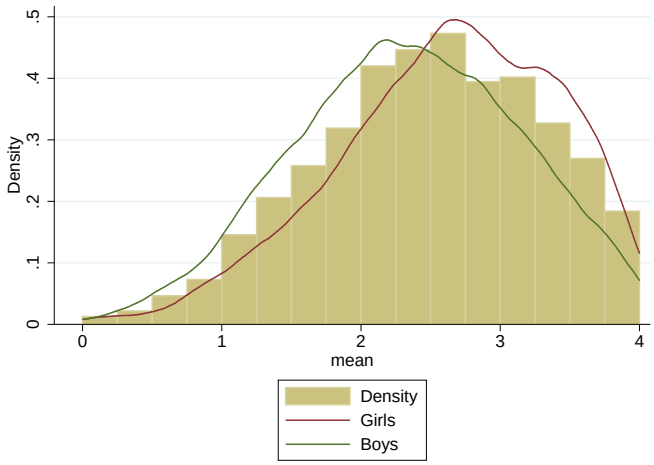
test scores and other behavioral measures on an annual basis. Between 1980 and 1983, transcript information was collected with data on each grade received during high school. See Appendix A for a detailed description of the transcript data.

The sample consists of three sub samples: (1) a cross sectional sample of 6,111 respondents from the non-institutionalized segment of the population (2) a supplemental sample of 5,295 Hispanic, Latino, black and economically disadvantaged non-black/non-Hispanic respondents, and (3) a sample of 1,280 respondents enlisted in the military as of September 30, 1978. Following the 1984 interview, most of sample (3) and parts of sample (2) were dropped from the survey. Following Heckman, Stixrud, and Urzua (2006), the main sample with 6,111 respondents is used in the analysis.

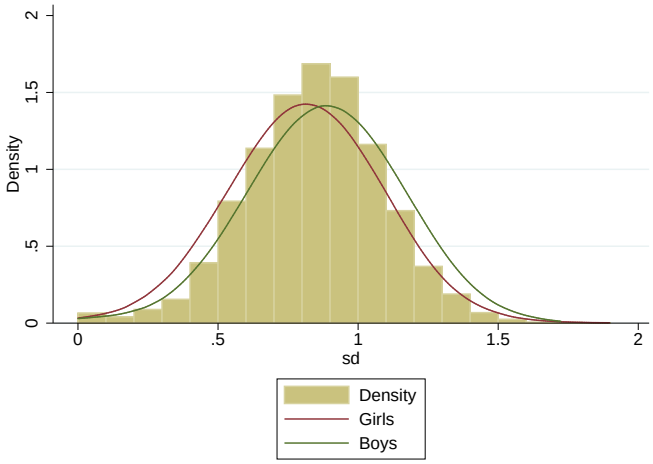
Grade point average (GPA) is measured as the unweighted mean of all grades received in all years of high school (grades 9-12), and is restricted to students with at least 10 valid grades. Grade variance is measured as the standard deviation of an individual's grades (GSD), using the same grades as were used to calculate the individual's grade point average. Descriptive statistics for the transcript data are reported in panel A of Table 1 with the last columns presenting descriptive statistics for girls and boys separately. In the regressions, both GPA and GSD are standardized with mean 0 and standard deviation 1 to facilitate interpretation. The average number of grades is 25.8 with a standard deviation of 6.32. Students either receive a pass/fail grade, or grades A-F, where grade A is coded to value 4, grade B to value 3, grade C to value 2, grade D to value 1. Grade F is a failing grade and is coded to value 0. Figures A1 and A2 in Section A display the distribution of grades and the distribution of number of grades respectively.

Figure 1a displays the distribution of GPA with the red and green lines displaying kernel densities with a bandwidth of 0.15 for girls and boys respectively. Average GPA is higher for girls (2.62) than for boys (2.33) while the spread is slightly higher for boys (standard deviation of GPA is 0.79 for girls and 0.81 for boys). These are both common findings in the literature (Herrnstein and Murray, 2010). Figure 1b displays the distribution of the GSD. Once again, red and green lines displaying kernel densities with a bandwidth of 0.15 for girls and boys respectively. Average GSD is higher for boys (0.88) than for girls (0.81) while the spread in GSD is the same (standard deviation of GSD is 0.24 for both girls and boys).

Figure 1: NLSY79



(a) Distribution of grade point average



(b) Distribution of grade standard deviation

Note: The figure includes 4,389 students from the main sample with 10 or more valid grades and with non-missing educational attainment at age 30. For grade point average, each bin has a width of 0.25, while each bin has a width of 0.1 for grade standard deviation. Lines display kernel densities with bandwidth 0.15 for each variable for girls (red) and boys (green).

The outcome of interest is educational attainment and is measured as years of education at age 30, measured from 1 in 1st grade to 20 in the 8th year of college. Average years of education is 13.5 with a standard deviation of 2.22 (Panel B of Table 1). Educational attainment is similar for boys and girls, while the standard deviation is higher for boys (2.33 for boys and 2.11 for girls). Socioeconomic characteristics include number of siblings, father's highest grade completed, mother's highest grade completed and family income in 1979 as well as a dummy for broken home at age 14, a dummy for living in the south at age 14 and a dummy for living in an urban area at age 14, race and ethnicity dummies. Cohort fixed effects are in-

cluded in all specifications where cohort corresponds to birth year. The measures of socioeconomic characteristics correspond to those in Heckman, Stixrud, and Urzua (2006). Descriptive statistics are listed in panel C of Table 1 with last columns of Table 3 presenting descriptive statistics for girls and boys separately.

3.3 Empirical strategy and results

Ideally, we would like to have exogenous variation in grade variance to capture the causal effect of grade variance on educational attainment. However, it is hard to find such variation. Instead, the association between GPA and GSD is estimated using an OLS model controlling for socioeconomic characteristics and including cohort fixed effects. In order to interpret this model as causal, all relevant variables that are correlated with both GSD and educational attainment must be included in the analysis, which is likely not the case. This model therefore expresses the association between GSD and educational attainment, conditional on socioeconomic characteristics and cohort fixed effects.

The outcome variable, y_{it}, is years of education by age 30 for individual i born in year t. GPA_{it} is grade point average and GSD_{it} is grade standard deviation, where each variable is standardized with mean 0 and standard deviation 1. The model includes individual socioeconomic characteristics, $X_t^{'}$, listed in Table 1, and cohort fixed effects, δ_t, in correspondence with Heckman, Stixrud, and Urzua (2006). The error term, ϵ_{it}, is clustered at the cohort level. The model can be expressed as

$$y_{it} = \alpha GPA_{it} + \gamma GSD_{it} + X_{it}^{'}\beta + \delta_t + \epsilon_{it} \tag{1}$$

The variable of interest is γ, which is the conditional correlation of GSD and outcome y, once GPA and other variables are controlled for. If γ is positive, a student with the same GPA but with higher GSD is expected to have more years of education by age 30 whereas a negative γ indicates the opposite.

The results are presented in Table 2 where all columns include cohort fixed effects. The first two columns present a simple OLS regression with GPA as an explanatory variable with and without socioeconomic characteristics. As expected, GPA is positively correlated with educational attainment, with a one standard deviation increase in GPA predicting 1.2 years more of education by age 30. This corresponds to 0.55 of a standard deviation increase in years of education. The estimate remains stable when controlling for socioeconomic characteristics.

In the next columns, the variable of interest, GSD, is added to the model. The coefficient for GSD in columns (3) and (4) tells us how grade standard deviation

Table 1: NLSY79 - Descriptive statistics

	Total		Boys		Girls	
	mean	(sd)	mean	(sd)	mean	(sd)
A. Transcript data						
Grade point average (GPA)	2.48	(0.81)	2.33	(0.81)	2.62	(0.79)
Grade standard deviation (GSD)	0.84	(0.25)	0.88	(0.24)	0.81	(0.24)
Number of grades	25.8	(6.32)	25.6	(6.44)	26.0	(6.20)
B. Outcome variable						
Years of education	13.5	(2.22)	13.5	(2.33)	13.6	(2.11)
C. Socioeconomic characteristics						
Girl	0.51	(0.50)	0	(0)	1	(0)
Black	0.11	(0.31)	0.11	(0.31)	0.11	(0.31)
Hispanic	0.061	(0.24)	0.061	(0.24)	0.061	(0.24)
Living in south	0.30	(0.46)	0.29	(0.45)	0.32	(0.47)
Living in urban area	0.76	(0.43)	0.76	(0.43)	0.76	(0.43)
Broken home	0.22	(0.41)	0.21	(0.41)	0.22	(0.41)
Number of siblings	3.20	(2.14)	3.14	(2.13)	3.25	(2.15)
Month of birth	6.45	(3.38)	6.49	(3.41)	6.41	(3.34)
Family income 1979 (thousands)	17.0	(15.1)	17.7	(15.3)	16.4	(15.0)
Mother: Years of education	11.3	(3.47)	11.3	(3.62)	11.4	(3.33)
Father: Years of education	11.2	(4.64)	11.3	(4.69)	11.1	(4.59)
D. Cognitive skills						
Arithmetic reasoning (ASVAB 1)	18.2	(7.19)	19.2	(7.34)	17.3	(6.92)
Word knowledge (ASVAB 2)	26.4	(7.12)	26.3	(7.35)	26.5	(6.90)
Paragraph comprehension (ASVAB 3)	11.2	(3.17)	10.8	(3.34)	11.5	(2.97)
Mathematical knowledge (ASVAB 4)	46.6	(15.2)	42.8	(14.8)	50.3	(14.7)
Coding speed (ASVAB 5)	14.1	(6.31)	14.4	(6.50)	13.8	(6.10)
Cognitive	0	(1.00)	-0.042	(1.05)	0.041	(0.95)
E. Non-cognitive skills						
Rotter locus of control scale	7.56	(2.38)	7.62	(2.36)	7.50	(2.39)
Rosenberg self-esteem scale	22.7	(4.05)	22.9	(3.96)	22.5	(4.12)
Non-cognitive	0	(1.00)	0.046	(0.98)	-0.045	(1.02)

Note: N=4,389 for the whole sample, with 2,234 girls and 2,155 boys. N=4,243 for the cognitive measure and N=4,225 for the non-cognitive measure. N=4,136 when combining the cognitive and non-cognitive measures.

Table 2: NLSY79 - Years of education by age 30

	(1)	(2)	(3)	(4)
Grade Point Average	1.223***	1.096***	1.092***	0.972***
	(0.029)	(0.026)	(0.032)	(0.028)
Grade Standard Deviation			-0.242***	-0.238***
			(0.033)	(0.034)
Socioeconomic Characteristics	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes
R-squared	0.304	0.391	0.312	0.399
N	4,389	4,389	4,389	4,389

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the cohort level.

predicts educational attainment when controlling for GPA. The coefficient for grade standard deviation is -0,242 without controlling for socioeconomic characteristics and -0.238 when controlling for socioeconomic characteristics, indicating that the result is not driven by some sub-sample of students. The coefficient for GPA is only slightly lower when including GSD in the specification.

In the NLSY79, results show that for a given grade point average, students with higher variance complete fewer years of education than students with low grade variance. If GSD increases by one standard deviation, educational attainment is reduced by 1/4 of a year. This corresponds to 0.11 of a standard deviation decrease in years of education.

4 Grade variance in Norway

To provide a similar and comparable investigation of Norway, I use Norwegian Register Data (NRD). Comparing results from the NRD to those from the NLSY79 indicates whether the results are country and context specific or more general. For instance, upper secondary and higher education in Norway has a high degree of tracking, which is not the case for the United States. According to the specialist/generalist hypothesis, high grade variance might therefore be associated with high educational attainment in Norway and the opposed to the United States.

Using the NRD has clear benefits. Firstly, the data cover the entire student cohort for three years, a sample of over 150,000 students. Secondly, the data include school identifiers so that school by cohort fixed effects can be added to the analysis. Thirdly, admission into upper secondary education and higher education is centralized and almost entirely based on GPA. It is therefore less likely that important variables are omitted from the analysis when including GPA and socioeconomic

characteristics along with GSD as the only measures determining educational attainment. Lastly, grading is monitored by the central government which reduces concerns of measurement error.

4.1 Institutional setting in Norway

There are clear institutional differences between Norway and the United States. In Norway, municipalities (428) are responsible for primary and lower secondary education, while counties (19) are responsible for the upper secondary education. Compulsory education consists of primary education (grades 1-7) and lower secondary education (grades 8-10), and ends the year the student turns 16 years of age, and entrance into primary and lower secondary education is determined by catchment areas. There is no possibility to fail a class in primary or in lower secondary education during the empirical period, implying that all students finish compulsory education on time.[4] There is no tracking, a common national curriculum for all students and very few private schools, with only 3.5 % of students attending a private elementary or lower secondary school in 2015 (The Norwegian Directorate for Education and Training, 2015).

Children do not receive grades in primary education.[5] In lower secondary education, students receive grades from their teachers every semester, primarily based on their performance in the subject. These grades have no consequences for the students prior to grade 10. Grades received in the last semester of grade 10, along with 2-3 externally graded oral or written exams, are used to determine acceptance to upper secondary education. Students are only tested in theoretical subjects on the exams, and the subject to be tested is decided by a draw. The written exams are the same nationally for all students taking the specific subject, while the oral exams are organized locally. The externally-graded grades are averaged with the teacher-graded grades in the corresponding subjects. The unweighted grade point average of the resulting grades is used to determine acceptance into upper secondary education.

Students may choose from 3 study tracks qualifying for higher education, and 12 vocational study tracks. When applying for upper secondary education, students

[4]In very few cases, students do not start primary education at the expected age, which implies that they finish lower secondary education at different age. If a child is not considered to be mature enough, the parents together with the school and psychologists can postpone enrollment one year. In addition, some older students return to improve their grades, and immigrants are often over-aged at graduation.

[5]Students in the highest grades of elementary education will in some cases receive grades as preparation for lower secondary education. The grades have no direct consequences for the students.

rank their preferred study tracks and schools within study tracks. All students have been guaranteed admission to upper secondary education since 1994, but whereas acceptance to one of their three ranked choices is guaranteed, the grade point average determines which school and study program the student is accepted to. How important grades are for entering the school or study program of their choice will vary from county to county as counties are free to determine how acceptance into upper secondary education is organized (Haraldsvik, 2003).

In upper secondary education, academic tracks have a duration of 3 years while vocational tracks typically last for 4 years, including 2 years of apprenticeship training. Subject requirements differ depending on the study program and there are both mandatory and elective subjects. If students from vocation tracks want to continue on to higher education, they can attend a year of supplementary studies qualifying for higher education.

The application system to higher education is centralized for the entire country and is solely based on grade points.[6] There are two application categories. In the first category, grade points are calculated using grade point average and any science or advanced placement credits if applicable. In the second category, grade points include any attempts at grade improvements and adds credits for e.g. age, military service, years of study in higher education. Students automatically apply in both categories, but most students are accepted in the first category. In both cases, grade point average is the major determinant of acceptance into higher education.

The major difference between Norway and the United States is that Norway has a much more centralized educational system. There is a national curriculum, in contrast to the United States where states and school districts have more influence. Although some students do attend private schools in Norway, they are highly regulated. In Norway, there is a centralized system for applying to higher education whereas each institution decides their admission criteria in the United States. Due to the centralized system, grading in Norway is monitored by the central government which reduces concerns of measurement error in the analysis.

In the following, grades from lower secondary school are used in the analysis. Educational outcomes are related to whether the student starts academic or vocational track in upper secondary education, grades in upper secondary education, whether the student completes upper secondary education and whether the student continues on to higher education.

[6]There are only some exceptions, such as music and architecture where admissions are determined by an entrance exam as well.

4.2 Norwegian register data

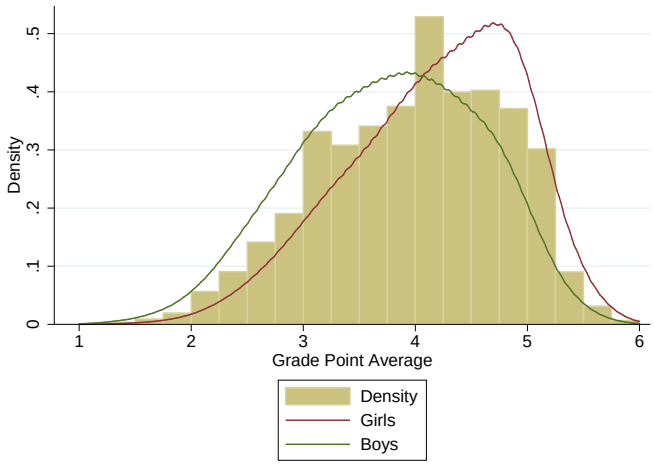
Using register data, provided by Statistics Norway for all individuals leaving lower secondary education during 2002-2004, allows for the combination of detailed information on individual's background and education, including grades, measures of educational attainment and socioeconomic characteristics. The sample is restricted to students with at least 10 valid teacher-assessed grades and only includes students graduating from lower secondary education at age 16.[7] Also, students must have non-missing information on the lower secondary school they attended. The data reduction is presented in Table B1.

Grade point average (GPA) in the NRD is measured as the unweighted mean of all 13 teacher-assessed grades received when leaving lower secondary education. The subjects are written and oral Norwegian, written and oral English, mathematics, natural science, social science, religion, home economics, music and arts, physical education and crafts. Grade variance is measured as the standard deviation of an individual's grades (GSD), using the same grades as were used to calculate the individual's grade point average. Descriptive statistics are presented in panel A of Table 3. In the regressions, both variables are standardized with mean 0 and standard deviation 1 to facilitate interpretation. About 90 % of students in the sample have 13 valid grades. Figure B1 in Section B displays the distribution of grades from one (the lowest) to six (the highest). The most common grade is four (34%), while the least common grade is one (0.86%).

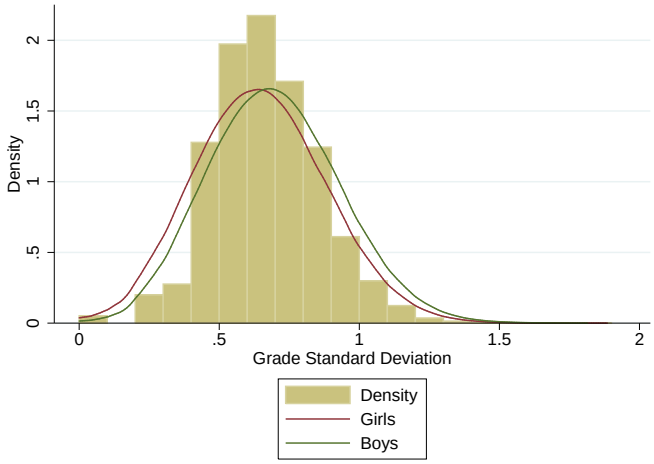
Figures 2a and 2b are equivalent to Figures 1a and 1b of Section 3.2. The distributions are remarkably similar to the NLSY79: The distributions of GPA are skewed to the right while the distributions of GSD are approximately normally distributed. Also, the gender differences are identical to the NLSY79. GPA is higher for girls (4.18) than for boys (3.77) and GSD is higher for boys (0.69) than for girl (0.64). This suggests that the measures of both GPA and GSD are comparable in general and for both genders across countries.

[7] In some cases, students do not start primary education at the expected age, which implies that they finish lower secondary education at a higher age. If a child is not considered to be mature enough, the parents together with the school and psychologists can postpone enrollment one year. In addition, some older students return to improve their grades, and immigrants are often over-aged at graduation.

Figure 2: NRD



(a) Distribution of grade point average



(b) Distribution of grade standard deviation

Note: For grade point average, each bin has a width of 0.25, while each bin has a width of 0.1 for grade standard deviation. Lines display kernel densities with bandwidth 0.15 for each variable for girls (red) and boys (green).

In the NLSY79, the main outcome variable was years of education at age 30. The analysis in the NRD uses cohorts leaving lower secondary school in Norway in 2002-2004, as 2002 is the first year grade data became available. In the last data point available, 2011, these students were still too young to have completed all years of higher education. Therefore, measures for educational attainment in the NRD are short-run measures and are measured as (1) Started academic track (Started ACA), (2) vocational track graduate (VOC grad) (3) academic track graduate (ACA grad) (4) grade point average upper secondary education (GPA UPE) and (5) started higher education (Started HE). Started academic track is an indicator variable equal

to one if the student started one of the three academic study tracks in the first year of upper secondary education. 97% of students go on to upper secondary education in the fall after completing lower secondary education, with 46% starting an academic track and 51% starting vocational tracks. Vocational track graduate and academic track graduate are indicator variables equal to one if the student starts vocational or academic upper secondary education and graduates within five years. Students have a legal right to five years of upper secondary education and this is the standard measure for upper secondary education completion used by the authorities. 70% of students graduate from upper secondary education within five years. Grade point average upper secondary education (GPA USE) is measured as the unweighted mean of all teacher-assessed grades on the upper secondary education transcript, standardized with mean 0 and standard deviation 1. The measure only includes students who complete the academic track and have at least 10 valid grades. Students who transfer from the vocational to the academic track are also included. GPA USE has a mean of 4.15 and a standard deviation of 0.68. The last measure, started higher education, is an indicator variable equal to one if a student has started, but not necessarily completed, a higher education program before 2012. 53% of the sample start higher education. Descriptive statistics are presented in panel B of Table 3.

Socioeconomic characteristics in the NRD are quite similar to the NLSY79. They include gender, birth month, immigration status,[8] parental employment status[9] and parental education.[10] Variables are measured the year the student turns 16. Descriptive statistics are presented in panel C of Table 3. The last columns of Table 3 present descriptive statistics for girls and boys separately. Boys are less likely to start the academic track, have lower GPA and higher GSD in upper secondary education, are less likely to complete upper secondary education and less likely to start higher education.

[8]Immigration status is divided into two categories, where the first indicates that you are a first generation immigrant born abroad with parents born abroad and the second indicates that you are a second-generation immigrant, born in Norway but with both parents born abroad.

[9]Parental employment status is an indicator for whether only the mother, only the father or both parents are working, where no parents working is the reference category.

[10]Parental education as measured as the highest completed education by one of the parents, with categories including having completed upper secondary education, a Bachelor's degree, a Master's degree or PhD and having an unknown education, with less than upper secondary education being the reference category.

Table 3: NRD - Descriptive statistics

	Total		Boys		Girls	
	mean	(sd)	mean	(sd)	mean	(sd)
A. Transcript data						
Grade Point Average (GPA)	3.97	(0.82)	3.77	(0.82)	4.18	(0.77)
Grade Standard Deviation (GSD)	0.67	(0.19)	0.69	(0.19)	0.64	(0.19)
Number of grades	12.87	(0.42)	12.84	(0.47)	12.90	(0.37)
B. Outcome Variables						
Started academic track	0.46	(0.50)	0.42	(0.49)	0.50	(0.50)
Vocational track graduate	0.60	(0.49)	0.56	(0.50)	0.64	(0.48)
Academic track graduate	0.85	(0.36)	0.81	(0.39)	0.88	(0.33)
GPA upper secondary education	4.15	(0.68)	4.06	(0.69)	4.21	(0.67)
Started higher education						
- complete sample	0.53	(0.50)	0.43	(0.49)	0.63	(0.48)
- academic track	0.88	(0.33)	0.88	(0.32)	0.88	(0.33)
C. Socioeconomic characteristics						
Girl	0.49	(0.50)				
Birth month	6.41	(3.36)	6.39	(3.35)	6.44	(3.37)
First generation immigrant	0.034	(0.18)	0.034	(0.18)	0.034	(0.18)
Second generation immigrant	0.020	(0.14)	0.020	(0.14)	0.021	(0.14)
Parental education: Upper secondary	0.47	(0.50)	0.47	(0.50)	0.47	(0.50)
Parental education: Bachelor	0.29	(0.45)	0.29	(0.45)	0.29	(0.45)
Parental education: Master +	0.10	(0.30)	0.10	(0.30)	0.10	(0.30)
Parental education: Unknown	0.042	(0.20)	0.042	(0.20)	0.042	(0.20)
Only mother working	0.13	(0.34)	0.13	(0.33)	0.13	(0.34)
Only father working	0.12	(0.33)	0.13	(0.33)	0.12	(0.33)
Both parents working	0.68	(0.47)	0.68	(0.47)	0.68	(0.47)
D. Cognitive and non-cognitive skills						
Cognitive skills	3.68	(1.06)	3.57	(1.08)	3.81	(1.04)
Non-cognitive skills	4.26	(0.72)	4.09	(0.73)	4.44	(0.67)

Note: N=158,308, with 80,701 boys and 77,607 girls. For Grade point average upper secondary education and grade standard deviation upper secondary, N=84,010 with 33,334 boys and 50,676 girls.

4.3 Empirical strategy and results

For the NRD, the estimated model is equivalent to the one estimated using the NLSY79 data, except that school by cohort fixed effects, $\delta_t \times \theta_s$, are added. y_{ist} is the outcome for student i from school s in year t. GPA_{ist} is grade point average and GSD_{ist} is grade standard deviation from lower secondary education, where each variable is standardized with mean 0 and standard deviation 1. X_{ist} is a vector of socioeconomic characteristics including gender, immigrant status, parental education, parental employment status and birth month. Socioeconomic characteristics

are listed in Table 3. The error term ϵ_{ist} is clustered at the school level. The model can be expressed as

$$y_{ist} = \alpha GPA_{ist} + \gamma GSD_{ist} + X'_{ist}\beta + \delta_t \times \theta_s + \epsilon_{ist} \quad (2)$$

Table 4 reports the results where the outcome is the indicator variable for whether the student has started higher education. The table is equivalent to Table 2 in Section 3.3, with the exception that school by cohort fixed effects are added to the last column. As with the NLSY79, GPA is as expected positively correlated with the educational outcome. Increasing GPA by one standard deviation increases the likelihood that one starts higher education by 30%, which is equivalent to 0.6 of a standard deviation and is similar to the finding for NLSY79.

GSD is added in Column (3) and is negatively correlated with starting higher education. A one standard deviation increase in GSD decreases the likelihood that one starts higher education by 3.2%. This is equivalent to 0.06 of a standard deviation increase in the likelihood of starting higher education. This is approximately half of the GSD estimate found for years of education in the NLSY79. The results remain remarkably stable when adding socioeconomic characteristics (Column (4)) and school by cohort fixed effects (Column (5)), indicating that neither student background nor school characteristics are driving the results.

Table 4: NRD - Started higher education

	(1)	(2)	(3)	(4)	(5)
GPA	0.306***	0.270***	0.292***	0.257***	0.266***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
GSD			-0.032***	-0.030***	-0.027***
			(0.001)	(0.001)	(0.001)
Soc. Char	No	Yes	No	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	No
CohortxSchool FE	No	No	No	No	Yes
R-squared	0.375	0.402	0.380	0.405	0.397
N	158,308	158,308	158,308	158,308	158,308
Number of groups					3,397

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Started higher education is an indicator variable equal to 1 if the student has started higher education before 2012. Standard errors are clustered at the school level.

Table 5 displays results for the outcome variables described in Section 4.2. The last column is equivalent to Column (5) of Table 4, expect that only students graduating from the academic track are included. All estimations include socioeconomic

characteristics and school by cohort fixed effects. Estimates show that GSD is negatively associated with starting the academic track, graduating from the academic track, upper secondary grade point average and starting higher education. The estimate for graduating from upper secondary for students starting the vocational track is small and insignificant. The estimate for GSD in Table 4 seems to be the combined result of students with higher GSD (1) having a higher probability of starting vocational track, where one is less likely to go on to higher education and (2) being less likely to graduate from the academic track and (3) receiving lower grades in the academic track.

Table 5: Main Results - NRD

	Started ACA	VOC grad	ACA grad	GPA USE	Started HE
GPA	0.244***	0.270***	0.210***	1.012***	0.137***
	(0.002)	(0.002)	(0.003)	(0.006)	(0.003)
GSD	-0.018***	-0.0003	-0.013***	-0.012***	-0.004**
	(0.001)	(0.002)	(0.002)	(0.003)	(0.001)
Soc. Char	Yes	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.301	0.242	0.223	0.542	0.111
N	158,308	80,725	72,839	83,740	83,740
Number of groups	3,397	3,306	3,194	3,208	3,208

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Academic is an indicator variable for whether the student goes on to academic track. VOC grad includes all students who start vocational track and complete upper secondary within 5 years. ACA grad includes all students who start academic track and complete upper secondary within 5 years. GPA USE is the GPA from upper secondary education for students who have graduated from the academic track of upper secondary school. This includes students who have transferred from the vocational track during upper secondary school. Started HE is an indicator variable for whether the student has started higher education before 2012 and includes the same sample as GPA USE.

Both the results from Norway and the United States show a negative association between grade variance and educational attainment when controlling for GPA. These findings do not support the hypothesis that being a specialist in compulsory education is beneficial for further education. However, it is still an open question whether the relationship between grade variance and educational attainment depends on the grading distribution, gender and cognitive and non-cognitive skills.

5 Grading distribution, gender and skills

How does the relationship between grade variance and educational attainment depends on the grading distribution, gender and cognitive and non-cognitive skills? In

answering this question, all analyses below are based on the regression in column (4) of Table 2 for the NLSY79 data and column (5) of Table 4 for the NRD.

Grading distribution

If high grade variance individuals are specialists, grade variance is expected to be positively associated with educational attainment, particularly in the upper end of the grade distribution. If high grade variance individuals are individuals with low non-cognitive skills, grade variance is expected to be negatively associated with educational attainment, particularly in the lower end of the grade distribution.

The following investigates whether the direction or strength of the relationship depends on where the student is located in the grading distribution. Regression results reported in Tables 2 and 5 might be masking such differences. To investigate this hypothesis in the NLSY79 data, separate regressions are run for observations above and below the median grade point average, and then separately for each quartile of grade point average. The results are presented in Table 6. The first column shows results for observations below the median grade point average, while the second column shows results for observations above. Both coefficients are negative and significant, but the coefficient is much more negative for the sample above the median. The same pattern emerges when the regression is run for each quartile, however results are no longer significant as the standard errors increase due to fewer observations.

Table 6: NYLS79: Years of education by age 30 - median and quartiles

	Below med.	Above med.	Q1	Q2	Q3	Q4
GPA	0.881***	1.222***	0.937***	1.298***	1.035*	1.017
	(0.042)	(0.161)	(0.128)	(0.232)	(0.379)	(0.455)
GSD	-0.095*	-0.245*	-0.130	-0.042	-0.206	-0.284
	(0.032)	(0.088)	(0.056)	(0.031)	(0.114)	(0.189)
Soc. Char	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.230	0.290	0.202	0.138	0.147	0.259
N	2,200	2,189	1,101	1,099	1,098	1,091

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the cohort level.

Table 7 displays the same regressions for the NRD. Once again, the coefficient is negative across all quartiles, and for the NRD, the coefficient is also strongly significant across all specifications. In the NRD, however, it seems to be that the strongest relationship between GSD and educational attainment is at the middle of

the grading distribution. The coefficient is -0.028 and -0.20 in the middle quartiles, while the coefficient is -0.015 in the lowest quartile and -0.012 in the highest quartile.

Table 7: NRD: Started higher education - median and quartiles

	Below med.	Above med.	Q1	Q2	Q3	Q4
GPA	0.202***	0.178***	0.069***	0.391***	0.315***	0.098***
	(0.003)	(0.004)	(0.003)	(0.010)	(0.012)	(0.004)
GSD	-0.031***	-0.019***	-0.015***	-0.028***	-0.020***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Soc. Char	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.174	0.111	0.046	0.112	0.075	0.048
N	84,085	74,223	41,309	42,776	37,138	37,085

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

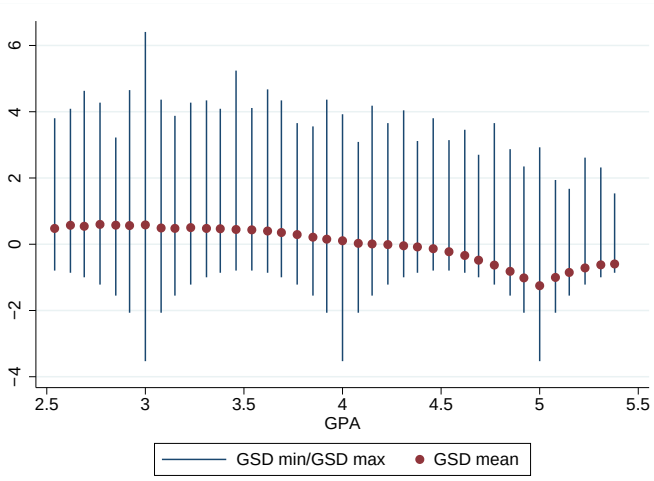
As grades in Norway can only take on integer values from one to six, students are bunched at certain values of GPA. When calculating the grade point average, receiving grades two and four is equivalent to receiving two three's which means that although students have exactly the same GPA, they can have different GSD. This feature not only makes it possible to investigate whether there are heterogeneous results across the grading distribution, it also makes it possible to remove any concern that the coefficient for grade standard deviation is picking up a mechanical correlation between grade standard deviation and grade point average due to for instance ceiling effects.

The analysis is restricted to values where there are at least 1000 students, leaving 38 unique GPA values. Figure 3a displays the mean, minimum and maximum value of GSD for each value of the 38 GPA values. There is a spread in GSD for each value of GPA, which is the variation used to identify how GSD is associated with educational attainment. A separate regression is run at each of these values, and results are reported in Figure 3b. The point estimates are always negative. Confidence intervals show that estimates are lower and significantly different from zero at the middle of the grading distribution, while they are typically not significantly different from zero at the lower and higher end of the grading distribution. This corresponds to the results found in Table 7. The results indicate a negative association between GSD and GPA across the grading distribution, and that this is not solely due to a mechanical correlation between the two variables.

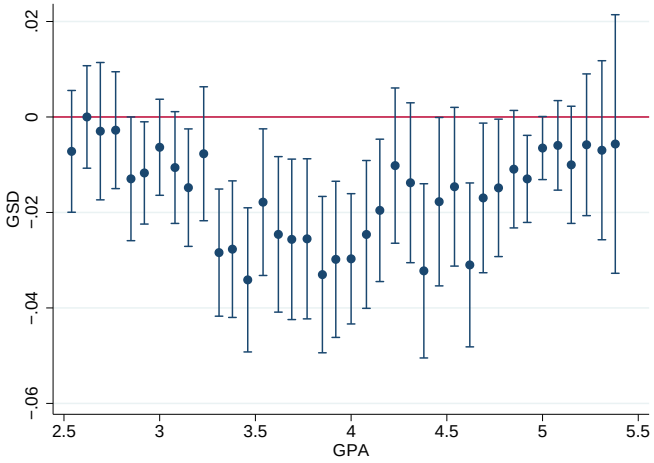
For both the United States and Norway there is no evidence of the direction of the estimates changing across the grading distribution. All point estimates are negative and are significantly lower than zero in most cases. There is also no evidence

Figure 3: NRD: Started higher education - grading distribution

(a) GSD - Descriptive figure



(b) GSD - Regression results



Note: GSD is standardized for the entire sample with mean 0 and standard deviation 1. GPA corresponds to the 38 values of grade point average where there are at least 1000 observations. Figure 3a: Dots indicate the mean value while the bars indicate the minimum and maximum vales of GSD for each regression. Figure 3b: Regressions include socioeconomic characteristics and cohort fixed effects. Dots indicate the coefficient for each regression while the bars indicate the 95% confidence interval.

Table 8: NYLS79 and NRD: Results by gender

	NYLS79		NRD	
	Girls	Boys	Girls	Boys
GPA	0.797***	1.131***	0.278***	0.259***
	(0.040)	(0.035)	(0.002)	(0.002)
GSD	-0.259***	-0.234***	-0.023***	-0.028***
	(0.034)	(0.040)	(0.002)	(0.002)
Soc. Char	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	-	-
CohortxSchool FE	-	-	Yes	Yes
R-squared	0.345	0.453	0.377	0.369
N	2,234	2,155	77,605	80,701
Number of groups	-	-	3,287	3,287

* $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$

Note: Standard errors are clustered at the cohort level for the NYLS79 and school level for the NRD.

that the relationship is stronger at the lower part of the grading distribution. In the non-cognitive skills hypothesis and the specialist/generalist hypothesis, grade variance is thought to be particularly important at the lower and upper end of the grading distribution respectively. There is no support for either in the data.

Gender

Does the relationship between GSD and educational attainment depend on gender? The results could differ by gender if for instance grade variance reflects being a generalist or specialist for one gender, while it reflects high or low non-cognitive skills for the other gender.

For both the NYLS79 and the NRD, the main estimation is run separately for boys and girls. The results are reported in Table 8. The estimates for GSD are not statistically different between genders in either the NYLS79 (columns (1) and (2)) or in the NRD (columns (3) and (4)). The negative association between GSD and educational attainment is the same direction and magnitude for both genders in the United States and Norway.[11]

These estimations show that the main results are not masking differences across boys and girls. Some might believe that high grade variance reflect low non-cognitive skills for boys while it reflects being a specialist for girls. There is no evidence to

[11] Another way to investigate whether results differ by gender is to see how the coefficient for gender in the regressions that includes socioeconomic characteristics changes when GSD is added to the estimation. For both the NYLS79 and the NRD, the coefficient for female stays the same when adding GSD to the regression. The estimate changes from -0.35 to -0.32 in the NYLS79 and from 0.066 to 0.065 in the NRD.

support this theory as the estimates are negative for both genders. Also, there is no evidence that grade variance is more important for one gender as the estimates are not statistically different.

Cognitive and non-cognitive skills

How are results affected by including measures of cognitive and non-cognitive skills to the analysis? If grade point average does not perfectly capture cognitive skills (Roth, Becker, Romeyke, Schäfer, Domnick, and Spinath, 2015) then adding improved measures of cognitive skills might reduce a potential bias in the estimate of GSD. Non-cognitive skills are added to see whether they explain part of the association between grade variance and educational outcomes. If they do, this suggests that grade variance is capturing a measure of non-cognitive skills and supports the non-cognitive skills hypothesis.

The analysis is conducted using all three data sources. In the NLSY79 data, measures of cognitive and non-cognitive skills previously used by Heckman, Stixrud, and Urzua (2006) are added to the analysis. In the NRD, measures of a student's skills in cognitive and non-cognitive subjects, based on a subset of subjects, are added to the analysis. Finally, data from the Development in Adolescence Project (CDAP) are used to investigate how non-cognitive skills relate to GSD when conditioning on GPA.

The measure for cognitive skills in the NLSY79 is a composite score of five measures from the Armed Services Vocational Aptitude Battery (ASVAB),[12] which includes scores for arithmetic reasoning, word knowledge, paragraph comprehension, mathematical knowledge and coding speed. Descriptive statistics are reported in panel D of Table 1. For each measure, the scores are standardized with mean 0 and standard deviation 1, and the sum of these five scores is then again standardized with mean 0 and standard deviation 1.

The measure for non-cognitive skills in the NLSY79 is a combination of the Rotter Locus of Control Scale (Rotter, 1966), and the Rosenberg Self-Esteem Scale (Rosenberg, 1965). The Rotter Locus of Control Scale is designed to measure the extent to which individuals believe they have control over their lives through self-motivation or self-determination (internal control) as opposed to the extent that the environment (chance, fate, luck) controls their lives (see Table A2). The Rosenberg Self-Esteem Scale describes ones degree of approval or disapproval toward oneself.

[12]The Armed Services Vocational Aptitude Battery (ASVAB) is a battery of tests administered to applicants to the United States military to determine their qualifications and job assignment. The Armed Forces Qualifying Test AFQT is comprised of test results from the batteries Arithmetic Reasoning, Math Knowledge, Word Knowledge and Paragraph Comprehension (ASVAB, 2015).

Table 9: NYLS79: Conditional correlations

	(1) GSD	(2) GSD	(3) GSD	(4) GSD
GPA	-0.521***	-0.523***	-0.464***	-0.462***
	(0.014)	(0.014)	(0.017)	(0.017)
Non-cognitive		-0.009		0.011
		(0.014)		(0.014)
Cognitive			-0.127***	-0.136***
			(0.019)	(0.020)
Soc. Char	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R-squared	0.302	0.309	0.315	0.318
N	4,389	4,226	4,243	4,136

Note: Standard errors are clustered at the cohort level.

(see Table A3). Descriptive statistics are reported in panel E of Table 1 above. Both scores are standardized with mean 0 and standard deviation 1, and the sum of these two scores is then again standardized with mean 0 and standard deviation 1.

Table 9 displays the conditional correlation between GPA and GSD when including cognitive and non-cognitive measures to the NLSY79 data. Column (1) is the conditional correlation between GPA and GSD when including school fixed effects and socioeconomic characteristics. Column (2) adds the measure of non-cognitive skills, column (3) adds the measure of cognitive skills and column (4) adds both measures. We see that the measure for non-cognitive skills is not significant while the measure for cognitive skills is negatively associated with GSD, conditional on GPA. Importantly, adding non-cognitive skills does not change the conditional correlation between GPA and GSD.

In Table 10, cognitive and non-cognitive measures are added to the main analysis. Descriptive statistics are presented in panel D of Table 3. The estimate for non-cognitive skills, as shown in column (2) is significant and positive, as expected, with a one standard deviation increase in non-cognitive skills predicting an increase in educational attainment by 0.26 of a year. However, the estimates for GPA and GSD are unchanged, suggesting that the measure of non-cognitive skills does not explain why GSD is negatively associated with educational attainment. The measure for cognitive skills, as shown in column (3), is significantly and positively associated with educational attainment and reduces both the estimate for GPA and GSD. A one standard deviation increase in cognitive skills predicts an increase in educational attainment by 0.8 of a year. Column (4) includes both measures, with estimates for GPA and GSD remaining stable from column (3) to column (4). The results

Table 10: NYLS79: Years of education age 30

	(1)	(2)	(3)	(4)
Grade Point Average	0.972***	0.913***	0.604***	0.598***
	(0.028)	(0.033)	(0.030)	(0.031)
Grade Standard Deviation	-0.238***	-0.250***	-0.176**	-0.182**
	(0.034)	(0.036)	(0.036)	(0.034)
Non-cognitive		0.263***		0.143**
		(0.047)		(0.039)
Cognitive			0.800***	0.752***
			(0.044)	(0.041)
Socioeconomic Characteristics	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R-squared	0.399	0.415	0.461	0.463
N	4,389	4,226	4,243	4,136

* $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$

Note: Standard errors are clustered at the cohort level.

correspond to those found for the conditional correlations. The main inference from these estimates is that there is no evidence that the relationship between GPA and GSD or the relationship between GSD and educational attainment can be explained by non-cognitive skills in the NLSY79 data. Also, the estimate for GSD remains negative and statistically significant in all specifications.

In the Norwegian data, measures of a student's skills in cognitive and non-cognitive subjects are added to the analysis. Falch, Nyhus, and Strøm (2014), using the same grade data from Norway as this paper, use the average grade in math and science as a proxy for cognitive skills and the average grade in physical education, food and health, arts and crafts and music as a proxy for non-cognitive skills. These same measures are standardized with mean 0 and standard deviation 1 and added to the analysis to investigate how cognitive and non-cognitive skills relate to GSD in the Norwegian data. Note that these measures are sub-samples of the grades used to calculate GPA and GSD. They are imperfect measures that do not add any new information, but rather take out some of the variation. This makes the results hard to interpret.

Table 11, comparable to Table 9, displays the conditional correlation between GPA and GSD when including these cognitive and non-cognitive measures. The non-cognitive measure is positively associated with GSD while the cognitive measure is negatively associated with GSD. For a given GPA, students with good grades in non-cognitive subjects have higher GSD, while students with good grades in cognitive subjects have lower GSD. The conditional correlation between GSD and GPA is greatly affected by the inclusion of measures of non-cognitive and cognitive skills.

Table 11: NRD: Conditional correlations

	GSD	GSD	GSD	GSD
GPA	-0.431***	-1.010***	0.054***	-0.586***
	(0.005)	(0.010)	(0.009)	(0.015)
Non-cognitive		0.659***		0.586***
		(0.010)		(0.010)
Cognitive			-0.524***	-0.389***
			(0.009)	(0.009)
Socioeconomic Characteristics	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.212	0.305	0.260	0.331
N	158,308	158,308	158,289	158,289
Number of groups	3,397	3,397	3,397	3,397

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

This is not surprising as these variables are subsets of grades used to calculate GSD and GPA. However, it is interesting to note that the different subsets do, in fact, seem to measure something different, but whether this is cognitive and non-cognitive skills is harder to determine.

As the conditional correlations are differentially affected by including measures of cognitive and non-cognitive skills, it is reasonable to assume that this will also be the case when including these measures to the estimations in Table 5. Tables B2 - B6 in Appendix B report the results and this is indeed the case. However, the results are hard to interpret as the measures of cognitive and non-cognitive skills are so closely related to GPA and GSD.

The results from the NLSY79 data show that, if anything, grade variance is associated with cognitive skills rather than non-cognitive skills, while the results from the NRD show no clear pattern. However, both measures of non-cognitive skills are quite simple and do not necessarily include the non-cognitive skills one would associate with low grade variance. To explore this further, data from the Character Development in Adolescence Project (CDAP), provided by Angela Duckworth, are used to investigate the non-cognitive skills in greater detail. The data include grades and a rich set of non-cognitive skills allowing me to investigate how non-cognitive skills relate to GSD when conditioning on GPA (see Section C1 for a description of the data). Non-cognitive skills are either self-reported by the student or reported by the student's teachers. The self-reported measure (Non-cognitive: SR) is a joint measure for the non-cognitive skills (1) delay discounting, (2) grit, (3) self-control: work, (4) self-control: interpersonal, (5) gratitude, (6) actively open-minded think-

Table 12: CDAP: Conditional correlations

	(1) GSD	(2) GSD	(3) GSD	(4) GSD
GPA	-0.343***	-0.350***	-0.340**	-0.365*
	(0.040)	(0.050)	(0.081)	(0.092)
Non-cognitive: SR		-0.012		-0.013
		(0.016)		(0.020)
Non-cognitive: TR			0.002	0.034
			(0.068)	(0.074)
Soc. Char	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Observations	1293	1021	1268	1015

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: SR denotes self-reported. TR denotes teacher-reported. Standard errors are clustered at the school level.

ing, (7) prosocial purpose and (8) internal locus of control. The teacher-reported measure (Non-cognitive: TR) is a joint measure for the non-cognitive skills (1) grit, (2) self-control: work, (3) self-control: interpersonal, (4) gratitude, (5) actively open-minded thinking and (6) prosocial purpose. The results are displayed in Table 12. Column (1) displays the conditional correlation between GPA and GSD which is negative and significant. Column (2) adds the self-reported non-cognitive measure, column (3) adds the teacher-reported non-cognitive measure and column (4) adds both measures. The estimate for GSD remains stable and the measures for non-cognitive skills are not statistically significant for all specifications. The results hold when regressions are run for each student and teacher reported non-cognitive skill separately (not reported here). Once again, it does not seem that grade variance is associated with non-cognitive skills. The main inference from these estimates is that the association between grade variance and grade point average cannot be explained by non-cognitive skills.

In all three data sets, non-cognitive skills do not change the size or direction of the GSD estimate in the conditional correlation tables. There is no evidence that the association between grade variance and educational attainment can be explained by non-cognitive skills. As a result, even though the estimate between GSD and educational attainment is negative, there is no support of the non-cognitive skills hypothesis.

6 Conclusion

Throughout all explorations of the importance of the second moment of individual grade distribution, I find that individual grade variance is negatively associated with educational attainment. For both the United States and Norway, this association holds across the grade distribution and for both genders and estimates are robust to controlling for socioeconomic characteristics and school fixed effects. In addition, estimates remain negative when including measures of cognitive and non-cognitive skills. My results suggest that the negative association between grade variance and educational attainment is a general finding that is not country or context specific.

The cognitive-skill hypothesis is that high grade variance is associated with low educational attainment because it reflects low non-cognitive skills. This hypothesis is supported by the main results. However, the grade standard deviation estimate is larger in the upper end of the grading distribution for the United States and in the middle of the grading distribution for Norway, which does not support Lindqvist and Vestman (2011) who find that non-cognitive skills are more important in the lower end of the grading distribution. More importantly, using three different data sets, it is not possible to find a systematic relationship between non-cognitive skills and grade variance.

The other hypothesis is that high grade variance reflects being a specialist rather than a generalist, and that this is positively associated with educational attainment. However, the main results rather support the reversed generalist/specialist hypothesis, that it is beneficial to be a generalist. Why could it be beneficial to be a generalist? Lazear (2004) suggests that it might be beneficial to have a span of skills for certain studies or occupations. This might be the case also for higher education, which is often based on general knowledge, particularly in the United States where there is a long tradition for a liberal arts education in four-year colleges. Another possible explanation is that being a generalist increases your adaptability which could be beneficial for long-run outcomes (Hanushek, Woessmann, and Zhang, 2011). Testing these hypotheses is a topic for future research.

If institutions are interested in students with high ability and effort, but only use grade point average in the admission decision, they may not be accepting the best students. Students with low grade variance who are just below the grade point average cutoff are likely to outperform student just above the cutoff with high grade variance. My findings support that institutions should take grade variance, or other measures of skill, into account in admission decisions.

References

ACT (2015): "What is the ACT?," http://www.actstudent.org/faq/what.html, Reading date: 07.08.2015.

ASVAB (2015): "ASVAB Fact Sheet," http://official-asvab.com/docs/asvab_fact_sheet.pdf, Reading date: 07.08.2015.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008): "The economics and psychology of personality traits," *Journal of Human Resources*, 43(4), 972–1059.

Carneiro, P., C. Crawford, and A. Goodman (2007): "The impact of early cognitive and non-cognitive skills on later outcomes," Discussion paper, CEE DP 92.

Clinedinst, M. E., and D. A. Hawkins (2011): "State of college admission," *Washington, DC: National Association for College Admission Counseling*.

Falch, T., O. H. Nyhus, and B. Strøm (2014): "Performance of Young Adults: The Importance of Different Skills," *CESifo Economic Studies*.

Hanushek, E. A., and D. D. Kimko (2000): "Schooling, labor-force quality, and the growth of nations," *American Economic Review*, pp. 1184–1208.

Hanushek, E. A., and L. Woessmann (2008): "The role of cognitive skills in economic development," *Journal of Economic Literature*, pp. 607–668.

Hanushek, E. A., L. Woessmann, and L. Zhang (2011): "General education, vocational education, and labor-market outcomes over the life-cycle," Discussion paper, National Bureau of Economic Research.

Haraldsvik, M. (2003): "Inntaksprosedyrer for den videregående skole: Grad av valgfrihet," Institutt for samfunnsøkonomi, NTNU.

Heckman, J. J. (1995): "Lessons from the bell curve," *Journal of Political Economy*, pp. 1091–1120.

Heckman, J. J., and T. Kautz (2013): "Fostering and measuring skills: Interventions that improve character and cognition," Discussion paper, National Bureau of Economic Research.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Hedges, L. V., and A. Nowell (1995): "Sex differences in mental test scores, variability, and numbers of high-scoring individuals," *Science*, 269(5220), 41–45.

Herrnstein, R. J., and C. Murray (2010): *Bell curve: Intelligence and class structure in American life*. Simon and Schuster.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014): "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success," OECD Education Working Papers 110, OECD Publishing.

Lazear, E. P. (2004): "Balanced skills and entrepreneurship," *American Economic Review*, pp. 208–211.

Lindqvist, E., and R. Vestman (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3(1), 101–28.

Murnane, R. J., J. B. Willett, and F. Levy (1995): "The growing importance of cognitive skills in wage determination," *Review of Economics and Statistics*, 77(2), 251–266.

National Center for Education Statistics (2015): "Private School Enrollment," `http://nces.ed.gov/programs/coe/indicator_cgc.asp`, Reading date: 10.07.2015.

National Center for Research in Vocational Education and The Center for Human Resource Research, The Ohio State University (1984): *NLSY79 High School Transcript survey: Overview and Documentation*.

OECD (2015): "About PISA," `http://www.oecd.org/pisa/aboutpisa/`, Reading date: 07.08.2015.

Rosenberg, M. (1965): *Society and the adolescent self-image*. Princeton University Press Princeton, NJ.

Roth, B., N. Becker, S. Romeyke, S. Schäfer, F. Domnick, and F. M. Spinath (2015): "Intelligence and school grades: A meta-analysis," *Intelligence*, 53, 118 – 137.

Rotter, J. B. (1966): "Generalized expectancies for internal versus external control of reinforcement.," *Psychological monographs: General and applied*, 80(1), 1.

SAT (2015): "About the SAT," https://sat.collegeboard.org/about-tests/sat, Reading date: 07.08.2015.

Segal, C. (2012): "Working when no one is watching: Motivation, test scores, and economic success," *Management Science*, 58(8), 1438–1457.

Stevenson, H. W., and R. Nerison-Low (2002): "To Sum It Up: Case Studies of Education in Germany, Japan, and the United States.," .

The Norwegian Directorate for Education and Training (2015): "Grunnskolens informasjonssystem," https://gsi.udir.no/application/main.jsp?languageId=1, Reading date: 29.12.2015.

U.S. Bureau of Labor Statistics (2015): "NLSY79 Appendix 21: Attitudinal Scales," https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-appendix-21-attitudinal-scales, Reading date: 10.07.2015.

A National Longitudinal Survey of Youth, 1979

A.1 Data description

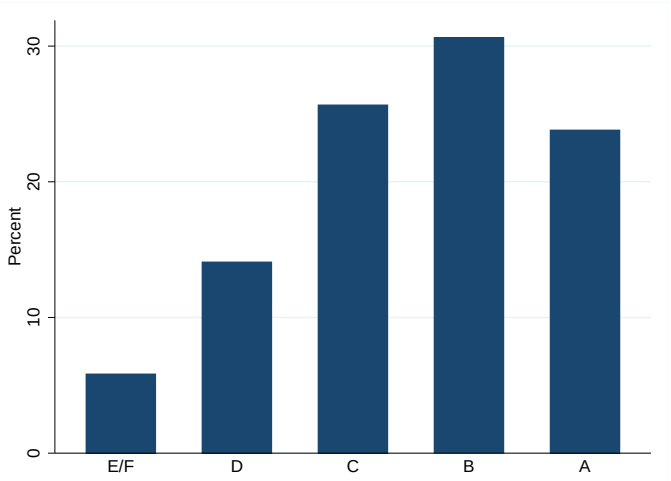
Between 1980 and 1983, transcript information was collected for respondents who were 17 years of age or older and expected to complete high school in the United States. The data include up to 64 courses. Of the 6,111 respondents, 5,009 have non-missing transcript data (see Table A4). Information for each course on the transcript includes (1) grade level for which the course was taken (2) a code for the high school course (3) the final or computed grade for that course (4) the source for the final grade and (5) the credits received. Courses are divided into 22 subject areas, listed in Table A1. For a complete list of course codes, see (National Center for Research in Vocational Education and The Center for Human Resource Research, The Ohio State University, 1984). Students either receive a pass/fail grade, or grades A-F, where grade A is coded to value 4, grade B to value 3, grade C to value 2, grade D to value 1. Grade F is a failing grade and is coded to value 0. Figure A1 shows the distribution of grades for the 214,507 grades in the sample. The analysis is restricted to students with 10 or more valid grades. Figure A2 shows the distribution of number of grades in the sample. The data reduction is presented in Table B1.

Table A1: NLSY79 - Course subject area in transcript data

	N	Percent
Agriculture	1718	0.79
Art	7405	3.40
Business	3058	1.40
Distributive education	1038	0.48
English	43119	19.80
Foreign Language	7830	3.59
Health occupations education	294	0.13
Health and physical education	25129	11.54
Home economics	9707	4.46
Industrial arts	7390	3.39
Mathematics	23496	10.79
Music	6517	2.99
Natural sciences	19926	9.15
Office occupations education	11287	5.18
Social studies	34354	15.77
Technical education	62	0.03
Vocational	2971	1.36
Safety and driver education	3827	1.76
Junior ROTC	450	0.21
Philosophy and religion	1500	0.69
Study skills	731	0.34
Career education	4120	1.89
Missing	1875	0.86
Total	217804	100.00

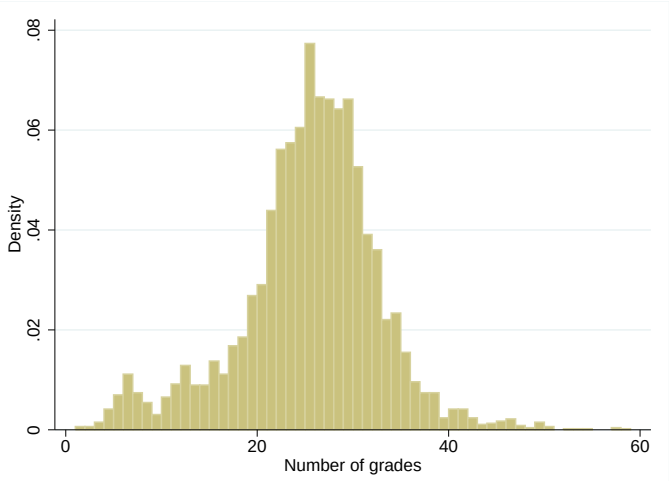
Note: Missing denotes missing course code but non-missing course grade. See National Center for Research in Vocational Education and The Center for Human Resource Research, The Ohio State University (1984) for a detailed list of the course codes.

Figure A1: NLSY79 Grade distribution



Note: The figure includes 113,213 grades ranging from E/F (non-pass, lowest) to A (highest) for 4389 students from the NLSY79 survey. The sample includes students from the main sample with 10 or more valid grades and with non-missing educational attainment at age 30.

Figure A2: NLSY79 - Number of grades, grades 9-12



Note: The figure includes 4,577 students from the NLSY79 survey who have grades reported transcript data, are from the main sample and have non-missing educational attainment at age 30. 205 students have less than 10 grades, and are dropped in the analysis. The final sample is thus 4,389 students (see Table A4).

A.2 Cognitive and non-cognitive skills

Table A2: The NLSY79 Rotter – Locus of control questions

1a	What happens to me is my own doing.
1b	Sometimes I feel that I don't have enough control over the direction my life is taking.
2a	When I make plans, I am almost certain that I can make them work.
2b	When I make plans, it is not always wise to plan too far ahead, because many things turn out to be a matter of good or bad fortune anyhow.
3a	Getting what I want has little or nothing to do with luck.
3b	Many times we might just as well decide what to do by flipping a coin
4a	Many times I feel that I have little influence over the things that happen to me.
4b	It is impossible for me to believe that chance or luck plays an important role in my life.

Note: The Rotter Locus of Control Scale is a four item forced choice questionnaire and is an abbreviated version of the 60-item Rotter scale. Scores are generated for each pair of items. Internal control: Much closer=1 Slightly closer =2 External control: Much closer=3 Slightly closer=4. Scores of 4 pairs were summed. Total score could range from 4 to 16 points. If one item is missing, the scale score is coded as missing(U.S. Bureau of Labor Statistics, 2015). In this paper, scores are reversed such that a higher score is more internal control, and thus reflects higher non-cognitive skills (values from 0 to 12). The test was administered in the NLSY79 in 1979.

Table A3: The NLSY79 Rosenberg Self-Esteem Scale questions

1	I am a person of worth.
2	I have a number of good qualities.
3	I am inclined to feel that I am a failure.
4	I am able to do things as well as most other people.
5	I felt I do not have much to be proud of.
6	I take a positive attitude toward myself.
7	I am satisfied with myself.
8	I wish I could have more respect for myself.
9	I certainly feel useless at times.
10	At times I think I am no good at all.

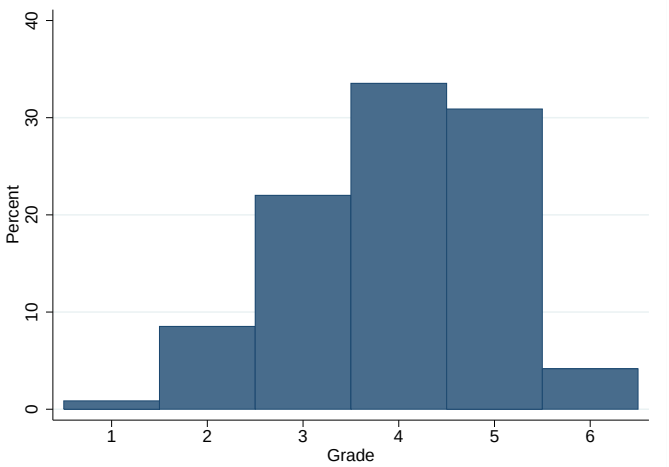
Note: The scale contains 10 statements about self-approval and disapproval to which the respondents are asked to strongly agree, agree, disagree or strongly disagree. Higher scores are associated with higher self-esteem. Scoring for items 3, 5, 8, 9, 10: strongly agree=0 agree=1 disagree=2 strongly disagree=3. Scoring for items 1, 2, 4, 6, 7 is reversed so that a higher score indicates higher self-esteem. Scores of 10 items were summed. Total score could range from 0 to 30 points. If one item is missing, the scale score is coded as missing(U.S. Bureau of Labor Statistics, 2015). The test was administered in the NLSY79 in 1979.

Table A4: Data Reduction NLSY79

	N	Reduction	% Reduction
1. Complete sample	12686		
2. Main sample	6111	6575	51.83 %
3. Non-missing transcript data	5009	1102	18.03 %
4. Non-missing educational outcome	4577	432	8.62 %
5. 10 or more valid grades	4389	188	4.11 %
6. Non-missing cognitive skills	4243	146	3.33 %
6. Non-missing non-cognitive skills	4226	163	3.71 %
6. Non-missing cognitive and non-cognitive skills	4136	253	5.76 %

B Norwegian register data

Figure B1: NRD



Note: 2,037,789 grades ranging from 1 (lowest) to 6 (highest) for 158,308 students leaving lower secondary education 2002-2004. About 90% of students have 13 valid grades.

Table B1: Data Reduction NRD

	N	Reduction	% Reduction
1. Sample 2002-2004	168,151		
2. 10 or more valid grades	162,831	5,320	3.16 %
3. 16 years old	159,077	3,754	2.31 %
4. Non-missing school information	158,308	769	0.48 %

Note: Restriction number 3 is that the student has to be 16 years old when graduating from lower secondary education.

Table B2: NRD: Academic track - cognitive and non-cognitive skills

	(1) ACA	(2) ACA	(3) ACA	(4) ACA
GPA	0.244***	0.233***	0.291***	0.286***
	(0.002)	(0.003)	(0.003)	(0.005)
GSD	-0.018***	-0.017***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Cognitive		0.012***		0.005
		(0.003)		(0.003)
Non-cognitive			-0.049***	-0.048***
			(0.003)	(0.003)
Soc. Char	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.301	0.301	0.303	0.302
N	158,308	158,289	158,308	158,289
Number of groups	3,397	3,397	3,397	3,397

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

Table B3: Vocational graduate - cognitive and non-cognitive skills

	(1) VOC graduate	(2) VOC graduate	(3) VOC graduate	(4) VOC graduate
GPA	0.270***	0.205***	0.179***	0.096***
	(0.002)	(0.004)	(0.004)	(0.006)
GSD	-0.000	0.009***	-0.023***	-0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
Cognitive		0.076***		0.087***
		(0.004)		(0.004)
Non-cognitive			0.091***	0.099***
			(0.004)	(0.004)
Soc. Char	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.242	0.246	0.248	0.253
N	80,725	80,710	80,725	80,710
Number of groups	3,306	3,306	3,306	3,306

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

Table B4: Academic graduate - cognitive and non-cognitive skills

	(1) ACA graduate	(2) ACA graduate	(3) ACA graduate	(4) ACA graduate
GPA	0.210***	0.174***	0.178***	0.130***
	(0.003)	(0.004)	(0.004)	(0.005)
GSD	-0.013***	-0.009***	-0.016***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)
Cognitive		0.038***		0.046***
		(0.004)		(0.004)
Non-cognitive			0.036***	0.043***
			(0.003)	(0.003)
Soc. Char	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.223	0.225	0.225	0.228
N	72,839	72,838	72,839	72,838
Number of groups	3,194	3,194	3,194	3,194

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

Table B5: Upper secondary education GPA - cognitive and non-cognitive skills

	(1) GPA USE	(2) GPA USE	(3) GPA USE	(4) GPA USE
GPA	1.012***	0.756***	1.152***	0.881***
	(0.006)	(0.008)	(0.008)	(0.011)
GSD	-0.012***	0.016***	0.004	0.026***
	(0.003)	(0.003)	(0.003)	(0.003)
Cognitive		0.280***		0.261***
		(0.007)		(0.007)
Non-cognitive			-0.162***	-0.125***
			(0.007)	(0.007)
Soc. Char	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.542	0.554	0.546	0.557
N	83,740	83,737	83,740	83,737
Number of groups	3,208	3,208	3,208	3,208

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

Table B6: Started higher education- cognitive and non-cognitive skills

	(1) Started HE	(2) Started HE	(3) Started HE	(4) Started HE
GPA	0.137***	0.118***	0.162***	0.145***
	(0.003)	(0.004)	(0.004)	(0.005)
GSD	-0.004**	-0.002	-0.002	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Cognitive		0.020***		0.016***
		(0.003)		(0.003)
Non-cognitive			-0.029***	-0.027***
			(0.003)	(0.003)
Soc. Char	Yes	Yes	Yes	Yes
CohortxSchool FE	Yes	Yes	Yes	Yes
R-squared	0.111	0.112	0.113	0.113
N	83,740	83,737	83,740	83,737
Number of groups	3,208	3,208	3,208	3,208

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are clustered at the school level.

C Data from the Development in Adolescence Project

The Development in Adolescence Project (CDAP) is a longitudinal survey of 1559 middle school students and their teachers from 8 different schools. The same students receive a survey in four rounds, the fall and spring of eighth grade and the fall and spring of ninth grade. Their teachers in math, science, English and social studies also receive a survey in each round. The data also include grades from math, science, English and social studies for each semester. I use data from rounds 1 and 2. Only students with one or no missing grades are included in the analysis. Two schools are dropped from the analysis, one due to missing grade data and another due to different grading practices. This leaves a sample of 1293 students.

Grade point average (GPA) is calculated as the average of all grades received during the two rounds. Grade standard deviation (GSD), used as a measure of grade variance, is calculated as the standard deviation of the same grades used to calculate grade point average. GPA and GSD are then standardized for the whole sample. Socioeconomic characteristics include gender, ethnicity (dummy variables for Hispanic, Asian, African American, multiethnic or other) birth date, being an English language learner, receiving reduced/free lunch and receiving special education. Rather than exclude students with missing values on control variables, dummy variables for missing are constructed and included in the regressions. Descriptive

statistics for GPA, GSD and socioeconomic characteristics are listed in Table C1.

Students' self-reported non-cognitive skills in each round include, among other things, (1) delay discounting, (2) grit, (3) self-control: work, (4) self-control: interpersonal, (5) gratitude, (6) actively open-minded thinking, (7) prosocial purpose and (8) internal locus of control. To create a joint measure of students' non-cognitive skills, each measure is standardized with mean 0 and standard deviation 1 before standardizing the sum of these measures with mean 0 and standard deviation 1. There are 272 students with missing information on one or more measures, reducing the sample to 1021. Teacher-reported non-cognitive skills for individual students in each round include (1) grit, (2) self-control: work, (3) self-control: interpersonal, (4) gratitude, (5) actively open-minded thinking and (6) prosocial purpose.

Teacher self-reported measures are averages across all teachers for each student. To create a joint measure of teacher-reported non-cognitive skills, each measure is standardized with mean 0 and standard deviation 1 before standardizing the sum of these measures with mean 0 and standard deviation 1. There are 25 students with missing information on one or more teacher-reported measures, reducing the sample to 1268. Descriptive statistics for student self-reported and teacher-reported non-cognitive skills are listed in Table C1.

Table C1: Development in Adolescence Project - Descriptive statistics

	Total		Boy		Girl	
	mean	(sd)	mean	(sd)	mean	(sd)
Girl	0.49	(0.50)	0	(0)	1	(0)
Hispanic	0.16	(0.37)	0.16	(0.37)	0.17	(0.37)
Asian	0.11	(0.32)	0.12	(0.33)	0.11	(0.31)
Multiethnic or other	0.0085	(0.092)	0.012	(0.11)	0.0047	(0.069)
African American	0.48	(0.50)	0.47	(0.50)	0.49	(0.50)
Birth month	6.68	(3.48)	6.70	(3.48)	6.66	(3.48)
English language learner	0.14	(0.35)	0.14	(0.35)	0.14	(0.35)
Special education	0.16	(0.36)	0.20	(0.40)	0.11	(0.32)
Free/reduced lunch	0.66	(0.47)	0.64	(0.48)	0.68	(0.47)
Non-cognitive: self-reported	0	(1.00)	-0.022	(1.01)	0.022	(0.99)
Non-cognitive: teacher reported	0	(1.00)	-0.22	(1.02)	0.23	(0.93)

Note: N=1293, with 659 boys and 634 girls. For Non-cognitive: self-reported, N=1021, with 514 boys and 507 girls. For Non-cognitive: teacher reported, N=1268, with 650 boys and 618 girls.