ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-731

Measuring grading standards at high schools: a methodology and an example

David R. Johnson
Wilfrid Laurier University

At schools with low grading standards, students receive higher school-awarded grades across multiple courses than students with the same skills receive at schools with high grading standards. A new methodology shows grading standards vary substantially, certainly enough to affect post-secondary opportunities, across high schools in Alberta. Schools with low grading standards are more likely to be private, rural, offer courses for students returning to high school, have smaller course cohorts, have a smaller percentage of lone parent households and a larger percentage of well-educated parents. Variation in grading standards changes post-secondary opportunities in systematic ways.

VERSION: February 2023

Suggested citation: Johnson, David R.. (2023). Measuring grading standards at high schools: a methodology and an example. (EdWorkingPaper: 23-731). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ak9a-va37

February 2023

Measuring grading standards at high schools: a methodology and an example

David R. Johnson Department of Economics djohnson@wlu.ca
Wilfrid Laurier University
Waterloo, ON, Canada

Abstract:

At schools with low grading standards, students receive higher school-awarded grades across multiple courses than students with the same skills receive at schools with high grading standards. A new methodology shows grading standards vary substantially, certainly enough to affect post-secondary opportunities, across high schools in Alberta. Schools with low grading standards are more likely to be private, rural, offer courses for students returning to high school, have smaller course cohorts, have a smaller percentage of lone parent households and a larger percentage of well-educated parents. Variation in grading standards changes post-secondary opportunities in systematic ways.

*Very useful comments were received at the 2020 Association for Education Finance and Policy meeting and from colleagues at Wilfrid Laurier University, the University of Waterloo, and the Spring 2022 National Bureau of Economic Research Education Meeting.

1. Introduction

Grading standards vary when students with the same skills receive different grades on evaluations that are designed to measure the same skills. When one form of evaluation systematically gives a higher grade than another evaluation, the evaluation awarding higher grades has lower grading standards. In Alberta, Canada, each high school student, in up to ten different courses, is awarded a school-awarded grade between 0 and 100 by the student's teacher and an examination-awarded grade between 0 and 100 on an anonymously graded provincial examination in the same course. In each course at each school in each year, the average, across the cohort of students enrolled in that course in that year, of the school-awarded grade and the average examination grade are reported publicly.

One key contribution of my paper is to derive and estimate a simple linear model of a single parameter for each school, the school grading standards parameter (SGSP) that measures the school's grading standard, high or low.[1] The second contribution of the paper is to show that the variation in grading standards by school, the values of the SGSP, is large and consequential in the lives of students. The third contribution of the paper is to link the estimated values of

[1] The methodology could also be used to evaluate grading standards of teachers with a sample of individual student grades in specific subjects from their teachers and an external anonymous examination grade for the same students in the same subjects. In this case the gradings standards parameter would measure relative grading standards of individual teachers in a subject. In Alberta, as in many jurisdictions, individual student grades are not available to researchers. A small contribution of my paper is to demonstrate measurement of grading standards using school-level data.

SGSP to observable characteristics of schools. Schools with lower grading standards are systematically schools where students already have other life advantages. This observation raises important equity concerns associated with grading in the school system in Alberta.

High school grades are the most important determinant of undergraduate university admission for Alberta students. Appendix A presents illustrative variation in average admission grades and some program cut-off grades for 2022 for the three universities in Alberta continuously open over the years studied. Students with low average high school grades have no chance of admission to the two more desirable universities. High and even very high average grades are required for admission to the more desirable programs in nursing, engineering, or business within any university in Alberta. The high school grades studied in this paper are for very high stakes. The SGSP is the average increase, if positive, in grades awarded by teachers at one school relative to grades awarded by teachers at another school where students with the same skills. Moving between a school one standard deviation below the mean of the various measures of the SGSP to one standard deviation above generates a three-percentage point change in average final grades scaled between zero and one hundred percentage points, the grading scale used for post-secondary admission (see Appendix A). That means it is common that the average final grade across a set of courses at one school could be three percentage points higher (or lower) than at another school where students have the same skills.[2] Such a difference in grades

[2] Grade variation is frequently scaled by the standard deviation of individual student grades when all grades are standardized. The derivation in Appendix C makes it clear why it is more helpful to measure grades in their percentage points from 0 to 100 in this study. Alberta Education reports the standard deviation of individual grades in most courses in most years (excluding

is easily large enough to move you in or out of your desired program or university (see the text below the table in Appendix A). The degree program from which a student graduates plays a significant role in their future income prospects.[3] Other possible consequences of low grading standards are discussed in Gershenson (2018).

2012-13 and 2013-14 for unknown reasons). In the non-excluded years, the range of standard deviation of individual final grades across all students in a course is between 8.5 and 16.1 percentage points in different courses and years. The average of the standard deviations of individual course grades is 12 percentage points. If school-awarded grades in one school are 6 percentage points higher than at another school (a two standard deviation distance in the values of SGSP) and school-awarded grades are 50% of the weight of final grades, then average final grades at the school with lower grading standards in the example above are three percentage points higher across all courses, about 0.25 of a standard deviation of individual final grades, than at the school with higher grading standards. This is a second way to scale the measure of school grading standards. A third way to scale the percentage measures of grading standards considers movement in letter grades. A three-percentage point change in final grades following the usual North American conventions, would frequently move a student within the D, C and B range by one category, from a B- (70-74) to a B (75-77). A- is usually 80-84; A 85-89; A+ 90-100. When the weight on school-awarded grades rises to 70% in the last 4 years of the study, the variation in school-awarded grades increases in importance. All grades studied in this paper can be found on Alberta Education website (https://www.alberta.ca/education.aspx).

[3] Finnie et.al (2019) document the large variation, by undergraduate program, in the earnings of graduates from the University of Ottawa over the years studied in this paper. There is no reason

The third contribution of this paper is to document that variation in the values of SGSP is clearly associated with observable school-level characteristics. Private schools have lower standards. Non-conventional high schools have lower standards. Non-conventional schools enrol mostly students that previously dropped out of high school and are returning to complete a high school diploma.[4] High schools located outside in the two large census metropolitan areas of Calgary and Edmonton, the only large cities in Alberta, have lower standards.[5] Three continuous variables describe schools: the percentage of lone parent households associated with the school; the percentage of parents with a university degree associated with the school and the average size

to think earnings patterns by program would be substantially different in Alberta then in the locations where University of Ottawa graduates live.

[4] The term non-conventional is invented for this paper, not a term used by Alberta Education. Some non-conventional schools are entirely online; others are a mix of online and in-person. They are mostly operated by public school boards. Conventional high schools have similar proportions of Grade 10, 11 and 12 students. In Alberta, high school starts in Grade 10. Non-conventional high schools are defined as having more than 50% of their students in Grade 12. That percentage is usually much higher, as high as 100% in some cases.

[5] In 2016 the Calgary CMA (Census Metropolitan Area) had a population of 1.4 M; the Edmonton CMA had a population of 1.2M and the next largest urban centre, Lethbridge, had a population of only 117,000.

of the diploma course cohorts taught in the school.[6] Schools with larger average cohorts have lower standards. School with a larger percentage of lone parent households have higher standards. Schools with a larger percentage of parents with completed university degrees have lower standards. Since schools with lower standards tend to be private, have a smaller proportion of lone parents and a larger proportion of parents with a completed university degree, students growing up with advantages appear to receive further advantages by being more likely to attend a school with lower grading standards.

Studying the grades from Alberta high schools has advantages and disadvantages relative to previous studies of grading standards in the literature. Comparing grading standards across gender, teacher or schools requires an independent measure of student ability from outside the student's school and the assumption that the teacher's grade and the outside grade measure the same skills.[7] In Alberta one grade is awarded to each student on an outside provincial examination in 10 end-of-course (EOC) "diploma courses." The other grade is assigned by teachers to the same student for mastering the same course-specific material in each course.[8] The measure of grading standards at each school come from observing grades in multiple courses

[6] The variables measuring the percent of lone parent households and the percentage of parents with university degrees are constructed by combining the locations of student homes with census data as described later in the text.

[7] Appendix C manipulates the relation between skills and grades to derive the estimate of SGSP.

[8] Ten courses are used in the analysis. There are 12 diploma courses, but very few schools offer the two French courses. Tyner and Larsen (2019) survey the use of EOC examinations in the United States.

from each school, up to 140 course-year observations at a school.[9] This improves on Gershenson (2018, 2020) and Tyner and Gershenson (2020) whose end-of-course (EOC) external examination is in only one course (Algebra I) in North Carolina. Betts (1998) and Betts and Grogger (2003) also have only an external mathematics grade available for inter-school comparisons. Their external measure of mathematics skills is the score on a national mathematics assessment, an assessment not directly tied to the curriculum in the different mathematics courses at different schools. Hurwitz and Lee (2017) use composite SAT scores as their independent measures of skills and compare the SAT score to an overall high school GPA. My measurement of SGSP is interesting precisely because it extends the measurement of grading standards within the school across a group of courses where, in each course, the intent is that the teacher and the examination are measuring the same subject-specific skill set. Allensworth and Clark (2020) measure grading standards using the variation in post-secondary success across students from different public high school in Chicago where the students have the same average grade (GPA). This is an indirect measure of school grading standards, a higher college graduation rate at the same GPA is interpreted to indicate high grading standards at that school

[9] This is the reason to use the term school grading standards parameter (SGSP); the unit of observation is the group of different courses taught at the school, not a class of students taught by a teacher in a specific course. The groups of students in a subject at a school do not correspond to classes except in a very small high school. Both average grades are reported by academic year and most schools would have classes in a subject in both semesters, sometimes with the same teacher and sometimes with multiple teachers. The methodology applies equally well to a situation with individual student data grouped by teacher.

relative to another school where a student has the same GPA. Another advantage of studying grading standards in Alberta is that the courses in Alberta high schools are taken by older students who clearly understand the stakes.[10] The EOC examinations in Alberta constitute either 50% or 30% of the final grade in courses that are required courses (in different combinations) to graduate high school as well as being the courses whose grades determine admission or non-admission to specific and often highly desired programs.[11]

This paper focuses on the measurement of SGSP across schools over a long time period.[12] There are papers in the literature that measure the change in the difference between the

[10] There are no known stakes relating to examination results for teachers or schools. Funding is provincial and allocated per student to boards. Teacher salaries reflect only seniority within their boards. The likelihood of a teacher being removed from their job for a poor student performance on examinations in a public or Catholic school is effectively zero. Teacher's unions in Alberta are very powerful.

[11] The comparison of teacher predicted grades on final examinations in English schools by Murphy and Wyness (2020) has some similarity to the situation in Alberta. In England, the external examination result is the entire final grade in the course. Murphy and Wyness (2020) do find school-level commonalities across grading variation in predicted grades relative to final grades at the school level but do not stress this result.

[12] Results are identical when the 14-year period is broken up into four separate time periods including a split across the change between a weight of 50% on the external examination in the final grade and a weight of 30% on the external examination in the final grade. These results are available on request.

high school grades and the single external measure of skills over time while sorting high schools by Social and Economic Status (SES) characteristics. Hurwitz and Lee (2017) compare the change in SAT scores to the change in average grades using a group of courses across a large sample of American high schools. Their key result is that there is a large increase in Grade Point Average (GPA) when there is no change in the SAT scores (dynamic grade inflation in their language) and that the largest amount of GPA growth is found in schools that are richer and whiter as well as private (including religious private). Betts and Grogger (2003) are also concerned with the impact of differential grading standards in mathematics at different points in the skill distribution and by race. Gershenson (2018) focuses on changes in differential grading standards when schools are sorted by income and race. Disadvantaged students experienced less of a reduction in grading standards over time in his relatively short sample. My finding that higher grading standards in Alberta are found in schools with more disadvantaged students mirrors these results.

There is a very large literature how variation in grading standards in a cross section of schools does or does not affect disadvantaged students more. Studies finding lower grading standards are applied to lower SES students include Botelho, Madiera and Rangel (2015), Gibbons and Chevalier (2008), Himmler and Schwager (2012) and Rauschenberg (2014). Studies finding higher grading standards applied to lower SES students include Marcenaro-Gutierrez and Vignoles (2015) and Rangvid (2015). Lavy (2018) finds no relation between SES measures of disadvantage and grading standards. Burgess and Greaves (2013), using very detailed English data, find relations between teacher assigned grades relative to examination grades and ethnic characteristics where some ethnic groups receive lower grades and other ethnic groups receive higher grades from teachers' conditional on the student's grade on the blindgraded external examination. My paper finds that higher SES students face lower grading standards in Alberta.

The paper proceeds as follows. The model of school assigned grades conditional on examination grades is explained. Estimation results are presented. An interpretive conclusion follows. There are three appendices.

2. Methodology

The relationship estimated between the average school awarded grade in course j at school k in year t, denoted $\bar{\mathrm{S}}_{jk(t)}$, and the average examination grade in course j at school k in year, denoted $\bar{\mathrm{E}}_{jk(t)}$, is

$$\bar{\mathrm{S}}_{jk(t)} = \mathrm{D}_{j(t)} + \mathrm{P}_{k} + \gamma_j \bar{E}_{jk(t)} + \varepsilon_{jk(t)} \qquad (1)$$

Both observations of average grades are scaled from 0-100 percent. Course fixed effects, $D_{j(t)}$, vary by year and course. Variation in $D_{j(t)}$ across courses j within a year is interpreted as mostly uninteresting variation in scaling of both school and examination assessments across different courses as skills are transformed into grades differently in different courses (equation (1) is derived in Appendix C). Variation in $D_{j(t)}$ within the same course "j" in different years is more controversial. The provincial examinations are scaled to be equal in difficulty from year to year so that variation in the average provincial grade is supposed to measure changes in the average skills of students in the province from year to year. But it is certainly possible, even likely, that the examinations are not calibrated perfectly. This means if students did not vary in quality as a group from year to year, then the same examination grade would correspond to different skill levels by subject between years and $D_{j(t)}$ would vary over time. School awarded grades could also vary in their average from year to year at the same student skill level.[13] These issues are further explored in Appendix C but are not important in this paper beyond requiring time-varying course fixed effects in the estimation of (1). Allowing the slope coefficients γ_j in (1) to vary by course is conceptually important (also see Appendix C). This coefficient allows the relationship between changes in the average school awarded grade and the average examination grade to differ by course but not over time within a course.

The parameter of central interest in this study is each school's fixed effect, P_k. Each estimate of a school's fixed effect is the estimated value of the school's grading standards

[13] This paper does not focus on the changes, if any, in school-level grading standards across time over all courses and schools. See Appendix B for aggregate evidence that average examination grades remained similar over time. There is a relatively small increase in school-awarded grades over time. This is dynamic grade inflation in the language of Gershenson (2018). These changes appear to be only a small part of the story in Alberta in diploma courses. It may be that the presence of the external examination acts to prevent or at least reduce grade inflation in diploma courses. There is some evidence that all school-awarded grades jumped in the four years after 2015-16. That period coincides with the re-weighting of school-awarded grades from 50% to 70% of the final grade. If a change in the level of school-awarded grades occurred at all schools with the reweighting, then the $D_{j(t)}$ terms would have a shared component across all courses within each of those four years or even across the last four years. Further analysis is left for future work. In this this paper, any common change in school-awarded grades or examination-awarded across a course within any academic year is modelled as part of changes in the course-year fixed effects $D_{j(t)}$.

parameter (SGSP). Note there is no time subscript on P_k. The SGSP measures grading standards at each school averaged over all courses taught over the period of the up to 14 years of available results.[14] If at school k, the estimated value of P_k is three, then average school-awarded grades in all courses offered at the school are predicted to be three percentage points higher than the school-awarded grades predicted for that school using the estimates of equation (1) of the year-specific course fixed effects and slope parameters in (1) and using that school's actual average grades on the examinations in each course in that year. A value of P_k of three, because it is positive, shows that school k has low grading standards across all courses. P_k is negative when the school has high grading standards. The average value of SGSP across schools is zero by construction so that the average school has neither low nor high grading standards.

The critical assumption in using Equation (1) to estimate the SGSP is that the school-awarded grade and the examination-awarded grade in each course reward a set of overlapping skills. The common skills included in the construction of the grade in each course have positive, generally different, weights on the school's evaluations and the provincial examination. The school-awarded grade and the examination-awarded grades are not comparable if, for example, a skill is rewarded in the school evaluation and is not rewarded in the examination-based evaluation. This assumption is found elsewhere, if less explicitly, in the literature. Lavy (2008) (page 2084) writes "As long as the two scores are comparable, i.e., as long as they measure the same skills and cognitive achievements," then the statistical relationship between grades on the school-based evaluation and the examination-based evaluation can be estimated. Breda and Ly

[14] Diploma examinations were cancelled due to Covid-19 in 2019-2020. There are no reported public results in the 2020-2021 academic year when examinations were optional.

(2015) (page 55) make a similar comment in noting that in previous comparisons of school-assigned and examination-assigned grades “scores by teachers may reflect both cognitive skills and the assessment of students' behavior in the classroom” and then imply that the comparison of the two grades is not valid in this situation. The derivation in Appendix C adds precision to the relation between average school-awarded and average examination-awarded grades. If the classroom behavior rewarded in the school-awarded grade and classroom behavior has no impact on the examination result, then Breda and Ly (2015) are correct in stating that the grades cannot be compared. However, it is straightforward to think of examples where classroom behavior, for example, arriving to class on time, would affect the examination grade indirectly since the student who arrives on time hears the entire lesson and is present in the class for longer periods of time. Thus, arriving on time, perhaps rewarded directly by teacher grades as a skill, would affect the examination grade even if that behaviour is itself not directly rewarded on the examination. Appendix C makes it clear that the assumption needed to use equation (1) to measure SGSP is that if there are skills like timely arrival, that are rewarded explicitly in the school-awarded grade, having more of that skill will also increase the examination grade indirectly.[15]

[15] It then becomes a conceptual game to decide if there are significant skills rewarded in the school-awarded grade where more of that skill has no impact on the examination-awarded grade. The many papers in the literature comparing school-awarded grades and examination-awarded grades depend on the maintained assumption that external examination grades and school grades reward an overlapping set of skills.

The last methodological point to emphasize from Appendix C is that even though it is entirely possible, and even likely, that parents (or students) would seek out and choose to attend schools with low grading standards, that sorting does not affect the estimates of the grading standards parameter within each specific group of years. The intuition: when the school-awarded grade and the examination-awarded grade linearly reward a common set of skills (Appendix C), the examination grade is a sufficient control variable that allows accurate estimates of the grading-standard parameters, the values of P_k from a group of years. It could be that once a school with low grading standards is known to exist, parents or students who want low grading standards flock to that school and having more parents with such preferences might lower further lower grading standards in a subsequent time-period. This case is not studied here beyond some robustness checks by breaking the 14 years into sub-samples.[16] Equation (1) allows for the case where the grading standards parameter is the result of like-minded parents wanting a school with low grading standards. The obvious example is a group of parents forming their own private school with low grading standards. Equation (1) simply estimates the grading standards parameter at that private school.

[16] Equation (1) was also estimated with 4 different sub-samples. The SGSP from the same school are strongly autocorrelated across the sub-samples, a very interesting result. These estimates can be provided. These dynamic issues are not addressed in the current paper: do schools with low grading standards attract more students over time; do schools lower grading standards to attract more students; do new schools with lower grading standards open in areas where other schools have high grading standards.

Equation (1) is estimated using the techniques created by Abowd, Kramarz and Margolis (1999) – henceforth AKM. The AKM model estimates a set of course fixed effects, the values of $\widehat{\mathrm{D}}_{j(t)}$, that vary by both course and year. The AKM model also estimates the slope coefficients estimating the values of $\hat{\gamma}_j$: the change in the average school-awarded grade per unit change in the average examination-awarded grade which vary by course. Finally, AKM estimates a set of school fixed effects, P_k, that are the values of the SGSP parameters. AKM estimates robust standard errors of the SGSP parameters and the slope coefficients.[17] In AKM the SGSP parameters are normalized to have an average value of zero across schools. If $\mathrm{P}_k = 3$ then the average school-awarded grade in all courses at school k is estimated to be three percentage points higher than the school-awarded grades predicted by the school's average examination grade on those courses. The school with a value of $\mathrm{P}_k=3$ has low grading standards.

To discover if school-level characteristics are systematically associated with larger and smaller values of P_k, the estimated values of P_k are regressed on school characteristics using[18]

$$\mathrm{P}_k = a_0 + a_1 I_k + a_2 SES_k + a_3 C_k + \varepsilon_k \qquad (2)$$

---

[17] The routine in Stata is felsdvregdm.

[18] To keep the notation less intensive, variation between the true value of P_k and the estimated value of P_k is incorporated into ε_k in expression (2). Equation (2) is similar to regressions used in the AKM literature for explanations of productivity that vary by workers across characteristics of firms. Dhuey and Smith (2018) estimate a similar regression to explain school fixed effects that vary with the tenure of specific school principals.

I_k is a series of indicator variables active when school k is a specific type of school: an unconventional school, a private school or an urban school. SES_k represents two continuous SES variables. One is a measure of the percentage of all households associated with the school that are lone parent households. The other is a measure of the percentage of adults associated with the school with a completed university degree. C_k is a measure of the average number of enrollees in diploma course cohorts in the school over the period studied. The error terms in (2) are treated as heteroscedastic and robust standard errors are estimated on the coefficients on the descriptive and indicator variables as a way of describing schools with characteristics associated with predictably higher or lower grading standards? [19]

---

[19] It is also possible to substitute equation (2) directly into equation (1) and estimate

$$\bar{S}_{jk(t)} = a_0 + D_{j(t)} + \gamma_j \; \bar{E}_{ijk(t)} + a_1 I_k + a_2 \, SES_k + a_3 \, C_k + \varepsilon_{jk(t)}$$

The course-specific time-varying constant term, the sum of a_0 and $D_{j(t)}$, is not of central interest. One advantage of the equation above is that, unlike the AKM estimator, the equation can be estimated with and without weights following the suggestions of Solon and Woodridge (2015). Standard errors can be generated clustering the observations in different ways. One difference between the equation in the footnote and equation (2) in the text is that equation (2) in the text uses one observation of SGSP per school from the first step in estimating equation (1), implicitly weighting all schools equally in the estimates of a_1, a_2 and a_3. In the footnote equation, schools vary by the number of examinations in the group of years included and schools with more examinations receive a larger weight in the estimates of a_1, a_2 and a_3. Estimates of a_1, a_2 and a_3 using this methodology are available on request. In my opinion, the two-step methodology in the

3. Results

Table 1 presents descriptive data on all 533 high schools with diploma course results operating in Alberta between 2005-06 and 2018-19. Many schools (certainly not all) operate in all years.[20] Conventional schools are high schools with roughly equal enrolment in Grades 10, 11 and 12. High school students in Alberta enter a conventional high school in Grade 10 and leave after obtaining both sufficient credits and the right combination of courses to graduate, usually, but not always, after three years.[21] Conventional schools educate most students. The remaining students are found in non-conventional schools, a concept already introduced. In these schools all or a very large proportion of their students are in Grade 12. These schools meet the needs of students who have dropped out and then return to complete high school courses. Table 1 shows there are many non-conventional schools, but their total enrolment is small. Table 1 also

---

paper is the preferred methodology since the unit of observation of interest in equation (2) is the school. The estimates of a_1, a_2 and a_3 are very similar across the two methodologies.

[20] Results from courses at a school with less than 6 students enrolled in a year are not publicly reported for privacy reasons, all other school-level results are reported. There is considerable population growth in Alberta over this period so new schools are opened every year.

[21] There is a significant drop out rate from Alberta high schools. There are also students who remain in conventional high schools for a fourth year and are counted as Grade 12 students. Such students remain in high school for an extra year to obtain additional credits, to increase their average grade and have better access to post-secondary programs or for social or athletic reasons.

shows a relatively large (for Canada) private school sector in Alberta.[22] In the last column of Table 1, the counts of schools of each type that have results for 10 or more courses are presented. Separate estimates of the SGSP values by school are presented for the schools with more than 10 courses as a robustness exercise. These estimates might be considered to have more meaningful standard errors because there are at least 10 courses uses to estimate the SGSP at every school.

The second last and third last columns of Table 1 present averages of examination and school awarded grades from the different types of schools. Averages of school-awarded grades are much higher than the average grades awarded on anonymously graded provincial examinations in every setting.[23] Average examination grades are much higher at private schools than at non-private schools. This is not surprising when the social and economic advantages of students attending private schools are presented later in the paper. Specific parental types select into private schools, but as noted earlier, this is not a problem in the unbiased estimation of the values of the SGSPs.

Table 2 presents AKM estimates of equation (1) using the three groupings introduced in Table 1. The estimates of the course-specific slope coefficients γ_j from equation (1) are

[22] In Alberta private schools take fees from parents and receive a lower provincial grant than public schools per student taught. It is unclear whether resources per student are higher or lower at private schools than public schools. The Catholic school system in Alberta receives full public funding, does not charge fees, and is not part of the private school system. I exclude the very small and fully publicly funded francophone system from this study.

[23] This is consistent with other literature comparing school-awarded grades to examination-awarded grades.

presented using all schools from three different groupings of schools and courses.[24] The slope parameters are precisely estimated and clearly different across courses. An increase in the average grade awarded to a cohort of students at a school on the provincial examination predicts an increase in the average grade school-awarded grade in the same cohort. It would be very surprising (perhaps even shocking) if such relationships did not exist. The provincial examination is based on the course's provincial curriculum in each subject. Teachers are supposed to teach and assess the provincial curriculum as tested on the examination. Table 2 introduces the concept of "university courses." The six "university" courses are the courses more closely associated with admission to the most competitive university post-secondary programs.[25]

---

[24] The estimates of the slope coefficients do not change if the schools with less than 10 courses are excluded. The excluded schools are a very small proportion of the total number of observations.

[25] The requirements to graduate from high school are credits in one of the two English courses, 30-1 or 30-2; one of the two mathematics courses 30-1 or 30-2; one of the two social studies courses 30-1 or 30-2 and one of the four science courses. However, admission to more competitive university programs generally require the 30-1 versions of English, Math and Social Studies courses as well as one or more of the three named sciences rather than the more general Science 30 course. There are 12 diploma courses in Alberta. There is a diploma course in the French Language, that is French for non-francophones, and a diploma course in French for francophones. Both French courses are taught at very few schools and are excluded from the analysis. The term "university" course is not a formal term used by Alberta Education. That term was created for this paper.

It is conceivable that behaviour around awarding school grades in the "university courses" could be different than in the other courses where the stakes for students relate more to completing high school rather than university admission. This is not the case.

The estimates of the SGSP parameters, the measures of grading standards at a school, are the parameters of central interest. Recall that an estimate of a school's SGSP of plus (minus) three says that average prediction of the school-awarded grade in all courses at that school is three percentage points higher (lower) than at another school with the same examination-awarded grade in each course in the group. Table 2 shows that for the set of schools including the schools with less than 10 course-year observations, the standard deviation of estimated values of the SGSPs is about three percentage points. Figure 1 (upper left) presents histograms of the SGSP estimates using all schools and all courses and then histograms that separate out conventional schools, private schools and rural schools, discussed later. The histogram in the upper left corner shows that for all schools, values of SGSP vary from -10 to +10 percentage points. As already noted in the introduction, these values are of practical importance to the lives of students. Such values move average letter grades across multiple letter grade categories. The changes in average final grades would certainly move students across the post-secondary cut offs presented in Appendix A and discussed in the introduction. The count of schools is on the vertical axes of Figure 1. The third last column of Table 2 reports more than half the estimated values of SGSP effects have robust t-statistics larger in absolute value than 2 in the AKM framework. The last two columns of Table 2 present the standard deviation of and the proportion of the estimates of SGSPs that are statistically different from zero when the sample is restricted to schools with more than 10 course-year observations. The SGSP estimates are less variable in the restricted sample but for the all-school and conventional school cases, a slightly higher percentage of the values of SGSPs are statistically different from zero.

It is easy to imagine a situation where most of the estimates of the SGSP parameters are small and are of no practical consequence.[26] One might even expect this to be the case. There are up to 140 courses at the same school in the 14-year period. Many different teachers would contribute to the school-awarded grades in such a situation. You would expect that within a school or even within a subject at a school, some teachers have low grading standards, and some teachers have high grading standards. The fact that many SGSP estimates are large and statistically different from zero over many groupings of courses and schools is strong evidence that low and high grading standards at Alberta high schools are shared by teachers within each school across subjects. The second key contribution of this study is the clear evidence that schools have common grading standards across groups of courses and over multiple years, a school "grading culture" is a possible term for this phenomenon.

Before moving to discussion of the estimates of equation (2), Table 3 summarizes the variables are used to describe characteristics of schools. Two school characteristics were previously presented in Table 1: 153 schools are non-conventional, and 48 schools are both conventional and private. Schools are also described by two continuous variables derived from the 2006 and 2016 Canadian censuses. For schools operating between 2005-06 and 2014-15 the

[26] The histogram(s) in Figure 1 would have schools bunched, for example, within one percentage point of zero. Estimates of SGSP parameters could still be statistically significant, just not of much practical consequence.

percentage of lone parent households in each school community and the percentage of adults with a completed university degree in each school community are calculated using lists of the number of students at each school in each year residing in each census dissemination area (DA) of the 2006 census.[27] Student locations by DA using the 2016 census are used for the same calculations for the years from 2015-16 to 2018-19.[28] The DA is the smallest geographic census unit reporting data, roughly 700 people normally live in a DA. The SES variables for the school in each year are weighted averages of DA characteristics, weighted by the number of school attendees from each DA at each school over all years. Then Table 3 describes the census-based variables in two ways. One measure uses each school as the unit of observation and presents statistics on means and standard deviations of the school-unit means. The second measure, instead of taking simple averages across the school unit observations, each school observation is weighted by the number of students enrolled at each school in each year. This calculation describes the population sending students to high schools in the two census time periods. Two facts emerge from this analysis. First, larger schools have more highly educated parents. Second, the level of education of the average adult increased substantially in Alberta between the 2006 and 2016 measure. The second fact makes it necessary to construct the SES variables to describe each school as follows. At each school in each year, the school's percent of adults with a

[27] Counts of persons within a DA with and without completed university degrees is the measure of education that is consistent between the 2006 and 2016 census.

[28] These are all students enrolled at the school as provided by Alberta Education, not the students enrolled in the diploma courses. Student locations by course or grades are not available. A student is counted in each year of their attendance at the school.

completed university degree (or percent of lone parent households) was subtracted from the school population percent of adults with a university degree (percent of lone parent households) sending students to all high schools.[29] Thus the continuous variables in equation (2) are expressed in percentage points of adults with a completed university degree (percentage points of single parent households) as deviations from population means of the relevant census. Then these variables are averaged across the years the school is operating among the 14 years. A positive (negative) value is a school with more (less) educated parents or more (less) lone parent households. The final note from Table 3 is that the percentage of parents with completed university degrees at private schools is much higher than at public schools while the percent of lone parent households is only slightly different at private schools than public schools.[30] Table 3 shows that 66% of all schools are rural and 40% of private schools are rural. Finally, Table 3 shows that the average size of a diploma course varies widely. While the average size of a diploma course at all schools or at conventional schools is between 39 and 46 students, the standard deviation of the average size of a diploma school-level cohort is larger than the mean.

---

[29] Nearly identical variables take the school mean of the variable, subtract the provincial mean and divide by the provincial standard deviation. Results using this method of joining the data measured across the two censuses yielded identical results.

[30] Recall Table 1 indicated much higher average examination results at private schools then at public schools. Part of the explanation is the stronger educational background, on average, of private school parents. Appendix C makes clear how parental background is appropriately taken into account in the unbiased estimates of the SGSPs.

There are a few schools with very large cohorts of diploma students.[31] In contrast, private schools have very small diploma cohorts and much less variation in average cohort size compared to non-private schools.

Table 4 presents estimates of equation (2) for each school grouping. There are two parts to the table, the lower part presents estimated coefficients from the samples excluding schools with less than 10 course-year observations where it is reasonable to think the estimates of the values of SGSP are more reliable. All estimates of equation (2) present similar relations between school characteristics and the estimated values of the SGSPs. The coefficient on the non-conventional school indicator variable is positive and significant with a value of about two. Non-conventional schools have lower grading standards than other schools. At the same examination grades and thus the same skill levels, the extra grades added by teachers average two percentage points higher on diploma courses at a non-conventional school than would be awarded at a conventional school. This effect can be seen clearly in the upper right panel of Figure 1 where the open bars represent non-conventional schools. The mass of bars representing non-conventional schools clearly fall to the right of zero.

Grades assigned by teachers at private schools average 2 to 3 percentage points higher across all courses than at non-private schools when students have the same skills as measured by the same examination results. The slightly smaller coefficient on the private school indicator variable is found in the lower panel of Table 4 where the sample is restricted to include schools

[31] Four very large non-conventional schools are operated by public and Catholic school boards in each of Calgary and Edmonton to allow students who have dropped out to obtain their credits toward graduation.

with 10 or more courses. The private school effect is readily observed in lower left panel of Figure 1 where estimates of the SGSPs at private schools are represented by the open bars. Private schools have lower grading standards.

Finally, average additional school awarded grades at schools outside Alberta's two urban areas with more than one million in population, where the rural indicator takes a value of one, are about one percentage point higher than grades at the urban schools when two schools have the same grades on the examinations. This effect is illustrated in the lower right panel of Figure 1 where the mass of rural schools, the open bars, lie to the right of zero. Rural schools have lower grading standards.

The three remaining estimated coefficients in Table 4 are the coefficients on continuous variables describing school characteristics. The first continuous variable is the average cohort size of all diploma courses offered at the school over the years in each group.[32] If the average size of a diploma course cohort offered at a school is larger, then the estimate of the school's SGSP is predicted to be smaller. Using a coefficient estimate on average course cohort of - 0.006, then reducing the average cohort size at a school by 50 students (roughly one standard deviation of cohort sizes in Table 3), predicts an increase in the estimate of SGSP of 0.3

[32] Cohort size is related to school size. However, if average total enrolment at the school is used in equation (2) in place of average enrolment in diploma courses, its estimated coefficient remains negative but is less frequently statistically significant. If both an enrolment variable and a diploma course variable are used, neither is significant. There is collinearity between the two variables. Information on class size is not available.

percentage points. Schools with smaller diploma course cohorts have slightly lower grading standards.

The coefficient on the higher education measure is roughly 0.05. Using the 0.05 value as an example, if the percentage of parents with a university degree at one school is larger than at another school by 10 percentage points, roughly one standard deviation of the variation in the percent of adults with completed university degrees in the school population average, the predicted value of the estimated SGSP increases average school awarded grades by 0.5 percentage points. Schools with a larger percentage of better educated parents are predicted to have lower grading standards.

The last continuous variable measures, in percentage points, the deviation of the school's average percentage of lone parent households from the population average. The coefficient on this variable is negative, a larger percentage of lone parent households predicts a reduction in school-awarded grades at the same skill level of students. The coefficient is larger when non-conventional schools are excluded from the sample, even larger when the schools with 9 or less diploma courses are excluded from the sample; and even larger when only university courses are in the sample. The last entry in the relevant column of coefficients in Table 4 is -0.13. Using a value of -0.10 as the example, if the percentage of lone parent households rises by 10 percentage points, the value of SGSP is predicted to be roughly 1.0 percentage point lower. Schools with a higher percentage of lone parent households have higher grading standards. A reasonable amount of the variation in SGSP is associated with the five or six observable characteristics of schools used as right-hand side variables in equation (2); the R^2 values range from 0.23 to 0.31. The larger values are in the estimates that exclude the non-conventional schools. This difference makes sense since the census variables are almost certainly more accurate representations of school characteristics for conventional schools where a larger proportion of students would live in the parental home.

4. Conclusions and Implications

This paper measures school-wide grading standards, that is, variation in grading standards common across all courses at a school over a multi-year period. In a school with low grading standards, the average school-awarded grade across courses is predicted to be higher (lower) than in another school with the same average examination-awarded grades. The systematic variation in school grading standards is easily large enough that it could affect both admission to high demand university programs. This affects the equity of admission to post-secondary programs. Individual student data would be needed to understand how differences in school grading standards are distributed across the grade distribution of students and how variation in grading standards might affect high school graduation. Individual student data are not made available for research in this province.

Descriptive characteristics of a school predict whether a school has low or high grading standards. Non-conventional schools, schools addressing the needs of students who previously dropped out and then return, are more likely to have low grading standards. This implies that students completing high school at non-conventional schools may be with lower skills than for students who complete high school in conventional schools. Rural schools are more likely to have low grading standards. Schools where the average cohort of all diploma courses is smaller are more likely to have low grading standards. Private schools, where parents pay fees, are more likely to have low grading standards. Schools with a larger percentage of parents with a university degree or a smaller percentage of lone parents have lower grading standards. These findings, particularly the latter three findings, suggest that students who already come to high school with some advantages in life are more likely to attend schools that have lower grading standards. This raises significant issues of equity across students. Some argue external or standardized examinations work against good outcomes for disadvantaged students. The results in this paper suggest a countervailing mechanism, the finding that school-awarded grading practices favour more advantaged students relative to disadvantaged students with the same examination grades. This is an issue for future study with individual student observations.

The measurement of school grading standards in this paper highlights a crucial maintained assumption in this study and other studies comparing school-based and examination-based grades: that the curriculum-based provincial examination and the school-awarded grades in each course measure a common set of skills with different course-specific weights on each common skill. The weights on the skills do not need to be identical for the school-awarded and examination-awarded grades. One advantage of the high school environment in Alberta is that teachers are supposed to teach and assess the provincial curriculum in the 10 courses used in this paper to measure school grading standards. Teachers know the content of the examination and of the curriculum. The common skills assumption seems more plausible in this setting than, for example, when the comparator examination for measuring school grading standards is an aptitude test or a very general standardized test in literacy or numeracy not tied to the curriculum.

It is possible that school-awarded grades measure significantly different skills than examination-assigned grades and these different skills are also similar enough to be rewarded in multiple courses. Higher school-awarded grades at the same examination-awarded grades, which I interpret as low grading standards, could thus be systematically different across schools and reward the significant skills mentioned above. Those skills, whether learned at school or at home, must be systematically different across schools and have no effect on examination-awarded grades. If this is your belief, then the results of this paper suggest that the additional skills measured by the school-assigned grade and not measured by the examination grade are systematically higher in private schools, at non-conventional schools, at rural schools, at schools with a smaller percentage of lone parent households, at schools with smaller diploma class sizes and at schools with a larger percentage of parents with completed university degrees. The systematic differences in school assigned grades associated with those school characteristics are then legitimate rewards for skills, they reward higher non-examination skills. An important question not answered in this paper is whether, after controlling for the portion of the value of the school's SGSP explained by social and economic factors, the residual SGSP has explanatory power for other measures of success. This would require individual observations linked to success measures not available in Alberta. Comparing success of two students with the same social and economic background, one with a high value of SGSP and one with a low value of SGSP, could identify whether the high SGSP student has acquired higher skills. Possible measures of success could be post-secondary grades, post-secondary graduation, employment, income, or periods on social assistance. Such a project could resolve whether the differences between school-awarded grades and examination-awarded grades reward different skills or are the result of variation in school grading standards, the preferred interpretation in this paper.

Here are some suggestive examples of how to interpret the results of this paper as variation in school grading standards. These are speculative narratives: perhaps when the class is smaller, teachers find it harder to give lower grades to students they likely know better; perhaps in rural settings, where teachers know families directly (and may teach their own children or the children of other teachers they know), it is harder for teachers to award lower grades; perhaps better educated parents and their children are more aware of the importance of high grades and lobby their teachers more effectively to raise school-awarded grades than do less educated parents; perhaps better educated parents congregate at school with lower grading standards due to a higher awareness of the importance of high school grades; perhaps teachers, as very well educated upper middle class persons, give higher grades to students more like themselves;[33] perhaps single parents have less time to lobby teachers for higher grades; perhaps parents are simply paying for higher school-awarded grades at private schools due to increased bargaining power in discussions of school-awarded grades with private school teachers and principals; perhaps the non-conventional schools focus on getting returning students to get their high school diploma and thus lower grading standards to facilitate that process. These simpler explanations are consistent with the observed variation in school grading standards. These interpretations raise important equity considerations because, in these interpretations, students in schools populated by students with more life advantages, tend to receive further life advantages in attending schools with lower grading standards.

[33] Teachers in Alberta earned at the 80th percentile of all employment earnings in Alberta in 2013/14, see Johnson (2015). Virtually all teachers have a completed university degree.

References

Abowd, John M., Francis Kramarz and David N. Margolis (1999) "High Wage Workers and High Wage Firms." Econometrica 67(2): 251-333.

Allensworth, Elaine M. and Kallie Clark (2020) "High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools" Education Research 49 (3): 198-211.

Betts, Julian R. (1998) "The Impact of Educational Standards on the Level and Distribution of Earnings" American Economic Review 88 (1): 266-275

Betts, Julian R. and Jeff Grogger (2003) "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings," Economics of Education Review 22 (4): 343–52.

Botelho, Fernando, Ricardo A. Madeira and Marcos A. Rangel (2015) "Racial Discrimination in Grading: Evidence from Brazil." American Economic Journal: Applied Economics 7(4):37-52

Burgess, Simon and Ellen Greaves (2013) "Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities." Journal of Labor Economics 31(3): 535-576.

Dhuey, Elizabeth and Justin Smith (2019) "How School Principals Influence Student Learning?" Empirical Economics 54(2):851-882.

David N. Figlio and Maurice E. Lucas (2004) "Do High Grading Standards Affect Student Performance?" Journal of Public Economics 88, No. 9–10:1815–34.

Finnie, Ross and Stephen Childs, Dejan Pavlic, Nemanjic Jevtovic (2019) “How Much do University Graduates Earn?” EPRI #Grad Earnings RESEARCH BRIEF #3 (VERSIONB,14-11-21)

Seth Gershenson (2018) “Grade Inflation in High Schools (2005–2016)” Washington, D.C.: Thomas B. Fordham Institute.

Gershenson, Seth (2020) “Great Expectations: The Impact of Rigorous Grading Practices on Student Achievement” Washington, DC: Thomas B. Fordham Institute (February 2020). https://fordhaminstitute.org/national/research/great-expectations-impact-rigorous-grading-practices-student-achievement

Gibbons, Stephen and Arnaud Chevalier (2008) “Assessment and age 16+ education participation.” Research Papers in Education 23 (2):113-123.

Godfrey, Kelly (2011) “Investigating Grade Inflation and Non-Equivalence.” College Board Research Report 2011-2.

Himmler, Oliver and Robert Schwager (2012) “Double Standards in Educational Standards – Do Schools with a Disadvantaged Student Body Grade More Leniently?” German Economic Review 14(2):166-189.

Hurwitz, Michael and Jason Lee (2017), “Grade Inflation and the Role of Standardized Testing,” in Measuring Success: Testing, Grades, and the Future of College Admissions, ed. Jack Buckley, Lynn Letukas, and Ben Wildavsky, Baltimore, MD: Johns Hopkins University Press, 64–93.

Johnson, David R. (2005) "Signposts of Success: Interpreting Ontario's Elementary School Test Scores." Toronto: C.D. Howe Institute.

Johnson, David R. (2015) "Value for Money? Teacher Compensation and Student Outcomes in Canada's Six largest Provinces" Toronto: C.D. Howe Institute Commentary No. 434. https://www.cdhowe.org/sites/default/files/attachments/research_papers/mixed/Commentary_434_0.pdf

Lavy, Victor (2018) "On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases." Journal of Public Economics 167: 263-279.

Marcenaro-Gutierrez, Oscar and Anna Vignoles (2014) "A comparison of teacher and test-based assessment for Spanish primary and secondary students." Education Research 57(1): 1-21.

Murphy, Richard and Gill Wyness (2020) "Minority Report: the impact of predicted grades on university admissions of disadvantaged groups." Education Economics 28(4):333-350.

Protivinsky, Tomas and Daniel Munich (2018) "Gender bias in teachers' grading: What is in the grade?" Studies in Educational Evaluation 59: 141-149.

Rangvid, Beatrice Schindler (2015) "Systematic Differences across evaluation schemes and educational choice." Economics of Education Review 48: 41-55

Rangvid, Beatrice Schindler (2019) "Gender Discrimination in Exam Grading" Double Evidence from a Natural Experiment and a Field Experiment." B.E. Journal of Economic Analysis and Policy

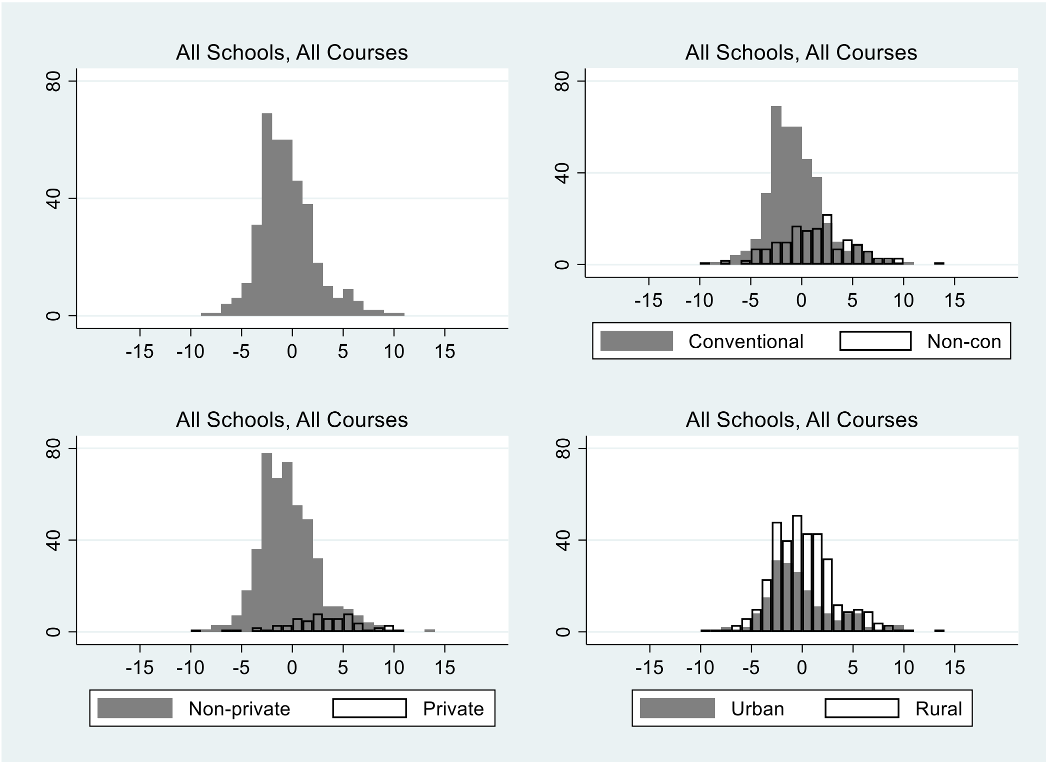
Rauschenberg, S. (2014) "How consistent are course grades? An examination of differential grading." Education Policy Analysis Archives, *22*(92) http://dx.doi.org/10.14507/epaa.v22n92.2014

Solon, Gary, Stephen J. Haider and Jeffrey M. Wooldridge (2015) "What are we weighting for?" Journal of Human Resources 50(2): 301-316.

Tyner, Adam and Seth Gershenson (2020) "Conceptualizing Grade Inflation" Economics of Education Review 78: 102037.

Tyner, Adam and Matthew Larsen (2017) End-of-Course Exams and Student Outcomes. Washington, DC: Thomas B. Fordham Institute (August 27, 2019). ttps://fordhaminstitute.org/national/research/end-course-exams-and-student-outcomes.

Figure 1: Histograms of the values of SGSP: the average increases or decreases in school-awarded grades in excess of grades warranted by student skills at all schools



Notes: Vertical axes measure the number of schools. Horizontal axes measure the values of the SGSPs estimated in the paper. The SGSP is the change in the average grade in all courses at a school controlling for the examination grade in the course. These values are increases (decreases) in average school-awarded grades in excess of grades warranted by student skills. Conventional schools offer all three grades; non-conventional schools (Non-con) have mostly Grade 12 students. Rural schools are outside the Calgary and Edmonton Census Metropolitan Areas. Private schools receive fees per student.

Table 1: School Assigned and Examination assigned grades 2005-06 to 2018-2019

Type of Schools	Number of Schools	Number of Diploma Courses	Total Enrolment in Diploma Courses	Examination Average Grades[1]	School Assigned Average Grade[1]	Number of schools that offer 10 or more courses
All Schools	533	40,934	2,389,923	65.8	72.0	461
Conventional Schools	380	35,734	2,116,788	66.4	72.2	357
Conventional Private Schools	48	2,469	51,333	70.7	78.1	39

Source: Author's calculations. 1. These are averages of grades weighted by course enrolment.

Table 2: Estimates of an AKM model of the relation between school-awarded grades and examination-awarded grades

Type of schools Courses[1]	Slope parameters measuring the relationship between the school assigned average grade and the examination assigned average grade (standard error)										SGSP estimates using all schools: no restriction on number of course-year observations		SGSP estimates using schools with more than ten course-year observations	
	English 30-1	Math 30-1	Social Studies 30-1	Phy-sics	Chem -istry	Biol-ogy	English 30-2	Math 30-2	Social Studies 30-2	Science 30	Standard deviation (number of schools)	% with t-stats > |2.0|	Standard deviation (number of schools)	% with t-stats > |2.0|
All, All	.29 (.01)	0.26 (.006)	.28 (.008)	.31 (.008)	.29 (.007)	.36 (.007)	.27 (.008)	.27 (.008)	.33 (.009)	.39 (.01)	3.2 (533)	66.4	2.69 (461)	71.8
Con , All	.24 (.01)	.25 (.006)	.25 (.009)	.30 (.008)	.29 (.007)	.38 (.008)	.24 (.01)	.28 (.009)	.34 (.01)	.39 (.01)	2.7 (381)	69.8	2.52 (358)	72.3
Con , Univ	.25 (.01)	.25 (.006)	.27 (.009)	.31 (.007)	.30 (.007)	.39 (.008)	NA	NA	NA	NA	2.9 (266)	70.3	2.61 (231)	65.8

1. The first "All" indicates all schools are included; Con indicates only conventional schools are included. The second "All" specifies that all non-French courses are included. "Univ" specifies that only the 6 courses more important for university admission are included. Non-university courses are the courses labelled NA.

Table 3: Characteristics of Schools in Alberta

Characteristic	Unit of observation	All Schools Mean (S.D.) ((number of units))	Conventional Schools Mean (S.D.) ((number of units))	Private Conventional Schools Mean (S.D.)((number of units))
Percent of adults associated with school with a completed university degree 2005-2006 to 2014-2015 Measured using 2006 census	Schools	13.01 (9.27) (493 schools)	13.38 (10.00) (365 schools)	21.32 (13.08) (44 schools)
	Weighted by enrolment	16.35 (9.75) (225,431 students)	16.44 (10.20) (192,904 students)	22.87 (13.62) (11,987 students)
Percent of adults associated with school with a completed university degree 2015-2016 to 2018-2019 Measured using 2016 census	Schools	16.78 (10.30) (486 schools)	17.19 (11.23) (355 schools)	28.32 (14.90) ((39 schools)
	Weighted by enrolment	21.01 (10.7) (228,160 students)	21.19 (11.43) (184,291 students)	31.13 (15.30) (12,157 students)
Percent of households that are lone parent 2005-2006 to 2014-2015 Measured using 2006 census	Schools	23.32 (9.21) (493 schools)	22.62 (8.74) (365 schools)	22.11 (8.92) (44 schools)
	Weighted by enrolment	24.41 (5.77) (228,160 students)	24.35 (7.39) (192,904 students)	22.61 (7.66) (11,987 students)
Percent of households that are lone parent 2015-2016 to 2018-2019 Measured using 2016 census	Schools	25.57 (6.83) (486 schools)	24.78 (6.92) (355 schools)	22.11 (8.92) (44 schools)
	Weighted by enrolment	24.47 (5.77) (228,160 students)	24.25 (6.02) (184,291 students)	19.07 (6.28) (12,157 students)
Percent of schools that are rural	Schools	66.0	66.6	40.8
Size of diploma cohort	Schools	38.80 (57.21)	46.03 (58.02)	15.77 (13.74)

Source: Author's calculations.

Table 4: Relationships between estimated values of School Grading Standards Parameter (SGSP) and School Characteristics

Grouping Schools, Courses	Number of Schools	Non-conventional school indicator	Private school indicator	Rural indicator	Average course size at school	Lone parent measure	Higher education measure	R^2
All, all	533	2.02 (.35)*	2.86 (.58)*	1.16 (.31)*	-0.007 (.002)*	-0.048 (.025)	0.054 (.019)*	0.23
Con, all	381	NA	2.72 (.53)*	0.98 (.31)*	-0.006 (.002)*	-0.095 (.02)*	0.052 (.017)*	0.28
Con, Univ	266	NA	2.78 (.61)*	1.01 (.38)*	-0.008 (.002)*	-0.089 (.034)*	0.045 (.020)*	0.28
	Estimates from samples restricted to schools with more than 10 courses taught at the school							
All, all	461	1.77 (.36)*	2.06 (.68)*	1.02 (.27)*	-0.006 (.002)*	-.083 (.022)*	0.063 (.019)*	0.25
Con, all	357	NA	2.21 (.49)*	0.85 (.27)*	-0.005 (.002)*	-0.117 (.016)*	0.059 (.015)*	0.31
Con, Univ	230	NA	2.05 (.54)*	0.88 (.35)*	-0.006 (.002)*	-0.130 (.024)*	0.048 (.019)*	0.31

Robust standard errors in parentheses. * indicates coefficient is statistically different from zero at 5 percent

Appendix A Table 1: High school grades at entry in three Alberta universities

Fall of Year	Average Entering Grade "University of			Percent of entrants with average grades less than 80 percent			Percent of Entrants with average grades less than 70%		
	Alberta	Calgary	Lethbridge	Alberta	Calgary	Lethbridge	Alberta	Calgary	Lethbridge
2019	88.7	86.7	79.9	4.3	11.0	50.8	0.0	0.0	9.1
2018	88.3	86.7	80.1	7.5	10.6	50.2	0.0	0.3	9.3
2017	88.8	86.0	79.0	4.7	13.4	50.5	0.0	0.3	8.7
2016	87.4	85.7	81.2	10.0	15.3	40.6	0.0	0.1	5.2
2015	87.1	85.8	81.5	10.7	14.2	40.5	0.0	0.2	4.9
2014	87.3	86.5	80.7	8.7	10.7	42.4	0.0	0.1	5.7
2013	86.5	86.2	80.7	10.3	12.6	53.4	0.0	0.0	6.9
2012	86.6	84.1	80.7	12.7	22.8	44.5	0.2	0.1	5.2
2011	86.2	82.7	80.7	14.2	32.1	45.0	0.1	0.4	5.6
2010	85.5	82.2	80.2	18.1	37.5	48.2	0.1	0.3	6.5
2009	85.7	82.2	80.2	15.1	36.1	49.2	0.1	0.5	6.5
2008	86.0	82.9	80.1	16.4	33.2	47.4	0.2	0.1	7.0
2007	86.1	83.1	79.6	15.3	29.9	53.0	0.4	0.0	9.1
2006	NA	NA	78.0	NA	NA	50.0	NA	NA	6.2
2005	86.2	84.0	79.7	NA	NA	NA	NA	NA	NA

Source: Macleans magazine annual guide to Canadian universities reports entering grades from the previous fall. The average entering grade and lowest entry grades are affected by the program mix within each university. In the missing years, universities did not participate in the survey.

Notes on specific programs entry grades in 2021.

The University of Alberta (January 2022) website reports admission average grade ranges: Arts, mid-70s to low 80s; Education, mid 70s to high 80s; Engineering, Low to high 80s; Kinesiology, High 80s to low 90s; Nursing, High 80s to low 90s; Science, High 80s to low 90s. https://www.ualberta.ca/admissions/undergraduate/admission/admission-requirements/competitive-requirements.html. The University of Calgary website reports admission grade ranges: Business, Mid 80s; Civil Engineering, low 90s; Computer Science, mid 80s; Economics, mid 80s; History, high 70s; Kinesiology, high 80s; Nursing, low 90s. https://www.ucalgary.ca/future-students/undergraduate/requirements. The University of Lethbridge actual admission averages in fall 2021 (provided from Lethbridge internal sources) by program: Bachelor of Arts, 80; Bachelor of Science, 85; Bachelor of Management, 81; Transfer Program, Engineering, 85; pre-Bachelor of Education/Science 85; pre-Bachelor of Education, Arts 83. The last three programs provide subsequent access to engineering or education. The last two universities MacEwan University (opened 2009) and Mount Royal University (opened 2011) both were conversions from non-university to university post-secondary institutions. They have much lower average entry grades and do not participate in the Macleans' survey. Mount Royal https://www.mtroyal.ca/Admission/_pdfs/ssdata_admission_requirements.pdf reports required averages on the Diploma courses of 85-90% in Education, 90-100% in midwifery and 95-100% in nursing. MacEwan University https://www.macewan.ca/academics/programs/bachelor-of-science-in-nursing/admissions/requirements/ has similar admission averages in nursing but seems to have lower requirements in other programs.

Appendix B: Average Diploma Course Grades 2005-06 to 2018-19

Year	Average Examination Grade	Average School Assigned Grade
2005-06	66.7	70.6
2006-07	66.9	70.7
2007-08	66.6	70.8
2008-09	65.5	70.9
2009-10	65.0	71.1
2010-11	64.6	71.2
2011-12	65.1	71.6
2012-13	65.5	71.7
2013-14	65.7	72.2
2014-15	65.8	72.6
2015-16	65.4	73.2
2016-17	65.8	73.5
2017-18	66.5	73.8
2018-19	66.6	73.9

Source: Calculations by author. Averages are across students from school-level data weighted by number of enrolments. Excluding grades in French courses.

Appendix C: A derivation of the equation relating school-awarded average grades to examination-awarded average grades

1 Skills

Skills acquired by student i are divided into two mutually exclusive groups measured by a vector of indexes K_{ei} and K_{oi} (examination skills and other skills respectively). Skills K_{ei} improve grades on both the externally graded examination and on the school-awarded grades given by teachers at schools, school-awarded grades. Skills denoted K_{oi}, “other” skills, increase only school-awarded grades. An example of this type of skill could be the ability to make an oral presentation with a loud clear voice, a skill that might or might not be useful in raising school-awarded grades but may not be useful in raising examination grades. The derivation in this appendix shows that a critical assumption in the literature and in this paper is that all skills with positive weights in school-awarded grades also have some positive weight in examination-awarded grades, a common skills assumption. If this assumption is made, then it is possible to derive equation (1) in the paper, a regression of school-course-year observations of average school-awarded grade on average examination-awarded grade incorporating school and course fixed effects. In equation (1), the school fixed effects are estimates of the parameter that describes school grading standards (SGSP), the increase in average school-awarded grades across courses at that school relative to the school-awarded grade merited by student skills measured at the “average” school where grading standards are neither high nor low.

The level of social and economic status (SES) of student i’s household is measured by index variable(s) Z_i. The skill levels of the i^{th} student attending the kth school at the time of assessment in a diploma course are

$$K_{eki} = \gamma_0 + \gamma_1 Z_i + Q_{ek} + \varepsilon_{1ki} \qquad (1)$$

and

$$K_{oki} = \alpha_0 + \alpha_1 Z_i + Q_{ok} + \varepsilon_{2ki} \qquad (2)$$

Skills are not course specific but, as will be made clear below, a given skill is not equally weighted in constructing the school awarded or examination grade in different courses.[1] Student skills are partly acquired from their home environment and increase in the SES variable Z_i.[2] Q_{ek} and Q_{ok} represent examination skills and other skills respectively as accumulated by a student enrolled at school k from their past sequence of schools including school k (and including any effect of possible peers at the past sequence of schools they may have attended) up to time t when we observe a school awarded grade and an examination grade. Q_{ek} and Q_{ok} take the same value for all students at school k and are time invariant.[3]

---

[1] Literacy could have a higher weight in an English course and mathematical skills a higher weight in physics but both courses would likely require some literacy skill and some mathematical skill.

[2] Z_i could be, for example, total years of education of both parents or the number of years of education of the most educated parent. It is important the student skills are learned both at school and at home.

[3] Forcing Q_{ok} and Q_{ek} to be identical for all students at a school may seem like a strong assumption. Since the empirical work uses average grades by school and course, if at this stage there was some random variation within students at the school in these skills acquired from this and their previous sequence of schools, it would be averaged out later. One source of variation could be that individual students enrolled at school k at time t had experienced a different

There are two mean zero random components of skills across all students at all schools that are uncorrelated with Z_i denoted ε_{1i} or ε_{2i}. The average value of ε_{1ki} or ε_{2ki} across students' i in course j in a school k need not be zero. In the most obvious case, students with a high value of ε_{1i} or ε_{2i} (there is no school subscript here) may find themselves directed into specific schools or specific courses. Parents of a high ε_{1i} or ε_{2i} student may seek out a specific school, perhaps even schools with higher values of Q_{ek} or $Q_{ok.}$ A positive or high epsilon student has a positive value of ε_{1i} (or ε_{2i}) and a higher level of skills conditional on their value of Z_i than the skill level predicted for the average student at each level of Z regardless of school of enrolment. The school quality terms Q_{ok} or Q_{ek} then add (or reduce) a student's skills from the student's skill level conditional on the student's Z_i and that student's realization of ε_{1i} or $\varepsilon_{2i.}$ A student can have a high level of skill because they come from a high SES background; because they come from a "good" school with a high Q_{ok} or Q_{ek}; or because they have a large value of ε_{1i} or ε_{2i}.[4]

---

sequence of schools in the period before time t. The notation is already sufficiently dense to let this possibility go for the time being.

[4] This broad set for the sources of student ability is important to describe variation in average school results on external examinations. Some schools where children come from disadvantaged backgrounds outperform schools where children come from advantaged backgrounds indicating high school quality when schools are not selecting on ability. Schools have a substantial range of examination results across schools attended by students of the same average SES background. Johnson (2005) documents these effects in elementary schools in Ontario. In some cases, some schools have entrance policies that select on high overall ability, some combination of SES and high epsilon values. These schools, in a very uninteresting way,

2 Grades

$E_{ijk(t)}$ is the examination grade in course j at time t for a student enrolled at school k. Skills at school k are transformed into examination grades $E_{ijk(t)}$ using

$$E_{ijk(t)} = \delta_{0j} + \delta_{1j} K_{ei} + D_{j(t)} + \varepsilon_{3ijk(t)} \qquad (3)$$

δ_{0j} and δ_{1j} are course specific parameters. $\varepsilon_{3ijk(t)}$ is the random component in (3). Any random component that hits all students in course j at all schools in the entire province in time t is absorbed into the parameter $D_{j(t)}$. There could also be a school-specific shock to all courses at school k in time t, a forest fire in school k's area. There could be a specific shock to a course that affected all students in that specific course at that school - a student vomiting in the exam room in that course but not in other courses at that school. There could be a student specific random component, student "i" just has a bad or good day. If (3) were to be estimated using individual observations across courses and years, the error structure implies clusters at the school-course-year as well as the school-year level. There are no direct measures of individual skills that allow (3) to be estimated and, in the Alberta data, no individual observations are available for research. Instead, the available data describes school average grades by course and year.

The term $D_{j(t)}$ in the grading equation is controversial. This formulation clarifies that a change in the value of $D_{j(t)}$ is a change in the difficulty of the examination; the same level of skill

---

invariably have very strong examination results conditional on the SES background of their students. The Calgary charter schools fall into this category in Alberta.

translates to a different grade in different years in the same course.[5] Provincial examinations are carefully constructed with the intention of being equally difficult across years and, if that goal were achieved, then $D_{j(t)}$ equals zero in all courses in all years. Then all variation in the average examination outcome across different years is variation in average examination-related skills of the participating students.

Skills for the same student in the same course are transformed into school-awarded grades $S_{ijk(t)}$ at school k using

$$S_{ijk(t)} = \beta_{0j} + \beta_{1j} K_{ei} + \beta_{2j} K_{oi} + P_{jk(t)} + \varepsilon_{4ijk(t)} \qquad (4)$$

$\varepsilon_{4ijk(t)}$ is a random component similar to $\varepsilon_{3ijk(t)}$ clustered with individual observations at both the school-year level and the school-course-year level. $P_{jk(t)}$, when positive, indicates that teachers at school k in course j award higher school-awarded grades to all students in course j than are consistent with the actual skill level of their students in that course. If $P_{jk(t)}$ is negative, teachers at school k in course j have high standards, if $P_{jk(t)}$ is positive, the teachers in course j at school k

[5] The economics literature typically adds a time fixed effect and or standardizes individual grades to capture or remove possible variation in the difficulty of the examination over time. If such an adjustment is not made, there is an implicit or explicit assumption that assigned grades across different years are comparable. Adding the $D_{j(t)}$ terms makes the adjustment for variation in difficulty explicit and the grades remain in the reported scale from 0 to 100. One advantage of that choice is that admissions to post-secondary are usually set in terms of cut off grades measured from 0 to 100. For the measurement of the school-level variation in grading standards in this paper, this is a useful and natural scaling.

have low standards. $P_{jk(t)}$ is zero when grading standards are neither low nor high in course j at school k.[6]

This model of grades states that within a school, at the same level of either ε_{1i} or ε_{2i}, higher SES students have higher grades within a school. Although it is likely schools with a larger proportion of high SES students will have both higher average school-awarded grades and higher average examination-awarded grades than schools with a lower proportion of high SES students, this may not happen if a school is either a higher quality school (a high value of Q_{ek} or Q_{ok}) or if the school has a large proportion of high epsilon students.[7]

$D_{j(t)}$ measures variation in examination difficulty over time in the simplest case. $P_{jk(t)}$ measures grading standards by school and course.[8]

---

[6] Equations like (5) are typically not estimated since skills are not directly observed. One exception is Cornwall, Mustard and Van Parys (2013) who add measures of specific non-academic skills to successfully explain gaps between external and teacher grades in kindergarten.

[7] League tables of school average examination results are often presented without any such nuance. In many but not all cases, schools with high average examination results are schools where more students come from stronger SES backgrounds. It is the schools with strong results and large proportion of students from lower SES backgrounds that are the most interesting schools. That is the focus of Johnson (2005) and his subsequent studies of the same type.

[8] Bonesronning (2008) estimates a parameter similar to $P_{jk(t)}$ as the average difference between internal grades and external grades on a 6-point scale within a classroom of students.

3 Expressing the school-awarded grade as a function of the examination-awarded grade

Equations (1), (2), (3) and (4) are manipulated to parameterize a relation between the school-awarded grade and the examination-awarded grade, first in the most general case and then in two special cases.

Combining expression (1) and expression (3) as

$$Z_{ijk(t)} = \frac{E_{ijk(t)}}{\delta_{1j}\gamma_1} - \frac{\delta_{0j}}{\delta_{1j}\gamma_1} - \frac{\gamma_0}{\gamma_1} - \frac{1}{\gamma_1} Q_{ek} - \frac{1}{\delta_{1j}\gamma_1} D_{j(t)} - \frac{1}{\gamma_1} \varepsilon_{1ijk(t)} - \frac{1}{\delta_{1j}\gamma_1} \varepsilon_{3ijk(t)} \quad (5)$$

shows that a student's examination grade in course j at school k; adjusted for the quality of the school k attended by student i; the difficulty of the examination; and the value of $\varepsilon_{1i;}$ creates a measure of the unobserved SES level of student i enrolled in course j at school k at time t.[9]

Substitution of equations (1) and (2) into (4) yields:

$$S_{ijk(t)} = \beta_{0j} + \beta_{1j} (\gamma_0 + \gamma_1 Z_{ijk(t)} + Q_{ek} + \varepsilon_{1ijk(t)}) + \beta_{2j} (\alpha_0 + \alpha_1 Z_{ijk(t)} + Q_{ok} + \varepsilon_{2ijk(t)})$$

$$+ P_{jk(t)} + \varepsilon_{4ijk(t)} \quad (6)$$

Expression (6) shows that because school-awarded grades increase in student skills, these skills increase with the student's SES, the student's level of ε_1 and ε_2 and the quality of the student's school, Q_{ek} and Q_{ok}. The examination grade with additional terms (5) measures the student's unobserved SES level. Substitution of (5) into (6) and simplifying yields

[9] Burgess and Greaves (2013) make a similar observation in their analysis of examination results in English schools.

$$\mathrm{S}_{ijk(t)} = \beta_{0j} + \beta_{2j}\alpha_0 - \frac{\beta_{2j}\delta_{0j}}{\delta_{1j}} - \frac{\beta_{1j}\delta_{0j}}{\delta_{1j}} - \frac{\beta_{2j}\gamma_0\alpha_1}{\gamma_1} - \left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\mathrm{D}_{j(t)}$$

$$+ \left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\left[\mathrm{E}_{ijk(t)}\right]$$

$$+ \mathrm{P}_{jk(t)} - \left(\frac{\beta_{2j}\alpha_1}{\gamma_1}\right)\left[\varepsilon_{1ijk(t)}\right] + \beta_{2j}\,\varepsilon_{2ijk(t)} - \left(\frac{\beta_{2j}\alpha_1}{\delta\gamma_1}\right)\mathrm{Q}_{ek} + \beta_{2j}\mathrm{Q}_{ok} -$$

$$\left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\varepsilon_{3ijk(t)} + \varepsilon_{4ijk(t)} \qquad (7)$$

Expression (7) is a useful positive relation between a student's school-awarded grade and a student's examination-awarded grade. Student i's school-awarded grade is predicted, up to the student level random component, from the student i's s examination-awarded grade adjusted for the difficulty of the examination, the quality of the school, school grading standards and the values of ε_1 and ε_2 of the student. If the examination was too easy, then $\mathrm{D}_{j(t)}$ is positive and the prediction of the skill-consistent school-awarded grade from the examination-awarded grade in time t is adjusted downwards. Similarly, if the school attended by the student is very good at producing examination results, the prediction of the skill-consistent school-awarded grade is adjusted downward for large values of Q_{ek}. If the student is a high ε_1 student, the prediction of the skill-consistent school-awarded grade from the examination-awarded grade is adjusted downwards. The skill-consistent school-awarded grade is adjusted upwards for a high ε_2 student or a high-quality school, a high value of Q_{ok}. These parameters affect skills that raise only school-awarded grades. Equation (7) is now aggregated to a school-course average observation for estimation.

4. Aggregation

The averages of the grading standards parameters across all j courses offered at school k at a time (t), $\mathrm{P}_{k(t)}$, (notice j disappears in the subscript) are defined by:

$$\mathrm{P}_{jk(t)} \equiv \mathrm{P}_{k(t)} + \left(\mathrm{P}_{jk(t)} - \mathrm{P}_{k(t)} \right) \tag{8}$$

A positive value of $\mathrm{P}_{k(t)}$ in time t describes school k that has low grading standards averaged across all its courses. A negative value of $\mathrm{P}_{k(t)}$ in time t describes a school with high grading standards averaged across all its courses.

In the most general case, (7) is aggregated to school-course-time period observations by averaging across student observations in course j at school k at time (t). Averages are denoted with a bar for course j at school k.[10] The relation between average school awarded grades and average examination grades by school and course is

$$\begin{aligned}\bar{\mathrm{S}}_{jk(t)} = {} & \beta_{0j} + \beta_{2j}\alpha_0 - \frac{\beta_{2j}\delta_{0j}}{\delta_{1j}} - \frac{\beta_{1j}\delta_{0j}}{\delta_{1j}} - \frac{\beta_{2j}\gamma_0\alpha_1}{\gamma_1} - \left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\mathrm{D}_{j\,(t)} \\ & + \left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\left[\bar{\mathrm{E}}_{jk(t)}\right] \\ & + \mathrm{P}_{k(t)} - \left(\frac{\beta_{2j}\alpha_1}{\delta\gamma_1}\right)\mathrm{Q}_{ek} + \beta_{2j}\mathrm{Q}_{ok} \\ & - \left(\frac{\beta_{2j}\alpha_1}{\gamma_1}\right)\left[\bar{\varepsilon}_{1jk(t)}\right] + \beta_{2j}\,\bar{\varepsilon}_{2jk(t)} \\ & + \left(\mathrm{P}_{jk(t)} - \mathrm{P}_{k(t)}\right) - \left(\frac{\beta_{1j}\gamma_1 + \beta_{2j}\alpha_1}{\delta_{1j}\gamma_1}\right)\bar{\varepsilon}_{3jk(t)} + \bar{\varepsilon}_{4jk(t)} \end{aligned} \tag{9}$$

[10] Larger schools would have more than one class group in a course group. Alberta reports average grades by school.

Equation (9) leads directly to the equation estimated in the paper. Variation in the average school-awarded grade in a course at a point in time is modelled as a time fixed effect specific to each course in that time period; a linear function of the school's average grade on the examination in that course with a coefficient that varies by course; and an average school fixed effect across all courses at the school that is the value of SGSP, the school grading standards parameter, as well as an error term.

One useful way to understand (9) is that $\overline{\mathrm{E}}_{jk(t)}$ is a control variable that captures the otherwise unobservable average SES value of students in the course and school as well as part of the selection of high $\bar{\varepsilon}_{1jk(t)}$ students into the school and part of the school quality effect denoted Q_{ek}. The course-specific time fixed effect captures a series of uninteresting constants in the conversion of skills to grades across courses as well as variation over time in the difficulty of the examination.

The parameters of the most interest are the values of $\mathrm{P}_{k(t)}$, the measure of grading standards at school k at time t, denoted SGSP in the text. In equation (9) SGSP is grouped with the schools' ability to increase student skills, both skills measured on examinations Q_{ek} and those skills not measured on examinations, Q_{ok} in the school fixed effect. There is no separation in the general case between a school that has low (high) grading standards and a school that produces higher (lower) skills assessed through the school-awarded grade (Q_{ok}) and a school where there is higher (lower) quality of skill production in the examination dimension (Q_{ek}) that reduces (increases) the skill-consistent school awarded grade for a given examination grade.

The fourth line shows an additional complication in the use of the average examination grade to predict the average school awarded in the general case with the isolation of the school fixed effect as the measure of over or under-grading. Part of the error term, $\left(\frac{\beta_{2j}\alpha_1}{\gamma_1}\right)\left[\bar{\varepsilon}_{1jk(t)}\right] + \beta_{2j}\,\bar{\varepsilon}_{2jk(t)}$ is the average values of ε_1 and ε_2 of students enrolled in each course at that school in that period. These components are certainly correlated with the school's average examination grade. This would bias the coefficient estimate on the average examination grade and more importantly bias the estimates of the school fixed effect coefficient. If parents of higher ability students, even after conditioning on social and economic characteristics, choose higher quality schools, then the school fixed effects will over-estimate the grading standards parameter at that school. What could be interpreted as low grading standards, a high value of $P_{k(t)}$, is partly higher ability students selecting into the school. Teachers do not have low grading standards, they are rewarding actual higher skills. In Section 5 an unbiased estimate of $P_{k(t)}$ is obtained by placing further restrictions on the model.

The error term in (9) is clustered at the school-time period level across the courses in that year if common shocks hit all students at a school in all courses in a specific period. An example from this period in Alberta is a forest fire that closed an entire city of schools for part of a year.[11]

5 The special case of the common skills assumption

[11] A clear discussion of clustering is found in Cameron and Miller (2015).

If there is a common set of skills measured by school-awarded grades and examination grades then β_{2j} is zero.[12] Equation (9) then rewrites as

$$\bar{\mathrm{S}}_{jk(t)} = \beta_{0j} - \frac{\beta_{1j}\delta_{0j}}{\delta_{1j}} - \left(\frac{\beta_{1j}}{\delta_{1j}}\right)\mathrm{D}_{j\,(t)}$$

$$+ \left(\frac{\beta_{1j}}{\delta_{1j}}\right)\left[\bar{\mathrm{E}}_{jk(t)}\right]$$

$$+ \mathrm{P}_{k(t)}$$

$$+ \left(\mathrm{P}_{jk(t)} - \mathrm{P}_{k(t)}\right) - \left(\frac{\beta_{1j}}{\delta_{1j}}\right)\bar{\varepsilon}_{3jk(t)} + \bar{\varepsilon}_{4jk(t)} \qquad (10)$$

Under the common skills assumption, the estimates of the time-period course-specific fixed effects vary with the difficulty of the examination within a course and across courses. The slope coefficients on the average examination score vary by course. The school fixed effects are measures of the average of additional (reduced) grades awarded by the school's teachers relative to their students' actual skills, averaged across courses. The error term is clustered at the school across courses within a period. This is the equation is estimated in the text.

6. A very strong assumption in the literature

[12] That type of language is mentioned in passing in some, but not all, of the literature. The restriction does not mean, as further emphasized in Section 6, that skills are rewarded identically by teacher grading and examination grading.

A very strong assumption is that identical skills are transformed into identical grades (up to a random component) by teachers and examination graders. This assumption requires that β_{2j}=0; β_{0j}= δ_{0j}; and $\beta_{1j} = \delta_{1j}$. In that case, equation (9) or equation (10) reduce to

$$\overline{\mathrm{S}}_{jk(t)} = \overline{\mathrm{E}}_{jk(t)} - \mathrm{D}_{j\,(t)} + \widehat{\mathrm{P}}_{k(t)} + \left(\mathrm{P}_{jk(t)} - \widehat{\mathrm{P}}_{k(t)} \right) - \bar{\varepsilon}_{3jk(t)} + \bar{\varepsilon}_{4jk(t)} \quad (11)$$

which is more commonly written as

$$\overline{\mathrm{S}}_{jk(t)} - \overline{\mathrm{E}}_{jk(t)} = -\mathrm{D}_{j\,(t)} + \widehat{\mathrm{P}}_{k(t)} + \left(\mathrm{P}_{jk(t)} - \widehat{\mathrm{P}}_{k(t)} \right) - \bar{\varepsilon}_{3jk(t)} + \bar{\varepsilon}_{4jk(t)} \quad (12)$$

The simple difference between school awarded grades and examination grades is the dependent variable in Bonnesrunning (2008), Gibbons and Chevalier (2008), Falch and Napier (2013) and Berg, Palmgren and Trufores (2020).

My paper presents estimates of equation (10) rather than equation (11) because the estimates of the coefficient on the examination grade are very far from unity as implied in (11).

Additional References for Appendix C

Berg, Petter, Ola Palmgren and Bjorn Tyrefors (2020) “Gender grading bias in junior high school mathematics.” Applied Economics Letters 27 (11): 915-919.

Bonesronning, Hans (2008) “The Effect of Grading Practices on Gender Difference in Academic Performance” Bulletin for Economic Research 60(3): 245-264.

Burgess, Simon and Ellen Greaves (2013) “Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities.” Journal of Labor Economics 31(3): 535-576.

Cameron, A. Colin and Douglas L. Miller (2015) “A Practitioner’s Guide to Cluster-Robust Inference.” Journal of Human Resources 50(2): 317-372.

Cornwell, Christopher, David B. Mustard and Jessica Van Parys (2013) “Non-cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School.” Journal of Human Resources 48 (1): 236-264.

Falch, Torberg and Linn Renee Naper (2013) “Educational evaluation schemes and gender gaps in student achievement.” Economics of Education Review 36: 12-25

Gibbons, Stephen and Arnaud Chevalier (2008) “Assessment and age 16+ education participation.” Research Papers in Education 23 (2):113-123.