Grading Bias and Young Adult Mental Health

Linder, Anna; Nordin, Martin; Heckley, Gawain

2022

Document Version:
Other version

Link to publication

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2022:7

Department of Economics
School of Economics and Management

Grading Bias and Young Adult Mental Health

Anna Linder
Martin Nordin
Ulf-G. Gerdtham
Gawain Heckley

March 2022

LUND UNIVERSITY

Grading Bias and Young Adult Mental Health

Anna Linder[a,b,*], Martin Nordin[b,c], Ulf-G Gerdtham[a,b,c] & Gawain Heckley[a,b]

[a] Health Economics Unit, Department of Clinical Science, Lund University, Malmö, Sweden
[b] Centre for Economic Demography, Lund University, Lund, Sweden
[c] Department of Economics, Lund University, Lund, Sweden

Abstract

Various grading reforms and trends of more lenient grading have contributed to grade inflation in Sweden and other countries. Previous research shows that over-grading increases higher education enrolment, achievements and earnings, but no study has yet addressed the potential impact of grading bias on health. In this paper, we hypothesize that over-grading has a protective impact on mental health, either through a direct effect of performance feedback, or through mechanisms such as self-efficacy and university admission distortions. We test this hypothesis using Swedish individual-level register data for individuals graduating from upper secondary school in the years 2001-2004. Grading bias, which we interpret as over-grading, is constructed as the residual of final upper secondary school grades having controlled for results in a standardised test, itself not subject to grading leniency. Over-grading is further isolated by considering only within-school variation in over-grading and controlling for prior grades and school production. We show that over-grading has substantial significant protective impacts on the mental health of young adults, but only among female students. That grades themselves, independent of knowledge, substantially impact the production of health highlights an important health production mechanism, and also implies that any changes to the design of grading systems must consider these wider health implications.

JEL codes: I21, I28, I10

Key words: Grading bias, grade inflation, mental health, human capital development

---

Declarations of interest: Financial support from the Swedish Research Council (DNR 2019-06292) and Tom Hedelius and Jan Wallander foundation (P2017-0081:1) (Heckley) is gratefully acknowledged. The Health Economics Program at Lund University also receives core funding from Government Grant for Clinical Research, and Region Skåne (Gerdtham).

We thank Anton Nilsson, Lund University, for helpful comments and suggestions.

*Corresponding author: Anna Linder, Health Economics Unit, Lund University, Medicon Village 301:5, Nytänkargatan 4, 223 63 Lund, Sweden. Tel: +46 72 467 48 80. E-mail: anna.linder@med.lu.se.

1. Introduction

How we grade students is hugely important because grades are used as key indicators of student performance and function as a sorting mechanism within the education system and the labour market. Systemic grading bias, over- or under-grading not reflecting the true level of knowledge, could create unfairness between students. For example, if grades are inflated for the benefit of students in grading lenient schools, this will be at the expense of other, higher-achieving students not exposed to over-grading. Grading bias may also lead to welfare costs due to inefficient allocations of higher education investments. Various grading reforms and trends of more lenient grading have contributed to grade inflation in many countries, for example in the U.S (Rojstaczer & Healy, 2012), Germany (Müller-Benedict & Gaens, 2020) and Sweden (Vlachos, 2011; Wikström & Wikström, 2005). As a result, consequences of systematic grading biases have received growing attention in the literature.

Earlier studies have found that over-grading increases higher education enrolment and achievements (Dee, Dobbie, Jacob, & Rockoff, 2019; Diamond & Persson, 2016; Maurin & McNally, 2008; Nordin, Heckley, & Gerdtham, 2019). Over-grading also impacts labour income, an effect that has been explained by easier access to later stages of the education system (Maurin & McNally, 2008; Nordin et al., 2019). Human capital development theory suggests that grades could impact both the production of skills and the production of health (Cunha & Heckman, 2007; Heckman, 2012). This means that health could be an outcome of grading bias that is important in itself, but health could also be a mechanism in the relationship between grading bias and labour market outcomes. Yet, to the best of our knowledge, there is no existing evidence on the potential health consequences of grading bias. Therefore, in this study, we investigate the impact of grading bias on health. Specifically, we investigate the impact on mental ill-health among young adults, which is one of the largest public health challenges in modern times (United Nations Children's Fund, 2021). We hypothesize that over-grading impacts the production of mental health through 1) a direct effect of performance feedback, 2) an impact on self-efficacy beliefs by which positive feedback/re-ranking increases self-confidence, motivation and performance thereby positively affecting mental health, or 3) through the admission distortion which increases opportunities for higher education investments, of which both attendance and years completed have been found to causally affect health (Heckley et al., 2022; Buckles et al., 2016).

Several reforms related to deregulation and decentralisation of the Swedish school system were implemented at the beginning of the 1990s including a major reform of the grading system. The reforms introduced competition in the Swedish school market which, in combination with a move from a relative to a goal- and criterion-referenced grading system, contributed to substantial grade inflation during the following decade (Vlachos, 2011; Wikström & Wikström, 2005). Our data includes all students who graduated from upper secondary schools in Sweden between the years 2001 – 2004, that is, students exposed to grade inflation following the 1990s school reforms. For these cohorts, one in eight students had a psychiatric-related diagnosis or prescription in the years after graduation. To identify grading bias at the school level we exploit the difference between upper secondary Grade Point Average (GPA) and standardised Swedish Scholastic Aptitude Test (SweSAT) results, a test not subject to grading leniency. On an individual level, the difference between grades and SweSAT results is a residual skill measure, reflecting grading error and also variation in non-cognitive skills. However, on the school level, the difference between upper secondary school grades and SweSAT results capture systemic changes in grades that can be used to identify grading bias, at least if we control for non-cognitive skills.

Within-school changes in grading bias, controlling for prior cohort knowledge, background characteristics and changes in school production, finds that being exposed to over-grading has a significant, protective, impact on mental health, but only among women. The changes in grading bias are assumed to reflect increased levels of over-grading since they are measured over a period when grade inflation was observed. A quarter of a standard deviation increase in over-grading, exhibited by more than a third of schools in our sample, is associated with an eight percent lower probability of being diagnosed for internalising disorders such as depression and anxiety, and a four percent lower probability of being prescribed mental ill-health related drugs such as antidepressants and anxiolytics. The effect appears a couple of years after graduation, which suggests that it is more likely to be a self-efficacy effect than a direct effect of performance feedback. It also appears as though this effect is driven by women with low SES fathers and by women with no migration background.

These findings are important for several reasons. First, we show that the consequences of systematic grading bias are complex and go beyond the schooling and labour market consequences previously suggested in the literature – over-grading also impacts mental health – at least for women, which means that grades themselves, independent of knowledge changes, are important for both health and skills formation among students. Secondly, this has important policy implications. Given that grading bias has a non-negligible impact on mental health, any policy that directly or indirectly impacts grading in schools, for example, blind/non-blind grading of tests and the choice between standardised or discretionary grading systems, may also impact health and life chances among these young individuals, potentially in unintended ways. Future policy should consider this and be careful when manipulating factors that have implications for student grades.

The paper is organised as follows: in the next section, we present a conceptual framework where we discuss the relationship between grading bias and mental health in the context of the theory of developmental origins. In section three, we give a brief background to the Swedish schooling system with regards to the grading system and the relevant school reforms that impacted grade inflation. Sections four and five present our data and empirical strategy, we describe how we identify over-grading and mental ill-health and provide some balancing tests to support this identification. In section six, we present the results, including the main estimates of our analysis as well as cause-specific results for different definitions of mental ill-health and heterogeneity results. Finally, in sections number seven and eight, we discuss and conclude our findings.

2. Conceptual Framework

To help conceptualise how systematic grading bias could impact the development of mental health in young adult life, we turn to the human capital development literature in economics, an investment framework that unifies the literature on skills formation and production of health, see Cunha & Heckman (2007) and Heckman (2012). This framework describes how future outcomes such as education or earnings depends on one's current stock of capabilities. Capabilities at any given time consist of cognitive and non-cognitive skills (e.g. motivation, time preferences and social competence), and health. The idea is that the stock of capabilities evolves and that higher skills or better health in one period can affect the returns to any investment and thereby the stock of these capabilities in the next period (Cunha & Heckman, 2007). They are also potentially cross-fertilizing, for example, that emotional stability, motivation and good health can support the learning process, or that self-regulation and conscientiousness can reduce health risks (Cunha & Heckman, 2007). School grades reflect each student's cognitive and non-cognitive skills, but grades may also include systematic grading bias, i.e., over- or under-grading. If grading bias has a direct impact on health or skills, or an impact on the possibility to make investments in health or skills, the potential impacts of school-level grading bias may be observed across the full vector of capabilities, including mental health.

There are several potential causal channels for grading bias to influence the stock of capabilities, including mental health. First, there may be a direct link between over-grading and mental health. Grades are a form of feedback and feedback has been linked to several dimensions of mental health (Gustafsson et al., 2010). This is based on theories of symbolic interactionism, that children's perceptions about themselves are based on other people's appraisals about them, and that frequent positive (or negative) feedback impacts children's self-image which inhibits (or promotes) depression (ibid.). This would mean that being exposed to over-grading would have an immediate protective effect on an individual's stock of capabilities, through its impact on mental health.

A second potential channel is through the impact of over-grading on self-efficacy beliefs. A central question in social cognitive theory is related to how self-efficacy beliefs regulate cognitive, motivational, affective, and decisional processes (Bandura & Locke, 2003). Such beliefs could, for example, affect the choices we make at important decisional stages and if we think and act in self-enhancing or self-debilitating ways. They could also impact how we motivate ourselves in demanding situations and our vulnerability to stress and depression. According to Bandura & Locke (2003), self-efficacy beliefs are linked to differences in motivation and performances between individuals, but they are also linked to such differences in the same person over time. This means that self-efficacy beliefs could be responsive to changes in feedback that is a result of grading bias and, in turn, have an impact on mental health. Any change in self-efficacy would result in a direct change in the stock of capabilities, but if self-efficacy only impacts mental health through later positive feedback that is a consequence of the increased self-efficacy, this may take a while to impact the stock of mental health, in which case we would find the impact of grading bias on mental health only after a couple of years.

Third, over-grading could impact mental health through increased higher education enrolment opportunities. It has been shown that over-grading increases university enrolment. Schooling is linked to better mental health, but schooling is also expected to have positive impacts on later labour market outcomes, which also are correlated with better mental health (Lund et al., 2018; Macintyre, Ferris, Gonçalves, & Quinn, 2018).

Another aspect of the human capital development framework is that an individual's background characteristics, for example, social environment, migration background, parents' education and income, can be important for the development of capabilities. This is because acquired skills, health and investments over different stages of childhood bolster each other, and children from lower resource families usually have fewer opportunities to benefit from such multiplier effects which, according to (Cunha & Heckman, 2007), is a potential contributor to the socioeconomic gradient in health. It is possible that the three paths we hypothesise as potential channels between grading bias and mental health vary with family background. They may also vary based on gender. If constraints to higher education differ between groups in the population, or if self-efficacy responses to grades manifest themselves differently based on background characteristics and between women and men, then grading biases may also contribute to health gaps in the population. Previous findings indicate that this could be the case. Lavy & Sand (2015) show that teachers' discretionary grading favouring boys in math and science has a lasting positive effect on achievements and enrolment in advanced level courses among boys, yet a negative effect on girls. They also show that grading bias has spill-over effects across subjects (Lavy & Sand, 2015). Terrier (2015) concludes the same but for the opposite sex, when girls are over-graded in comparison to boys, they progress more, which indicates that grades themselves impact the development of skills also among women. Lavy & Sand (2015) moreover showed that the negative and positive impacts of grading bias on later school performance were intensified by both a socioeconomic and an ethnic gradient. Maurin & McNally (2008) showed that grading bias increased higher education enrolment among students on the margin of passing secondary school, reducing education gaps (since those on the margin of passing were more likely to be low socioeconomic status). Nordin et al. (2019) showed that over-grading increased university enrolment among men but not among women, and moreover that low-skilled women were affected negatively, potentially suggesting a negative self-efficacy effect of not being exposed to over-grading on school achievements among women.

3. The Swedish Schooling System

Several large reforms were implemented in the Swedish school system at the beginning of the 1990s. One reform, enacted in 1992, was related to a decentralization of Swedish schools which meant that the responsibility for primary and secondary schools was shifted from the central state to the local municipalities. Another reform introduced in the same year involved deregulation of the school financing which meant that privately run schools could now receive public funding through a voucher system. A third reform was related to school choice and increased the possibility for students to choose schools outside of their catchment area. Moreover, in 1994, the grading system was reformed from a norm-referenced relative system to a criterion-based goal-oriented system. In the prior system, grades were based on a scale from 1 – 5 and on a national level, the average grade should be a 3 with a standard deviation of 1. National standardized tests were given to help teachers grade the students according to the norm distribution, the class mean GPA should not diverge from the class mean of the national tests. With the reform, grades were based on the scale *fail* (IG), *pass* (G), *pass with distinction* (VG) and *pass with special distinction* (MVG). The distribution of grades was now based on how well the student met specific learning criteria in each subject, instead of a relative knowledge/performance comparison to one's peers. National course tests were also used in some subjects after the reform to guide the teachers in the grading process but the responsibility for grading is completely decentralized to the schools and the teachers, which means that the transparency of the grading in schools is quite low.

In Sweden, both upper secondary GPA and the SweSAT results function as ranking instruments for admission to higher education. Since upper secondary grades are based on the teacher's comprehensive assessment of student performance in relation to goals and grading criteria for each course, generally, grades reflect both cognitive and non-cognitive skills (e.g. presentation skills, ability to collaborate). Contrarily, the SweSAT is taken outside of schools and is administered by the Swedish Council for Higher Education. The SweSAT is not subject to teacher discretion and measures mainly cognitive skills, which means that the results are less likely to be biased. The deregulation and the marketization of the Swedish school system at the beginning of the 1990s introduced competition between schools. Because of the increased inter-school pressures, in combination with the restructuring of the Swedish grading system in 1994, teachers now had both the incentive and the possibility to over-grade, which resulted in substantial grade inflation in upper secondary schools in Sweden during the following decade (Vlachos, 2011; Wikström & Wikström, 2005).

4. Data

We use a population sample of students graduating from upper secondary schools in Sweden in the years 2001 – 2004 (n=299,459). To make sure that the grading bias measure is reliable we exclude students in schools that have less than 240 students in the four cohorts, i.e. less than on average 60 students per cohort (n=26,126 excluded). Only around 15 percent of students on the vocational tracks take the SweSAT, therefore, the sample is also restricted to students graduating from an academic track in upper secondary school, resulting in a substantial drop in our working sample (n=136,915 excluded). Finally, since we need compulsory school GPA to control for non-cognitive abilities, only students who received a 9th grade GPA in Sweden in the years 1998 – 2001 are included (n=4,554 excluded). The final sample comprises 131,864 individuals; 98.8 percent include cohorts born in 1982 – 1985, 0.4 percent is born in 1981 and 0.8 percent is born in 1986.

To this sample, we match register data from Statistics Sweden (SCB) on several background factors, compulsory and upper secondary school attainment and performance, as well as parents' socioeconomic and migration background. We also match data on municipal school quality indicators from the Swedish National Agency for Education, as well as psychiatric-related diagnoses and prescriptions from the National Patient Registers and the Prescribed Drug Register administered by the Swedish National Board of Health and Welfare.

5. Empirical strategy

In this section we describe our empirical strategy; first, we describe how we measure mental ill-health in our population sample of students; second, we describe how we obtain our school-level grading bias estimate; and third, we present our empirical model.

5.1 Measuring Mental Ill-Health

Internalising disorders[1] and substance use disorders[2] are the most common psychiatric diagnoses with onset in adolescence and young adulthood in Sweden (The National Board on Health and Welfare, 2020). Our main outcome measure, mental ill-health, is a dummy indicating the probability of at least one internalising- or substance use disorder diagnosis (inpatient or specialised outpatient care) during the five years following upper secondary school (including graduation year), or at least one prescription

[1] Internalising disorders include mood (affective) disorders (ICD-10 codes F30-39) and neurotic, stress-related and somatoform disorders (ICD-10 codes F40-48). In general, these are emotional or behavioural mental problems that are internalised – kept to themselves – by the individual, often characterised by a change in mood or affect, for example depression, a neurotic reaction to an environmental or stressful situation, or a general manifestation of anxiety.

[2] Substance use disorders include mental and behavioural disorders related to psychoactive substance use (ICD-10 codes F10-19), diagnoses that are attributable to acute intoxication, harmful use, dependence syndromes and other conditions related to the use of one or more psychoactive substances such as alcohol, narcotics or tobacco.

for a psychopharmaceutic drug[3] during 2005 – 2009[4]. For inpatient diagnosis, both the primary diagnosis as well as the top three secondary diagnoses are included.

The probability of being defined as mentally ill – any internalising or substance use diagnosis or prescription for a psychopharmaceutic drug – is 12.5 percent in our sample, see table 1. Prescriptions for psychopharmaceutic drugs are more common than internalising and substance use disorder diagnoses, 11.1 percent compared to 2.9 percent and 0.9 percent respectively. This relates to the fact that diagnoses only capture patients who are diagnosed in inpatient and outpatient care, while prescriptions capture patients who are prescribed drugs from anyone in the healthcare system, including even primary care. Thus, prescriptions for psychopharmaceutic drugs is likely a more comprehensive measure of internalising mental ill-health than inpatient and outpatient diagnoses.

Table 1: Descriptive Statistics of Mental Ill-Health

	Total *(n=131 864)*	Women *(n=72 292)*	Men *(n=59 572)*
	Mean	**Mean**	**Mean**
Mental ill-health (any of the below)	0.125	0.156	0.087
Psychopharmaceutic drug prescription	0.111	0.141	0.075
- Anxiolytics	0.054	0.069	0.036
- Antidepressants	0.077	0.099	0.049
- Hypnotics/sedatives	0.041	0.050	0.029
Internalising disorder diagnosis	0.029	0.038	0.018
- Inpatient	0.006	0.009	0.004
- Outpatient	0.027	0.035	0.017
Substance use disorder diagnosis	0.009	0.009	0.008
- Inpatient	0.004	0.004	0.004
- Outpatient	0.005	0.006	0.005

Notes: This table shows the mean probability of our mental ill-health outcomes among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). Diagnosis is measured in any of the five years following graduation (including graduation year) and prescription is measured in 2005 – 2009. Psychopharmaceutic drug prescription includes anxiolytics (anti-anxiety drugs; ATC code N05B), antidepressants (ATC code N06A), and hypnotics/sedatives (sleeping-pills; ATC code N05C); internalising disorders diagnosis includes mood (affective) disorders (ICD-10 codes F30-39) and neurotic, stress-related and somatoform disorders (ICD-10 codes F40-48), and substance use disorders include mental and behavioural disorders related to psychoactive substance use (ICD-10 codes F10-19).

[3] Psychopharmaceutic drugs are commonly used to treat internalising disorders and include anxiolytics (anti-anxiety drugs; ATC code N05B), hypnotics/sedatives (sleeping-pills; ATC code N05C) and antidepressants (ATC code N06A).

[4] The reason we use a different time-period to measure prescriptions is that register data on prescriptions was not available prior to 2005. Given that the follow up period is the same for all cohorts, if prescriptions vary with age, it is possible that the probability for prescription varies between the cohorts. However, this is controlled for by graduation year dummies.

In our sample, the probability of being defined as mentally ill is roughly twice as large among women compared to men. This is true both for inpatient and outpatient diagnosis for internalising disorders, as well as for prescriptions for anxiolytics, antidepressants and hypnotics/sedatives. The probability of getting diagnosed with substance use disorders is however quite equal between the sexes, inpatient diagnosis for substance use disorder is only slightly more common among men while outpatient diagnosis is slightly more common among women.

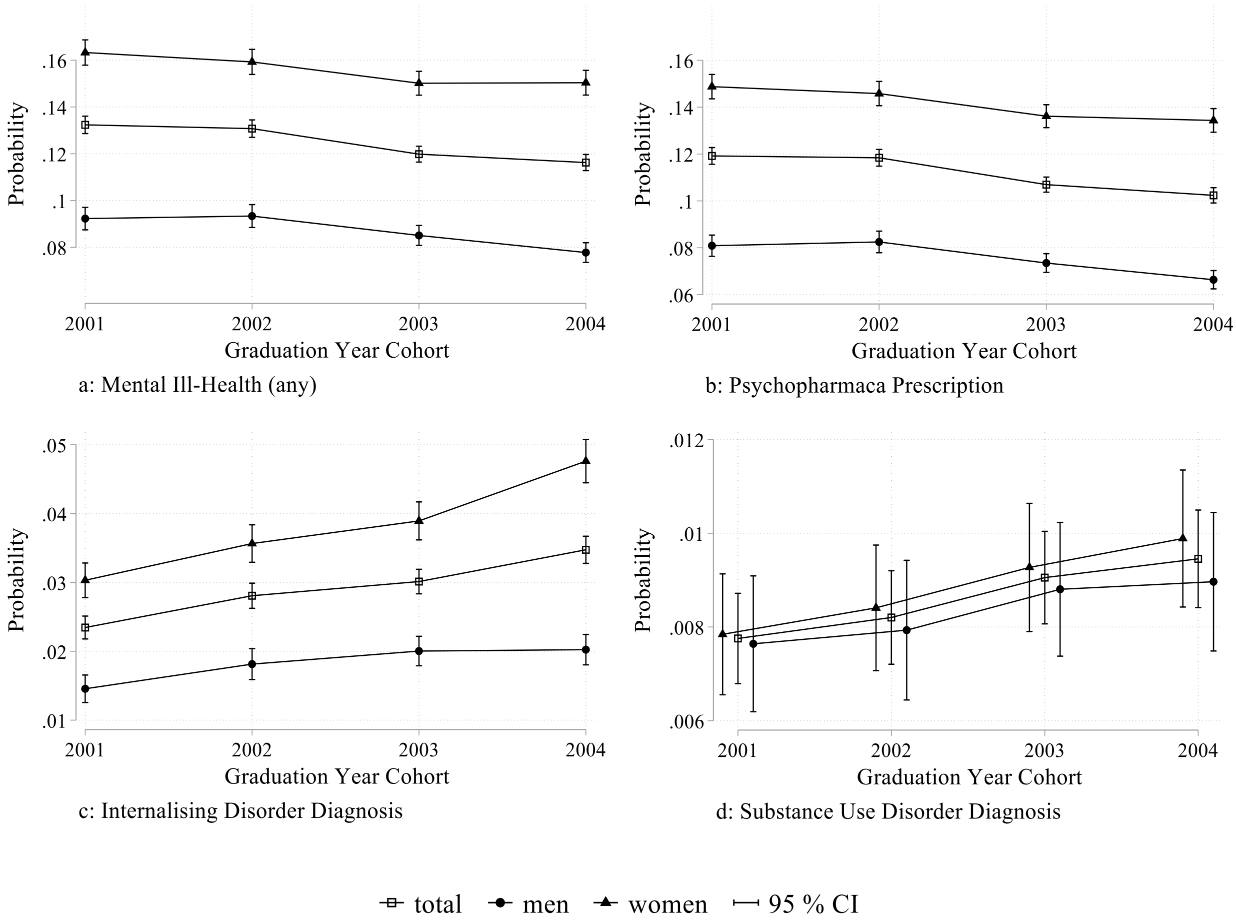

Figure 1: Trends in Mental Ill-Health

Notes: This figure shows the trends in mean probability of our mental ill-health outcomes among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). Diagnosis is measured in any of the five years following graduation (including graduation year) and prescription is measured in 2005 – 2009. Psychopharmaceutic drug prescription includes anxiolytics (anti-anxiety drugs; ATC code N05B), antidepressants (ATC code N06A), and hypnotics/sedatives (sleeping-pills; ATC code N05C); internalising disorders diagnosis includes mood (affective) disorders (ICD-10 codes F30-39) and neurotic, stress-related and somatoform disorders (ICD-10 codes F40-48), and substance use disorders include mental and behavioural disorders related to psychoactive substance use (ICD-10 codes F10-19). In panel d, the markers for men and women are slightly shifted sideways to improve visualisation, but these belong to the same graduation year cohorts.

In figure 1, the trends in mental ill-health are shown for women and men in the different graduation year cohorts. Similarly, as for the overall mean probabilities, these trends show that prescriptions for psychopharmaceutic drugs and internalising disorder diagnosis are significantly larger among women compared to men, and while a slight difference in substance use disorder diagnosis emerges over time between women and men, this difference is not statistically significant. The trend of any mental ill-health decreases over time. This appears to be driven by a significant decrease in prescriptions among women (statistical tests show that the change over time is significant among women but not among men). At the same time, internalising disorder diagnosis increases significantly among women, and substance use disorder diagnosis increases significantly among women and men. The fact that the probability of mental ill-health differs between women and men, and between the different mental ill-health indicators, will be considered in our empirical analyses.

5.2 Measuring Grading Bias

We identify grading bias following Nordin et al. (2019), which in turn is a development of Wikström & Wikström (2005). We measure grading bias as the difference between upper secondary GPA and SweSAT results, on school level, for each cohort[5]. Some students take the SweSAT more than once. Number of times taking the test (learning effect) and the time before or after graduation (age effect) could impact the test score (Nordin et al., 2019). Therefore, we regress each student *i*'s test score, $SweSAT_{i,s,l}$, on two sets of dummies; the first dummy, $\gamma_{i,s}$, indicating how many times (s=1,2,…,s) the student has taken the test and the second dummy, $\delta_{i,l}$, indicating the time to graduation *l* when the test is taken:

$$SweSAT_{i,s,l} = \alpha + \gamma_{i,s} + \delta_{i,l} + SweSAT_{i,s,l}^{res} \tag{1}$$

The residual test score, $SweSAT_{i,s,l}^{res}$, is saved and a mean residual test score, $\overline{SweSAT_{i,t}^{res}}$, is calculated and standardised for each student in graduation year *t*. Next, we regress each student's upper secondary grade in graduation year *t*, $GPA_{i,t}$, on his or her mean residual test score:

$$GPA_{i,t} = \overline{SweSAT_{i,t}^{res}} + \pi_{i,t} \tag{2}$$

The standardized version of this residual, $\pi_{i,t}$, is the individual level divergence between upper secondary GPA and mean SweSAT. In figure 2, we show mean upper secondary GPA ($GPA_{tt,t}$) and

[5] In Nordin et al. (2019) the grading bias measure was averaged up separately for women and men with the motivation that grading bias might differ between the sexes. In this study, we have calculated the grading bias measure on school level for women and men together, but we have checked the robustness of our results to this alternative identification of grading bias and there is no substantial difference in our results or conclusions.

mean difference between GPA and SweSAT ($\pi_{tt,t}$) for the test-taking population, *tt*, in each graduation year *t*. The trends show that mean GPA and the divergence between GPA and SweSAT follow each other closely, but that the divergence between GPA and SweSAT increases relatively more than mean GPA, suggesting that over-grading becomes more common during this period.

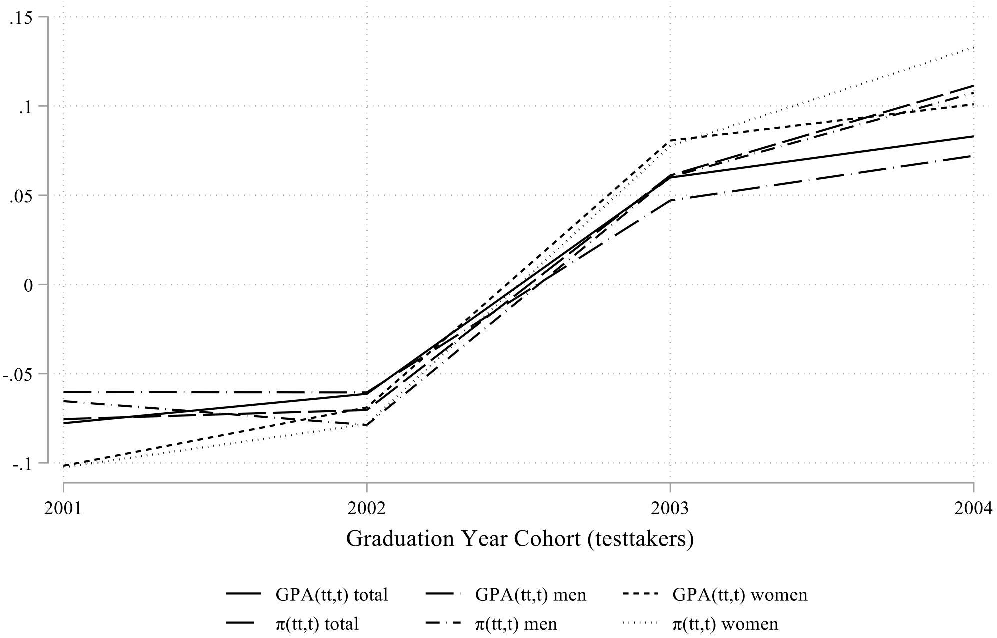

Figure 2: Trends in Upper Secondary GPA and Divergence GPA-SweSAT

Notes: This figure shows the standardised upper secondary GPA mean (GPA_{tt},t) and grading bias (standardized mean difference between GPA and SweSAT, π_{tt},t) among test-taking, tt, students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample).

By averaging the divergence between upper secondary GPA and mean SweSAT on the school level for each school j, we get our measure of grading bias, $\pi_{j,t}$. Since this measure is calculated on the school level, students who have not yet taken the SweSAT test can also be included. In this paper, we investigate the impact of grading bias over a period when grade inflation is observed in Sweden (Vlachos, 2011; Wikström & Wikström, 2005), which is why we assume that the estimated impact reflects over-grading.

5.3 Our Empirical Model

We estimate the impact of over-grading on mental ill-health using a school fixed effects model:

$$Y_{i,j,t} = \alpha_j + \delta_t + \beta\overline{\pi_{j,t}} + \boldsymbol{X'_{j,t}}\gamma + \varepsilon_i \quad (3)$$

The school fixed effects, α_j, controls for all time-invariant differences between the schools such as systematic sorting of students into different schools and differences in educational investment quality. We also use graduation year dummies, δ_t, to capture cohort variation in mental health. β is our coefficient of interest and can be interpreted as the impact of within-school (j) over-time (t) changes in grading bias, $\pi_{j,t}$, on the probability of receiving a mental ill-health-related diagnosis or prescription, $Y_{i,j,t}$. A vector of control variables is included in $\boldsymbol{X}$. We control for changing inflow of non-cognitive abilities by including the mean of 9th grade GPA on school level (assuming that the SweSAT results controls for changes in cognitive abilities). Moreover, if the inflow to different tracks with varying difficulty levels in the same school changes over time, it could have an impact on the grades in that school. Therefore, we control for the proportion of students on each different track for each school and year. To account for changing student characteristics, we also include measures for sex balance and the share with any foreign background (foreign-born or second-generation with at least one foreign-born parent) on school level for each year. As mentioned above, the school fixed effects will deal with differences in quality or resources between schools, however, resources may change over time. For example, an increased inflow of students with foreign backgrounds would result in increased reimbursements from the municipality. To control for such changes, we include several school resources indicators on the municipal level: teacher/pupil ratio, the share of qualified teachers and the share of teachers permanently employed.

Equation (3) thereby identifies the impact of grading bias on mental health, given that we have adequately controlled for all potential variables correlated with both grading bias and mental health. We assume this is to be true and present some tests, that are in the spirit of balancing tests, to support our assumption below. In sensitivity analyses, we also include linear time-trends for each municipality as a test of robustness to modelling choice. Standard errors are clustered on the school level in all regressions.

5.4 Adjusting Grading Bias

The unadjusted measure of grading bias $\pi_{j,t}$ is adjusted for each school using the following equation:

$$\overline{\pi_{j,t}} = \alpha_j + \delta_t + \boldsymbol{X}'_{j,t}\gamma + u_{j,t} \tag{4}$$

Saving the residuals $u_{j,t}$ we can then re-write equation (3) as:

$$Y_{i,j,t} = \widehat{u_{j,t}} + \varepsilon_i \tag{5}$$

In figure 3, to help visualise the variation we are using, we chart our unadjusted measure of grading bias $\overline{\pi_{j,t}}$ and the residuals $\widehat{u_{j,t}}$ after adjusting for the included controls in a histogram. This shows that the included controls adjust for some of the variation over time, and when all the controls are included, there is variation left, which we use to study the impact of grading bias on mental health.

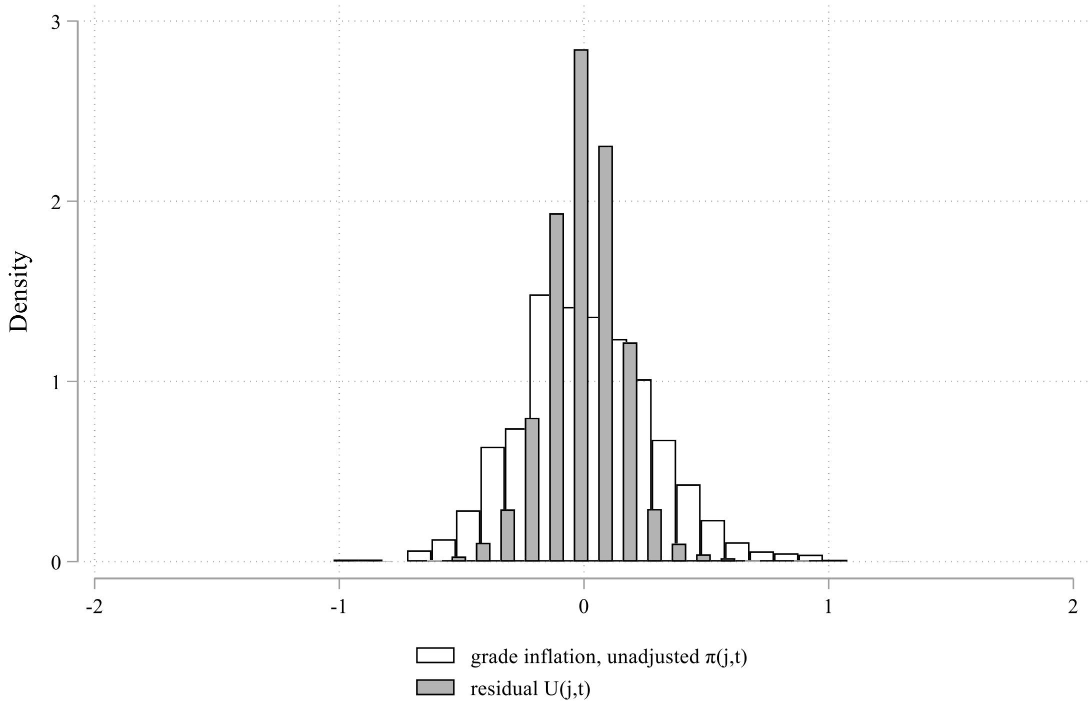

Figure 3: Adjusting the Grading Bias Measure

Note: This figure shows our unadjusted grading bias measure and the residual grading bias measure adjusted for school and graduation year fixed-effects and additional controls on compulsory grades, tracks, sex and migration share (all on school level for each year), as well as school quality indicators (on municipal level). School-level grading bias is measured among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample).

5.5 Balancing Tests

To provide support for the validity of our identification strategy we assess the balance of our sample with respect to grading bias using so-called balancing tests. We examine whether several predetermined characteristics – father's income and education, migration background and below-median 9th grade GPA – are predicted by grading bias ($\pi_{j,t}$) by including them as dependent variables in equation (3). The results of these tests are shown in table 2. When only school and graduation year fixed effects are used as controls (column (1) of table 2) grading bias is positively associated with fathers' years of schooling and associated with having below median GPA in 9th grade. However, the inclusion of compulsory school grades yields estimates very close to zero with no substantial effect on standard errors. Further additional controls for tracks, sex, migration share, and school quality indicators have no additional effect (see columns 3, 4 and 5 respectively). This implies that the additional controls are not necessary, and supports our assumption of no omitted variable bias.

Table 2 Balancing tests:

The Impact of Grading Bias on Predetermined Characteristics

	(1)	(2)	(3)	(4)	(5)
			Fathers' education		
Grading bias	0.171**	0.00792	0.0169	0.0874	0.0752
	(0.0787)	(0.0832)	(0.0851)	(0.0739)	(0.0761)
Constant	11.56***	11.56***	11.74***	12.52***	12.46***
	(0.000500)	(0.000529)	(0.110)	(0.119)	(0.737)
			Fathers' income		
Grading bias	0.0398	-0.0155	-0.0216	0.0144	0.0159
	(0.0339)	(0.0348)	(0.0342)	(0.0281)	(0.0287)
Constant	6.952***	6.953***	6.956***	7.358***	7.229***
	(0.000215)	(0.000221)	(0.0449)	(0.0488)	(0.224)
			Foreign-born		
Grading bias	-0.00201	0.00599	0.00868	0.00668	0.00671
	(0.00605)	(0.00627)	(0.00619)	(0.00610)	(0.00632)
Constant	0.0744***	0.0743***	0.0683***	0.0462***	0.100*
	(3.85e-05)	(3.98e-05)	(0.00748)	(0.0105)	(0.0514)
			Foreign background		
Grading bias	-0.00204	0.0118	0.0177*	0.0150*	0.0165*
	(0.00875)	(0.00885)	(0.00911)	(0.00907)	(0.00934)
Constant	0.207***	0.207***	0.197***	0.167***	0.183**
	(5.57e-05)	(5.62e-05)	(0.0130)	(0.0166)	(0.0833)
			Below median 9th grade Grade Point Average (GPA)		
Grading bias	-0.107***	0.00683	0.00255	-0.00345	-0.00345
	(0.0180)	(0.00598)	(0.00613)	(0.00560)	(0.00570)
Constant	0.501***	0.500***	0.497***	0.430***	0.373***
	(0.000115)	(3.80e-05)	(0.00753)	(0.0117)	(0.0582)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*

Note: This table presents the results for the impact of grading bias for selected predetermined characteristics among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year and an indicator for missing outcome variable, additional controls (on school level for each year) are added for each specification with 9th grade GPA in (2), tracks in (3), sex and migration share (not for "Foreign-born" and "Foreign background") in (4) and school quality indicators (on municipal level) in (5). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

6. Results

Our main results are presented in table 3 and show the impact of grading bias on mental ill-health. Grading bias is negatively associated with our broadest measure of mental ill-health – any internalising or substance use diagnosis or prescription for a psychopharmaceutic drug – among women, but not in the full sample (women and men together) or among men separately. One standard deviation (sd) increase in grading bias is significantly associated with between 1.85 and 2.82 percentage points (pp) lower probability of mental ill-health among women, corresponding to relative effects between 11 and 14 percent. However, only around one percent of schools experienced increases in grading bias of that size, while 39 percent of schools experienced a quarter of an sd increase in grading bias. Therefore, we interpret our results by corresponding increases in grading bias. A quarter of an sd increase in grading bias is significantly associated with a relative decrease in the probability of mental ill-health among women just above three percent. The results are quite stable over the different specifications (columns 1 – 5), but only weakly significant (significant on the 10 percent level) when tracks are added in specification (3).

The descriptive data show that the probability of mental ill-health differs substantially between the different mental ill-health indicators; being prescribed psychopharmaceutic drugs is more common than internalising diagnosis which in turn is substantially more common than substance use disorder diagnosis. We also find that the trends over time go in varying directions for these different measures of mental ill-health, prescriptions decrease while diagnoses increase. Therefore, we conduct our analyses on the mental ill-health outcomes – psychopharmaceutic drug prescription, internalising disorder diagnosis and substance use disorder diagnosis – separately. Moreover, since the results so far indicate that the impact of grading bias on mental health differs between the sexes, all analyses from now on are performed for men and women separately. We present only our preferred model estimates in table 4, which includes all the control variables (specification (5) in table 3) leaving full regression output tables to the Appendix (see Tables A2-A4). The results show that an sd increase in grading bias is significantly associated with a 1.95 pp lower probability of psychopharmaceutic prescription and a 1.66 pp lower probability of internalising disorder diagnosis, among women. For a quarter of an sd increase in grading bias, this relates to relative effects of four and eight percent lower probability of psychopharmaceutic prescription and internalising disorder diagnosis, respectively, among women. There is no association between grading bias and substance use disorders, neither among men nor women.

Table 3 Main Results:

The Impact of Grading Bias on Probability of Mental Ill-Health

	(1)	(2)	(3)	(4)	(5)
			Mental ill-health (any)		
Full sample					
Grading bias	-0.00899	-0.00372	-0.00434	-0.00939	-0.0101
	(0.00585)	(0.00609)	(0.00629)	(0.00587)	(0.00614)
Constant	0.125***	0.125***	0.119***	0.0630***	0.120**
	(3.72e-05)	(3.86e-05)	(0.00882)	(0.0115)	(0.0585)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*
Women					
Grading bias	-0.0230***	-0.0185**	-0.0165*	-0.0195**	-0.0232**
	(0.00812)	(0.00876)	(0.00908)	(0.00895)	(0.00933)
Constant	0.156***	0.156***	0.168***	0.135***	0.171*
	(0.000191)	(0.000222)	(0.0133)	(0.0207)	(0.0932)
Observations	*72,272*	*72,272*	*72,272*	*72,272*	*68,927*
Men					
Grading bias	-0.00239	0.00502	0.00345	0.00205	0.00479
	(0.00729)	(0.00751)	(0.00781)	(0.00781)	(0.00794)
Constant	0.0867***	0.0869***	0.0722***	0.0566***	0.125*
	(0.000105)	(0.000129)	(0.0105)	(0.0154)	(0.0725)
Observations	*59,569*	*59,569*	*59,569*	*59,569*	*56,973*

Note: This table presents the results for the impact of grading bias on the probability of mental ill-health among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year, additional controls (on school level for each year) are added for each specification with 9th Grade Point Average (GPA) in (2), tracks in (3), sex and migration share in (4) and school quality indicators (on municipal level) in (5). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

Table 4 Main Results:

The Impact of Grading Bias on Probability of Psychopharmaceutic Drug Prescription, Internalising- and Substance Use Disorder Diagnoses

	(1)	(2)	(3)
	Psychopharma-ceutic drug prescription	**Internalising disorder diagnosis**	**Substance use disorder diagnosis**
Women			
Grading bias	-0.0195**	-0.0166***	0.00104
	(0.00872)	(0.00550)	(0.00295)
Constant	0.125	0.0518	0.0175
	(0.0907)	(0.0455)	(0.0207)
Observations	*68,927*	*68,927*	*68,927*
Men			
Grading bias	0.00526	-0.00361	0.00216
	(0.00731)	(0.00420)	(0.00251)
Constant	0.108	0.00607	-0.00204
	(0.0683)	(0.0352)	(0.0218)
Observations	*56,973*	*56,973*	*56,973*

Note: This table presents the results for the impact of grading bias on the probability of psychopharmaceutic drug prescription, internalising- and substance use disorder diagnoses among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

6.1 Timing

A direct effect of grading bias on mental health would likely occur immediately after the performance feedback received at graduation, while mechanisms through self-efficacy beliefs or increased schooling likely would show health impacts only after a couple of years or so, depending on how the development of capabilities is affected. To understand these potential underlying mechanisms, it is important to understand when the health impact occurs. Therefore, we investigate how grading bias impacts mental ill-health in each of the five years following graduation.

The results are presented in figure 4 and show that the protective effect of over-grading on mental health appears among women three to four years following graduation. The trends are similar for psychopharmaceutic prescriptions and internalising disorder diagnosis, but statistically significant only for the latter. Similarly as in our baseline results, grading bias is not related to substance use disorders among women or any of the mental ill-health outcomes among men.

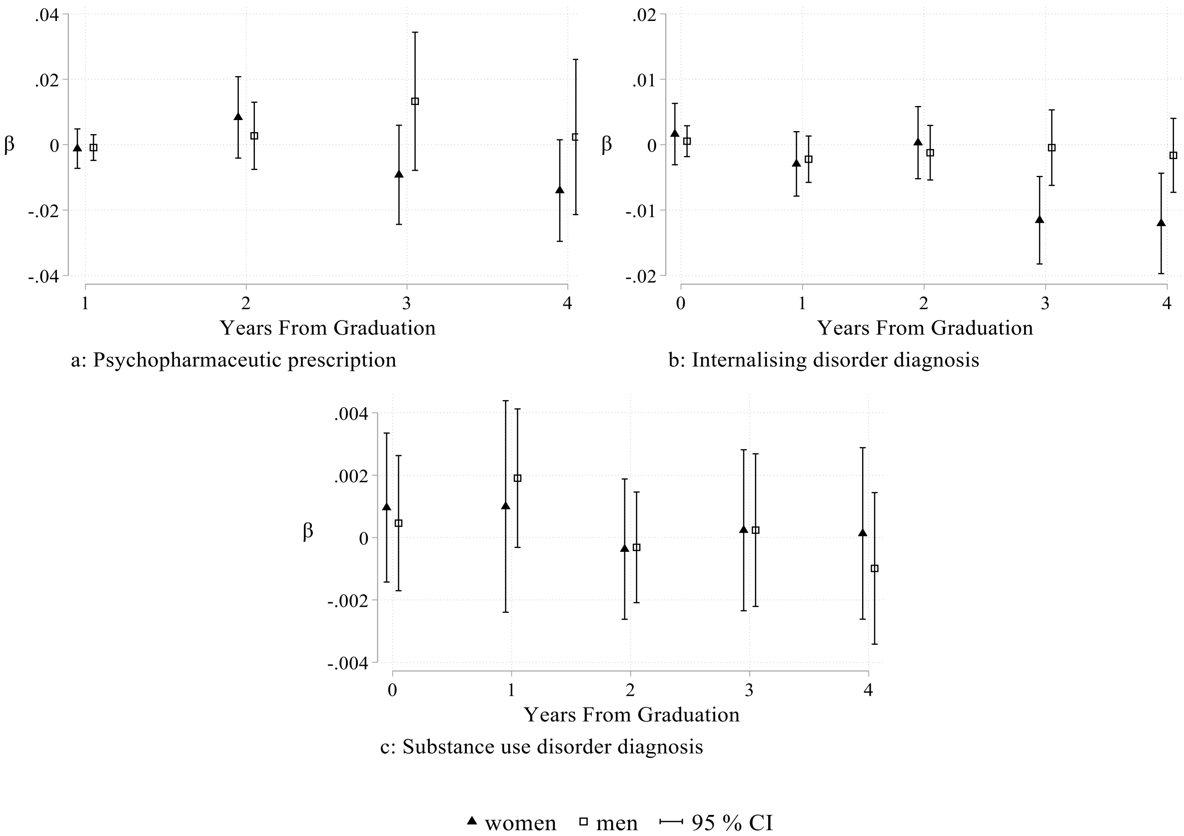

Figure 4: Timing of the Impact of Grading Bias on Mental Ill-Health

Notes: This figure shows the impact of grading bias on our mental ill-health outcomes (β in Eq. 3) for each year following graduation among male and female students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). Data on pharmaceutic prescriptions is only available from 2005 which is why we cannot measure this outcome in the year of graduation (year 0). Psychopharmaceutic prescription in year 1 is measured in graduation cohort 2004, year 2 is measured in cohorts 2003 and 2004, year 3 in cohorts 2002 – 2004, and year 4 in cohorts 2001 – 2004.

6.2 Heterogeneity

The human capital development framework of Cuhna et al. (2007) suggests that background characteristics and skills in the early stages of life are important for the development of capabilities (skills and health). Our main results show that over-grading has a protective impact on mental health among women but not among men. In this section, we test if grading bias impacts mental health differently in different subgroups of the population stratified by skills (above/below median compulsory school GPA), migration background (foreign-born or any foreign background), father's education level (higher education or not) and father's income (above/below median income).

The heterogeneity results indicate that the relationship between grading bias and mental ill-health differs with fathers' SES (see table 5A). The protective relationships between over-grading and psychopharmaceutic drugs and internalising disorder diagnosis are significant only among women with low SES fathers. Among women with high SES fathers, the estimate sizes for these relationships are substantially smaller, and even reversed for the impact of over-grading on psychopharmaceutic prescriptions among women with high-income fathers, but the effect estimates are not significantly different from zero. Also, the results indicate that there may be heterogeneity related to fathers' SES among men; over-grading is related (significant on the 10 % level) to a higher probability of prescriptions for psychopharmaceutic drugs among men with high-income and high educated fathers.

The heterogeneity results indicate differences based on migration background as well. The protective relationships between over-grading and psychopharmaceutic drugs and internalising disorder diagnosis are significant only among women who are born in Sweden and among women with no foreign background. Interestingly, while our main results show no relationship between grading bias and substance use disorders, the heterogeneity results indicate that there may be a link between over-grading and substance use disorders among men, but that this relationship differs with migration background. Over-grading is related (significant on the 10 % level) to a lower probability of substance use disorders among men with foreign backgrounds and a higher probability of these disorders among Swedish-background men.

Finally, there is heterogeneity also based on skills. The protective relationship between over-grading and internalising disorder diagnosis is significant only among women who had above median GPA in compulsory school (see Table 5C), for below-median GPA women the impact is substantially smaller but not significantly different from zero. The effect estimates for the relationship between over-grading and psychopharmaceutic prescription that we find in the full group of women is not significant in either subgroup when stratified by compulsory school GPA, but they are of similar size.

Table 5 Heterogeneity analysis:

The Impact of Grading Bias on Probability of Mental Ill-Health

A: By Socioeconomic Background

	(1) Psychopharma-ceutic drug prescription	(2) Internalising disorder diagnosis	(3) Substance use disorder diagnosis
Women			
Grading bias (low educated father)	-0.0234**	-0.0184***	-0.00186
	(0.0108)	(0.00686)	(0.00331)
Grading bias (high educated father)	-0.0102	-0.00977	0.00761
	(0.0171)	(0.0110)	(0.00521)
Constant	0.126	0.0495	0.00330
Observations: 68,925	(0.0912)	(0.0454)	(0.00408)
Grading bias (low-income father)	-0.0455***	-0.0195**	-0.00118
	(0.0119)	(0.00796)	(0.00410)
Grading bias (high-income father)	0.0136	-0.0121	0.00330
	(0.0139)	(0.00861)	(0.00408)
Constant	0.111	0.0553	0.0146
Observations: 68,925	(0.0898)	(0.0459)	(0.0205)
Men			
Grading bias (low educated father)	-0.00520	-0.00862	-0.000159
	(0.00941)	(0.00533)	(0.00247)
Grading bias (high educated father)	0.0245*	0.00546	0.00665
	(0.0133)	(0.00671)	(0.00449)
Constant	0.128*	0.00309	-0.00339
Observations: 56,971	(0.0682)	(0.0350)	(0.0216)
Grading bias (low-income father)	-0.00903	-0.00996	0.00424
	(0.0105)	(0.00668)	(0.00352)
Grading bias (high-income father)	0.0189*	0.00177	-0.000185
	(0.00998)	(0.00517)	(0.00317)
Constant	0.114*	0.00806	-0.00279
Observations: 56,972	(0.0688)	(0.0356)	(0.0219)

Note: This table presents heterogeneity by socioeconomic background for the impact of grading bias on the probability of psychopharmaceutic drug prescription, internalising- and substance use disorder diagnoses among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions in each subgroup. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Robust standard errors clustered at the school level are shown in the parentheses.
***p<0.01, **p<0.05, *p<0.1

B: By Migration Background

	(1)	(2)	(3)
	Psychopharma-ceutic drug prescription	**Internalising disorder diagnosis**	**Substance use disorder diagnosis**
Women			
Grading bias (born in Sweden)	-0.0224**	-0.0169***	0.000299
	(0.00913)	(0.00572)	(0.00316)
Grading bias (foreign-born)	0.00426	-0.0134	0.00988
	(0.0281)	(0.0187)	(0.00908)
Constant	0.132	0.0556	0.0164
Observations: 68,921	(0.0903)	(0.0461)	(0.0205)
Grading bias (Swedish background)	-0.0258**	-0.0229***	-0.00151
	(0.0104)	(0.00680)	(0.00346)
Grading bias (foreign background)	0.000911	0.00251	0.00862
	(0.0189)	(0.0123)	(0.00554)
Constant	0.137	0.0583	0.0189
Observations: 68,923	(0.0909)	(0.0469)	(0.0207)
Men			
Grading bias (born in Sweden)	0.00438	-0.00312	0.00381
	(0.00733)	(0.00411)	(0.00270)
Grading bias (foreign-born)	0.0159	-0.00405	-0.0103
	(0.0258)	(0.0127)	(0.00660)
Constant	0.102	0.00963	-0.00406
Observations: 56,962	(0.0695)	(0.0348)	(0.0220)
Grading bias (Swedish background)	0.00701	-0.00298	0.00554*
	(0.00855)	(0.00460)	(0.00299)
Grading bias (foreign background)	-0.000878	-0.00522	-0.00733*
	(0.0135)	(0.00702)	(0.00422)
Constant	0.114	0.00631	-0.000650
Observations: 56,971	(0.0697)	(0.0357)	(0.0222)

Note: This table presents heterogeneity by migration background for the impact of grading bias on the probability of psychopharmaceutic drug prescription, internalising- and substance use disorder diagnoses among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions in each subgroup. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Robust standard errors clustered at the school level are shown in the parentheses.
***p<0.01, **p<0.05, *p<0.1

C: By Compulsory School GPA

	(1) Psychopharma-ceutic drug prescription	(2) Internalising disorder diagnosis	(3) Substance use disorder diagnosis
Women			
Grading bias (below-median GPA)	-0.0165	-0.0105	-0.00141
	(0.0125)	(0.00730)	(0.00507)
Grading bias (above-median GPA)	-0.0218	-0.0235***	0.00512
	(0.0144)	(0.00822)	(0.00314)
Constant	0.147	0.0541	0.0161
Observations: 68,924	(0.0915)	(0.0473)	(0.0213)
Men			
Grading bias (below-median GPA)	-0.00402	-0.00554	0.00214
	(0.0101)	(0.00525)	(0.00373)
Grading bias (above-median GPA)	0.0157	-0.000845	0.00259
	(0.0115)	(0.00666)	(0.00321)
Constant	0.123*	0.00390	-0.000675
Observations: 56,970	(0.0689)	(0.0356)	(0.0222)

Note: This table presents heterogeneity by compulsory school GPA for the impact of grading bias on the probability of psychopharmaceutic drug prescription, internalising- and substance use disorder diagnoses among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions in each subgroup. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Robust standard errors clustered at the school level are shown in the parentheses.
***p<0.01, **p<0.05, *p<0.1

6.3 Cause-Specific Results

In our main analyses, the diagnosis probability is measured by combined inpatient and outpatient diagnoses. Moreover, the probability of prescriptions for psychopharmaceutic drugs is measured by anxiolytics, antidepressants and hypnotics/sedatives (sleeping pills) combined, preparations that are mainly used to treat internalising disorders such as anxiety and depression. Likely though, inpatient and outpatient diagnoses reflect somewhat different manifestations of mental problems, especially regarding symptom severity. Thus, the impact of grading bias may differ between inpatient and outpatient diagnoses, and between the different drugs.

In table 6 we investigate the impact of grading bias separately across inpatient and outpatient diagnoses and find that over-grading is associated with a lower probability of outpatient diagnoses among women. For inpatient diagnoses, the relationship is only weakly significant. Our cause-specific results moreover show the protective impact of over-grading on prescriptions for psychopharmaceutic drugs among women is driven by antidepressants. Over-grading is associated with a 2.01 pp lower probability of prescription for antidepressants among women. We also find weak indications that over-grading is associated with a 0.844 pp higher probability of prescriptions for anxiolytics among men. Similar to our baseline results, grading bias is not related to in- or outpatient substance use disorder diagnoses (results not shown).

Table 6 Cause-Specific Analyses:

The Impact of Grading Bias on Inpatient-/Outpatient Internalising Disorder Diagnoses and Different Psychopharmaceutic Drugs

	(1)	(2)	(3)	(4)	(5)
	Inpatient internalising disorder diagnosis	**Outpatient internalising disorder diagnosis**	**Anxiolytics**	**Anti-depressants**	**Hypnotics/ Sedatives**
Women					
Grading bias	-0.00472*	-0.0140***	-0.00592	-0.0201**	0.000207
	(0.00263)	(0.00529)	(0.00577)	(0.00831)	(0.00573)
Constant	0.000326	0.0494	0.153**	0.0106	0.112*
	(0.0265)	(0.0442)	(0.0642)	(0.0803)	(0.0577)
Observations: 68,927					
Men					
Grading bias	-0.00261	-0.00365	0.00844*	0.000402	-0.00271
	(0.00216)	(0.00404)	(0.00511)	(0.00609)	(0.00465)
Constant	-0.0104	0.0103	0.0734	0.0817	0.0115
	(0.0176)	(0.0333)	(0.0515)	(0.0526)	(0.0429)
Observations: 56,973					

Note: This table presents the results for the impact of grading bias on the probability of in- and outpatient internalising disorder diagnoses and different psychopharmaceutic drugs among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

6.4 Sensitivity Analysis

Our baseline model assumes a linear relationship between grading bias and mental ill-health. We assess the linear functional form assumption of our model between grading bias and mental ill-health by estimating the relationship between grading bias and mental ill-health in quintiles of grading bias from low bias level (Q1) to high bias level (Q5), see figure 5. Compared to the reference group (Q3, reflecting close to median levels of grading bias) psychopharmaceutic prescription is lower among women in the slightly higher bias quintile (Q4), but the difference is only significant on 10 % level. For internalising diagnosis, it is only women in the lowest bias quintile (Q1) that differs from Q3 with a significantly higher probability of diagnosis. These results indicate that the relationship between grading bias and mental ill-health might not be completely linear, however, the direction of the sign for the β estimates in the different quintiles indicate that the protective effect we find between grading bias and psychopharmaceutic prescription and internalising disorder diagnosis in our main results is present for women in schools with higher bias levels, that is, where we know over-grading took place.

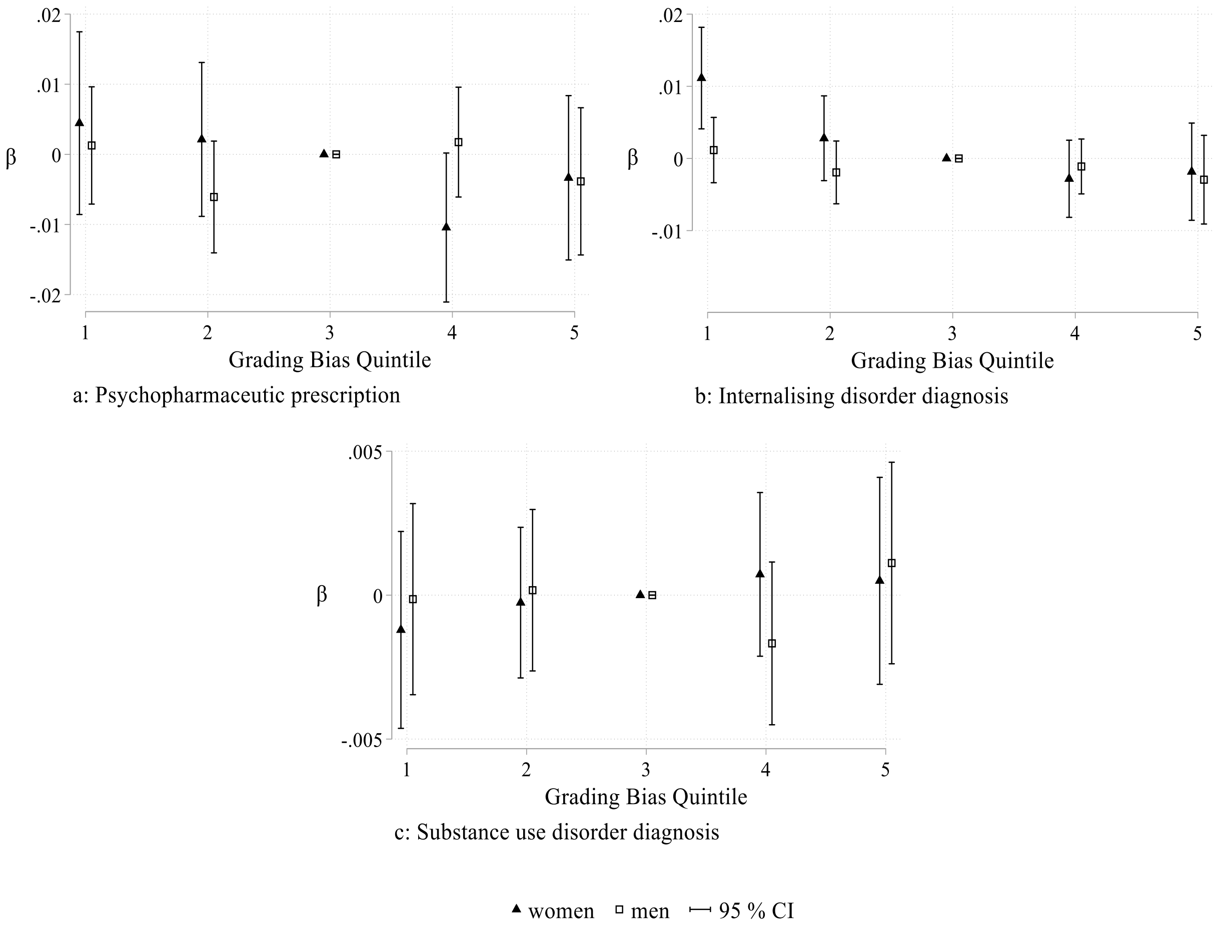

Figure 5 Sensitivity Analysis:

The Impact of Grading Bias in Quintiles on Mental Ill-Health

Note: This figure presents the impact of grading bias in quintiles on the probability of psychopharmaceutic drug prescription, internalising- and substance use disorder diagnoses among students graduating from academic track in upper secondary schools in Sweden in the years 2001 – 2004 (see data section for specific information about the sample). Grading bias is divided into quintiles 1 – 5 from lowest bias (Q1) to highest bias (Q5), Q3 is the reference group and represents the quintile of around median levels of bias. All models include school fixed-effects and the following included controls: graduation year; on school level for each year: 9th grade Grade Point Average (GPA), the share of students on different tracks, sex- and migration share; as well as school quality indicators on the municipal level. Standard errors are clustered at the school level.

In additional sensitivity analyses, we have also controlled our main results for linear time trends at the municipality level which are shown to have negligible impact on our results, see Tables A1-A4 in Appendix.

7. Discussion

Our results show that women who were exposed to over-grading in Sweden were substantially less likely to suffer from mental ill-health in the following years. Being exposed to a quarter of an sd increase in grading bias, observed in more than a third of the schools, is associated with an eight percent lower probability of being diagnosed for internalising disorders such as depression and anxiety, and a four percent lower probability of being prescribed mental ill-health related drugs such as antidepressants and anxiolytics. We find no effects of grading errors among men. We show that the measure we use to estimate grading bias does not predict several predetermined characteristics such as migration background, father's SES and compulsory school grades. The results are also robust to the inclusion of important control variables and even municipality level trends. This indicates that our independent variable of interest is balanced, and provides support that our model is, in fact, a valid model to investigate grading bias effects.

We hypothesized that over-grading impacts mental health, either through a direct effect of changes in feedback on mental health or through a self-efficacy effect by which over-grading increase self-confidence, motivation and performance and thereby impact mental health. We also hypothesize that over-grading impacts mental health through increased higher education opportunities which are correlated to better mental health. Our results on the timing of the protective effect on mental health among women find that this protective effect occurs three to four years after graduation. This indicates that there is not a direct effect of grading bias on mental health. Moreover, we know from previous research that over-grading increases higher education enrolment among men only (Nordin et al., 2019). This indicates that it is less likely that the grading bias effect on mental health among women is driven only by higher education enrolment. Instead, this supports the hypothesis that self-efficacy may be a relevant mechanism in the protective relationship between over-grading and mental health among women.

Our finding of clear differences across sexes raises the question if women and men are differently exposed to over-grading. Previous work by Lindahl (2007) suggests that girls are over-graded in comparison to boys in Sweden. She compares grades between girls and boys who have the same score on national tests in different subjects and finds that girls receive higher final grades in these subjects at the end of compulsory school. Possibly, teachers include other criteria besides the national test score in the final grade and if girls outperform boys in these tasks, it could be an explanation of this result. In another study from Sweden, Hinnerich, Höglin, & Johannesson (2011) uses both blind and non-blind examples of the same test to compare grading between boys and girls, and show that boys are not discriminated against in grading. Our findings suggest that grades and grading biases increased both among men and women during our study period (see figure 2), suggesting trends of inflated grades and exposure to over-grading in both groups. However, grades and grading biases increased relatively more among young women, so it is possible that over-grading was more intense among women.

Besides heterogeneity between women and men, we also find that the impact of over-grading on mental health appears to differ based on socioeconomic and migration background, and somewhat also by skills. Over-grading is protective against mental ill-health mainly among women with low-SES fathers and among women with no foreign background. Given that there is a documented socioeconomic gradient in mental health (see e.g. Linder et al. (2020)), this heterogeneity would actually reduce mental health gaps, all other things equal. There are also some (weaker) indications that over-grading increases the risk for psychopharmaceutic drug prescriptions among men with higher education and high-income fathers, and that over-grading increases the risk of substance use disorders among men with no foreign background and reduces the risk of substance use disorders among men with foreign backgrounds. This is interesting because in the main analyses we do not find any association between over-grading and any of our mental ill-health outcomes in the full sample of men. Intuitively it seems reasonable that the admission distortion resulting in competition with higher ability peers could induce more stress and anxiety-related disorders if the father is high educated if the expectations on performing well then also are higher. But it is not as straightforward as to why foreign background alters the relationship between over-grading and substance use diagnosis. Substance use is in general lower among adolescents with foreign backgrounds (Hjern & Allebeck, 2004), but could the mechanisms through self-efficacy or university enrolment be stronger among young men with foreign backgrounds? Other studies have also investigated heterogeneity in grading bias consequences on school performance. For example, (Lavy & Sand, 2015) finds that over-grading of boys in primary school had positive impacts on later school achievements among boys with high educated parents and among boys with European or North-American backgrounds, and that this re-ranking between the sexes had negative impacts on achievements among girls with low-educated parents and among girls with Asian or African background. Socioeconomic and migration background thus appears to be important factors to consider in the consequences of grading bias.

In general, grading bias and especially grade inflation is seen as something bad. Systemic grading bias creates welfare costs due to an inefficient allocation of investments in higher education and, on an individual level, grading bias creates unfair labour market opportunities between students. Our findings show that the consequences of systematic grading bias are complex and go beyond the schooling and labour market consequences previously suggested in the literature – over-grading even impacts mental health – which means that grades themselves, independent of knowledge changes, are important for both health and skill formation among students, and therefore impacts life chances. This means that policies that directly or indirectly impact grading in schools may also impact health and life chances among young individuals. This has implications for highly relevant policy questions such as blind/non-blind grading of tests and the choice between standardised or discretionary grading systems. For example, prior to the 90's school reforms in Sweden, the norm-referenced relative grading system was to some extent compensating; grades were often normally distributed within a class (The Swedish National Agency for Education, 2005), which meant that high-performing students were under-graded and low-performing students were over-graded. Recent reports have also shown that grading in Sweden still to some extent reflect the general level of performance at schools, that grading is more restrictive in schools with many high-performing students and conversely less restrictive in schools with many low-performing students (The Swedish National Agency for Education, 2019). Since there is a well-documented socioeconomic gradient in mental health these examples of grading bias would likely reduce mental health inequalities, all other things equal, which raises the question if interventions to prevent systematic grading bias always are desirable. Our results also relate to how we understand and respond to events and shocks. Recent findings show that mental ill-health is rising among the young and the pandemic appears to have exacerbated this trend (Thorisdottir et al., 2021). At the same time, we are facing a situation where distance schooling during the pandemic appears to have induced grade inflation (Karadag, 2021). While these trends individually are of concern, our results imply that the protective effects of over-grading may mitigate some of the challenges students are facing due to the pandemic.

We identify grading bias as the school-level divergence between mean GPA and SweSAT over a period when we know that there was grade inflation in Sweden (Vlachos, 2011; Wikström & Wikström, 2005). We can also see that in our sample both grades and grading bias increase over the study period, and the bias increases relatively more, which is why we assume that what we measure is over-grading. However, one potential shortcoming with this strategy is that we cannot say anything about the absolute levels of grading bias. We show that the relationship differs with level of bias and that the protective impact on mental ill-health is driven by median- to higher-bias levels, where we assume that students are exposed to over-grading. Another potential disadvantage is that we are not able to fully investigate mechanisms through increased higher education enrolment and advantageous labour market outcomes. If grading bias would affect mental health later in life through these mechanism, they are not captured by the sample period of our analysis. The long-term health consequences of over-grading are left to future research.

8. Conclusion

We show that over-grading, receiving a higher grade not reflecting one’s true knowledge or performance, has a non-negligible protective impact on mental health among women. Our findings contribute to a more comprehensive picture of the consequences of grading bias and grade inflation impacts. Grading bias impacts health and skills formation in the years following graduation, but the consequences differ based on sex, socioeconomic and migration background, which means that grading bias has the potential to impact health inequality.

References

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87-99. doi:10.1037/0021-9010.88.1.87

Buckles, K., Hagemann, A., Malamud, O., Morrill, M. S., & Wozniak, A. K. (2016). The effect of college education on mortality. Journal of Health Economics, 50, 99–114.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review, 97*(2), 31-47. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.97.2.31

Dee, T. S., Dobbie, W., Jacob, B. A., & Rockoff, J. (2019). The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations. *American Economic Journal: Applied Economics, 11*(3), 382-423. doi:10.1257/app.20170520

Diamond, R., & Persson, P. (2016). The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. *National Bureau of Economic Research Working Paper Series, No. 22207*. Retrieved from http://www.nber.org/papers/w22207

Gustafsson, J.-E., Allodi Westling, M., Alin Åkerman, B., Eriksson, C., Eriksson, L., Fischbein, S., . . . Gustafsson, J.-E. (2010). *School, Learning and Mental Health : A systematic review*. Stockholm: Kungl. Vetenskapsakademien.

Heckley, G., Nordin, M., & Gerdtham, U. G. (2022). The health returns of attending university for the marginally eligible student. *Health Economics*.

Heckman, J. J. (2012). The developmental origins of health. *Health Econ, 21*(1), 24-29. doi:10.1002/hec.1802

Hinnerich, B. T., Höglin, E., & Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education Review, 30*(4), 682-690. doi:https://doi.org/10.1016/j.econedurev.2011.02.007

Hjern, A., & Allebeck, P. (2004). Alcohol-related disorders in first- and second-generation immigrants in Sweden: a national cohort study. *Addiction, 99*(2), 229-236. doi:https://doi.org/10.1046/j.1360-0443.2003.00562.x

Karadag, E. (2021). Effect of COVID-19 pandemic on grade inflation in higher education in Turkey. *Plos One, 16*(8), e0256688. doi:10.1371/journal.pone.0256688

Lavy, V., & Sand, E. (2015). On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases. *National Bureau of Economic Research Working Paper Series, No. 20909*. doi:10.3386/w20909

Lindahl, E. (2007). Comparing teachers' assessments and national test results – evidence from Sweden. *IFAU - Institute for Evaluation of Labour Market and Education Policy*(2007:24). Retrieved from http://hdl.handle.net/10419/78672

Linder, A., Spika, D., Gerdtham, U.-G., Fritzell, S., & Heckley, G. (2020). Education, immigration and rising mental health inequality in Sweden. *Social Science & Medicine, 264*, 113265. doi:https://doi.org/10.1016/j.socscimed.2020.113265

Lund, C., Brooke-Sumner, C., Baingana, F., Baron, E. C., Breuer, E., Chandra, P., . . . Saxena, S. (2018). Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *The Lancet Psychiatry, 5*(4), 357-369. doi:https://doi.org/10.1016/S2215-0366(18)30060-9

Macintyre, A., Ferris, D., Gonçalves, B., & Quinn, N. (2018). What has economics got to do with it? The impact of socioeconomic factors on mental health and the case for collective action. *Palgrave Communications, 4*(1), 10. doi:10.1057/s41599-018-0063-2

Maurin, E., & McNally, S. (2008). Vive la Révolution! Long-Term Educational Returns of 1968 to the Angry Students. *Journal of Labor Economics, 26*(1), 1-33. doi:10.1086/522071

Müller-Benedict, V., & Gaens, T. (2020). A new explanation for grade inflation: the long-term development of German university grades. *European Journal of Higher Education, 10*(2), 181-201. doi:10.1080/21568235.2020.1718516

Nordin, M., Heckley, G., & Gerdtham, U. (2019). The impact of grade inflation on higher education enrolment and earnings. *Economics of Education Review, 73*, 101936. doi:https://doi.org/10.1016/j.econedurev.2019.101936

Rojstaczer, S., & Healy, C. (2012). Where A Is Ordinary: The Evolution of American College and University Grading, 1940-2009. *Teachers College Record, 114*.

The National Board on Health and Welfare (2020). Statistikdatabasen: Läkemedel. Retrieved from: https://sdb.socialstyrelsen.se/if_lak/val.aspx

Terrier, C. (2015). *Giving a Little Help to Girls? Evidence on Grade Discrimination and its Effect on Students' Achievement*. Retrieved from https://ideas.repec.org/p/cep/cepdps/dp1341.html

The Swedish National Agency for Education. (2005). *National Assessment and Grading in the Swedish School System*. Retrieved from Stockholm: https://www.skolverket.se/publikationsserier/ovrigt-material/2005/national-assessment-and-grading-in-the-swedish-school-system

The Swedish National Agency for Education. (2019). *Analyser av likvärdig betygssättning mellan elevgrupper och skolor*. Retrieved from https://www.skolverket.se/publikationsserier/rapporter/2019/analyser-av-likvardig-betygssattning-mellan-elevgrupper-och-skolor

Thorisdottir, I. E., Asgeirsdottir, B. B., Kristjansson, A. L., Valdimarsdottir, H. B., Jonsdottir Tolgyes, E. M., Sigfusson, J., . . . Halldorsdottir, T. (2021). Depressive symptoms, mental wellbeing, and substance use among adolescents before and during the COVID-19 pandemic in Iceland: a longitudinal, population-based study. *The Lancet Psychiatry, 8*(8), 663-672. doi:https://doi.org/10.1016/S2215-0366(21)00156-5

United Nations Children's Fund. (2021). *The State of the World's Children 2021: On My Mind – Promoting, protecting and caring for children's mental health*. Retrieved from UNICEF: https://www.unicef.org/media/108161/file/SOWC-2021-full-report-English.pdf

Vlachos, J. (2011). *Betygets värde. En analys av hur konkurrens påverkar betygssättningen vid svenska skolor*. Retrieved from Stockholm: Konkurrensverket: https://www.konkurrensverket.se/globalassets/aktuellt/nyheter/betygets-varde.pdf

Wikström, C., & Wikström, M. (2005). Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. *Economics of Education Review, 24*(3), 309-322. doi:https://doi.org/10.1016/j.econedurev.2004.04.010

Appendix

Table A1:

The Impact of Grading Bias on Probability of Mental Ill-Health

	(1)	(2)	(3)	(4)	(5)	(6)
			Mental ill-health (any)			
Full sample						
Grading bias	-0.00899	-0.00372	-0.00434	-0.00939	-0.0101	-0.0101*
	(0.00585)	(0.00609)	(0.00629)	(0.00587)	(0.00614)	(0.00614)
Constant	0.125***	0.125***	0.119***	0.0630***	0.120**	0.126**
	(3.72e-05)	(3.86e-05)	(0.00882)	(0.0115)	(0.0585)	(0.0590)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*	*125,900*
Women						
Grading bias	-0.0230***	-0.0185**	-0.0165*	-0.0195**	-0.0232**	-0.0233**
	(0.00812)	(0.00876)	(0.00908)	(0.00895)	(0.00933)	(0.00940)
Constant	0.156***	0.156***	0.168***	0.135***	0.171*	0.185**
	(0.000191)	(0.000222)	(0.0133)	(0.0207)	(0.0932)	(0.0937)
Observations	*72,272*	*72,272*	*72,272*	*72,272*	*68,927*	*68,927*
Men						
Grading bias	-0.00239	0.00502	0.00345	0.00205	0.00479	0.00484
	(0.00729)	(0.00751)	(0.00781)	(0.00781)	(0.00794)	(0.00795)
Constant	0.0867***	0.0869***	0.0722***	0.0566***	0.125*	0.121
	(0.000105)	(0.000129)	(0.0105)	(0.0154)	(0.0725)	(0.0732)
Observations	*59,569*	*59,569*	*59,569*	*59,569*	*56,973*	*56,973*

Note: This table represents Table 3 in the main text including a column controlling for linear time trends. The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year, additional controls (on school level for each year) are added for each specification with 9th Grade Point Average (GPA) in (2), tracks in (3), sex and migration share in (4), and (on municipal level) school quality indicators in (5) and linear time trends in (6). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

Table A2:

The Impact of Grading Bias on Probability of Psychopharmaceutic Prescription

	(1)	(2)	(3)	(4)	(5)	(6)
			Psychopharmaceutic prescription			
Full sample						
Grading bias	-0.00641	-0.00189	-0.00244	-0.00713	-0.00787	-0.00780
	(0.00571)	(0.00589)	(0.00602)	(0.00561)	(0.00586)	(0.00586)
Constant	0.112***	0.111***	0.103***	0.0506***	0.0859	0.0904
	(3.63e-05)	(3.74e-05)	(0.00824)	(0.0112)	(0.0572)	(0.0575)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*	*125,900*
Women						
Grading bias	-0.0198**	-0.0158*	-0.0139	-0.0169**	-0.0195**	-0.0194**
	(0.00784)	(0.00826)	(0.00846)	(0.00830)	(0.00872)	(0.00879)
Constant	0.142***	0.142***	0.146***	0.113***	0.125	0.131
	(0.000184)	(0.000208)	(0.0129)	(0.0202)	(0.0907)	(0.0912)

Observations	*72,272*	*72,272*	*72,272*	*72,272*	*68,927*	*68,927*
Men						
Grading bias	-2.28e-05	0.00603	0.00499	0.00401	0.00526	0.00532
	(0.00671)	(0.00688)	(0.00711)	(0.00711)	(0.00731)	(0.00729)
Constant	0.0754***	0.0756***	0.0628***	0.0519***	0.108	0.105
	(9.70e-05)	(0.000118)	(0.0104)	(0.0152)	(0.0683)	(0.0687)
Observations	*59,569*	*59,569*	*59,569*	*59,569*	*56,973*	*56,973*

Note: This table shows the full regression output for Table 4 column (1) in the main text. The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year, additional controls (on school level for each year) are added for each specification with 9th Grade Point Average (GPA) in (2), tracks in (3), sex and migration share in (4), and (on municipal level) school quality indicators in (5) and linear time trends in (6). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

Table A3:

The Impact of Grading Bias on Probability of Internalising Disorder Diagnosis

	(1)	(2)	(3)	(4)	(5)	(6)
			Internalising disorder diagnosis			
Full sample						
Grading bias	-0.00914***	-0.00904***	-0.00855**	-0.00995***	-0.0105***	-0.0104***
	(0.00319)	(0.00338)	(0.00339)	(0.00343)	(0.00358)	(0.00356)
Constant	0.0292***	0.0292***	0.0225***	0.00695	0.0218	0.0272
	(2.03e-05)	(2.15e-05)	(0.00401)	(0.00639)	(0.0295)	(0.0295)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*	*125,900*
Women						
Grading bias	-0.0163***	-0.0156***	-0.0143***	-0.0152***	-0.0166***	-0.0168***
	(0.00515)	(0.00521)	(0.00528)	(0.00539)	(0.00550)	(0.00554)
Constant	0.0385***	0.0385***	0.0344***	0.0248**	0.0518	0.0640
	(0.000121)	(0.000128)	(0.00647)	(0.0104)	(0.0455)	(0.0457)
Observations	*72,272*	*72,272*	*72,272*	*72,272*	*68,927*	*68,927*
Men						
Grading bias	-0.00353	-0.00347	-0.00383	-0.00424	-0.00361	-0.00342
	(0.00372)	(0.00399)	(0.00402)	(0.00407)	(0.00420)	(0.00420)
Constant	0.0183***	0.0183***	0.0137***	0.00893	0.00607	0.00247
	(5.38e-05)	(6.66e-05)	(0.00486)	(0.00735)	(0.0352)	(0.0357)
Observations	*59,569*	*59,569*	*59,569*	*59,569*	*56,973*	*56,973*

Note: This table shows the full regression output for Table 4 column (2) in the main text. The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year, additional controls (on school level for each year) are added for each specification with 9th Grade Point Average (GPA) in (2), tracks in (3), sex and migration share in (4), and (on municipal level) school quality indicators in (5) and linear time trends in (6). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1

Table A4:

The Impact of Grading Bias on Probability of Substance Use Disorder Diagnosis

	(1)	(2)	(3)	(4)	(5)	(6)
	Substance use disorder diagnosis					
Full sample						
Grading bias	0.000250	0.00107	0.00123	0.00131	0.00168	0.00144
	(0.00182)	(0.00192)	(0.00197)	(0.00197)	(0.00202)	(0.00203)
Constant	0.00864***	0.00863***	0.00567**	0.00656**	0.00965	0.0100
	(1.16e-05)	(1.22e-05)	(0.00222)	(0.00326)	(0.0139)	(0.0140)
Observations	*131,841*	*131,841*	*131,841*	*131,841*	*125,900*	*125,900*
Women						
Grading bias	0.000201	0.000755	0.000826	0.00124	0.00104	0.000758
	(0.00250)	(0.00269)	(0.00279)	(0.00280)	(0.00295)	(0.00296)
Constant	0.00885***	0.00883***	0.00746**	0.0117**	0.0175	0.0215
	(5.89e-05)	(6.70e-05)	(0.00347)	(0.00499)	(0.0207)	(0.0207)
Observations	*72,272*	*72,272*	*72,272*	*72,272*	*68,927*	*68,927*
Men						
Grading bias	-0.000113	0.000909	0.00112	0.000953	0.00216	0.00209
	(0.00238)	(0.00249)	(0.00253)	(0.00255)	(0.00251)	(0.00252)
Constant	0.00838***	0.00841***	0.00294	0.00142	-0.00204	-0.00328
	(3.44e-05)	(4.03e-05)	(0.00267)	(0.00472)	(0.0218)	(0.0220)
Observations	*59,569*	*59,569*	*59,569*	*59,569*	*56,973*	*56,973*

Note: This table shows the full regression output for Table 4 column (3) in the main text. The results in each column and for each outcome are from separate regressions. All models include school fixed-effects and controls for graduation year, additional controls (on school level for each year) are added for each specification with 9th Grade Point Average (GPA) in (2), tracks in (3), sex and migration share in (4), and (on municipal level) school quality indicators in (5) and linear time trends in (6). Robust standard errors clustered at the school level are shown in the parentheses. ***p<0.01, **p<0.05, *p<0.1