WORKING PAPER 2019:8

Explaining gender gap variation across assessment forms

Georg Graetz
Arizo Karimi

IFAU INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Explaining Gender Gap Variation across Assessment Forms[a]

Georg Graetz[b] Arizo Karimi[c]

April 18, 2019

Abstract

In Sweden, females outperform males on compulsory and high school GPAs by a third of a standard deviation, while males outperform females on the Swedish SAT by the same magnitude. We establish that GPAs capture different attributes and skills compared to SAT scores. Differences in motivation and effort explain up to 60 percent of the female advantage in GPAs, while cognitive skills explain 40 percent of the male advantage in SAT scores. The latter is accounted for by differential self-selection into taking the SAT. Our findings imply large effects of the choice of university admission criterion on admitted students' characteristics.

Keywords: gender gaps, student assessment, cognitive skills, non-cognitive skills, university admissions
JEL Classification: I21, I24, J16

[a]We thank Helena Holmlund, Erica Lindahl, Oskar Nordström Skans, Tuomas Pekkarinen, and seminar participants at the Uppsala Center for Labor Studies for helpful comments.

[b]Corresponding author. Department of Economics, Uppsala University; Centre for Economic Performance, CESifo, IZA, and Uppsala Center for Labor Studies. Email: georg.graetz@nek.uu.se.

[c]Department of Economics, Uppsala University; Institute for Evaluation of Labor Market and Education Policy (IFAU), and Uppsala Center for Labor Studies. Email: arizo.karimi@nek.uu.se.

1 Introduction

Women outnumber men in university attendance rates in the majority of OECD countries (Goldin, Katz, and Kuziemko, 2006), a fact that has spurred both public debate and research into why females systematically outperform males in compulsory and high school grade point averages (GPAs), and are less likely to drop-out of high school (Murnane, 2013; OECD, 2017; SCB, 2017). However, a challenge for understanding gender gaps in educational outcomes is that differences in measured performance consistently vary with assessment form. In particular, despite the female advantage in overall GPAs from school, males tend to perform at least as well as females on standardized aptitude tests (see for instance Duckworth and Seligman, 2006), and typically outperform females on tests measuring quantitative skills (see for instance Fryer Jr and Levitt, 2010).[1] As both GPAs and standardized tests are widely used in university admission procedures, and as academics and policy makers debate their relative merits (Rothstein, 2004; Diamond and Persson, 2016), it is important to understand what attributes each type of assessment captures.[2]

Using administrative data on the Swedish population, we document that females, on average, outperform males on both compulsory school and high school GPAs by about a third of a standard deviation, but that the reverse is true for the Swedish SAT, where females *under*-perform by a third of a standard deviation relative to males. These gaps are stable across the cohorts covered by our data, born between 1977 and 1996. The pattern of a flipped gender gap across school GPAs and SAT scores is present also within subject areas (verbal and quantitative), and there is a sizeable gender gap in within-individual score differences: a female student's position in the score distribution deteriorates by half a standard deviation, relative to males, when moving from school grades to the SAT.

We test for two potential explanations for the flipping gender gap across GPAs and SAT scores. First, we investigate whether school grades capture different individual attributes or skills than do SAT scores, and if so, whether gender differences in the endowment of these attributes are large enough to account for some portion of the gaps. While standardized achievement and aptitude tests and course grades both measure students' acquired knowledge and skills, previous evidence suggests that the two assessment formats differ in important ways. In particular, relative to standardized tests, course grades seem more strongly associated with personality traits like conscientiousness, which is generally higher in girls than in boys.[3] Second, we ask

[1]The latter finding is particularly puzzling as there appears to be no systematic gender difference in early-age mean numerical ability (Kersey, Braham, Csumitta, Libertus, and Cantlon, 2018), nor in regions or countries with a more gender-equal culture (Guiso, Monte, Sapienza, and Zingales, 2008; Pope and Sydnor, 2010; Gevrek, Neumeier, and Gevrek, 2018). Some evidence, however, points to greater variance in test scores in the male population (see, for instance Machin and Pekkarinen, 2008).

[2]In particular, one of the main rationales for standardized university admissions tests is to encourage a diversified student body (Dynarski, 2017). See Edwards, Coates, and Friedman (2012) for an overview of countries using central admissions tests.

[3]See Almlund, Duckworth, Heckman, and Kautz (2011). Duckworth and Seligman (2006) find that female

whether non-random selection into taking the SAT differs across the genders. In any setting where there is choice involved and potential for systematically different decisions made by males and females, gender gaps cannot be taken at face value (see for instance Card and Payne (2017) for evidence on the sources of the gender gap in STEM entry).

We test these explanations using data on cognitive skills, motivation, and effort measured at age 13 for a representative sample of Swedish students born in 1992, which we link to population-wide individual-level data on compulsory school and high school GPAs (measured at ages 16 and 19, respectively) as well as SAT scores (typically measured at ages 19-20). We are thus able to improve on previous studies, which focus on the importance of personality traits and other non-cognitive factors for school grades, but lack the data to explore the sources of a differing gender gap in standardized achievement tests.

We find that standardized indices of cognitive skills, motivation, and effort are strongly positively related to compulsory school GPA (CSGPA) and high school GPA (HSGPA). While SAT scores are also highly informative about cognitive skills, they show no correlation with motivation or effort. This suggests that school-level assessments capture different attributes than the SAT. We find a pronounced female advantage in motivation and effort in the representative sample, which accounts for over 60 percent of the female advantage in CSGPA, and for 30 percent in the case of HSGPA. Among SAT takers in the sample, males have higher cognitive skills than females, especially along the dimensions important for the test. These differences account for more than 40 percent of the gender gap in SAT scores.

However, in the representative sample as a whole there are no gender differences in cognitive skills. This means, first, that cognitive skills do not help explain the female advantage in GPAs. And second, the explanatory power of cognitive skills for the SAT gender gap is entirely due to the fact that males select into taking the test, based on the skills that predict performance, more systematically than females. In other words, if all individuals took the SAT, we would not predict a gender gap in scores based on differences in cognitive skills.

Taken together, our results show that understanding the origins of gender gaps in non-cognitive traits will be key to informing the debate about why boys under-perform, relative to girls, in school; and understanding why the genders select differentially into taking the SAT will be necessary to explain the male advantage there. (The fact that overall, females are more likely to take the SAT, is largely accounted for by their higher motivation.)

Finally, we note that even after controlling for individual differences in attributes and skills, there are sizeable gender gaps in performance: of 0.14 and 0.24 standard deviations (in favor

advantage in self-discipline among eighth graders accounts for a larger portion of the female dominance in report card grades than in achievement test scores. Similarly, Cornwell, Mustard, and Van Parys (2013) find that non-cognitive factors account for the female advantage in teacher assessed grades among primary school students in the US. Fortin, Oreopoulos, and Phipps (2015) focus on the shift in the mode of girls' high school GPA from B to A that occurred between the 1980s and 2000s in the US, leaving boys behind, and conclude that gender differences in expectations for attending higher education are the most important factors accounting for this trend.

of females) in CSGPA and HSGPA, respectively, and of 0.19 standard deviations (in favor of males) in SAT scores. Possible explanations for these (in the context of our study) 'unexplained' portions of the gender gaps could be differences in unobserved skills and attributes, or that those we do observe are measured with error. Other possible explanations for the unexplained gaps may also be found in related strands of literature. First, there is evidence of women performing worse than men, on average, in multiple-choice formats compared to on free-response exams.[4] In general, evidence from the lab and the field suggest a gender gradient in performance in competitive environments, which has potential implications for gender gaps in SAT scores or other high-stakes achievement tests.[5] In a second strand of literature, teacher-student interactions—such as teacher gender effects due to role models or teacher discrimination—are studied as potential explanations for the gender gap in grades and cumulative GPAs. The evidence from this literature is, however, inconclusive on the empirical relevance of teacher (gender) effects for gender gaps in school performance.[6]

This paper is organized as follows. In Section 2 we describe the institutional setting and discuss the features by which the SAT differs from teacher-assessed school grades. In Section 3 we describe the data sources, samples, and variables used in our study. Section 4 presents the results. Finally, Section 5 concludes the paper by discussing the policy implications of our findings.

2 Setting

The Swedish education system consists of nine years of compulsory schooling, followed by three years of (voluntary) high school, the completion of which is required for university eligibility.[7] For oversubscribed high school programs, slots are allocated among applicants based on compulsory school GPA (CSGPA). Similarly, slots to oversubscribed university programs are

[4]See e.g. Bolger and Kellaghan (1990). However, the male advantage in multiple-choice tests seems more prevalent when wrong answers are penalized with negative points as women exhibit a lower likelihood of guessing relative to men in such tests, which may be attributed to differences in risk preferences (Pekkarinen, 2015; Akyol, Key, and Krishna, 2016; Baldiga, 2013). Wrong answers are not penalized on the Swedish SAT.

[5]Gneezy, Niederle, and Rustichini (2003) find that men's performance increases as the competitiveness of the test increases, while that of females does not. Similarly, Niederle and Vesterlund (2007) find that males show a stronger preference for competitive tasks than females. In the context of education, results presented by Jurajda and Münich (2011) suggest that men perform better than women in entrance exams for more prestigious schools, but not in the exams for less competitive schools. Similarly, Ors, Palomino, and Peyrache (2013) find that females tend to perform worse in more competitive examinations with high future payoffs than do men.

[6]For instance, Holmlund and Sund (2008), Puhani (2018), and Lindahl (2016) find no evidence in support of the hypothesis that a same-sex teacher improves student outcomes, while Dee (2005, 2007) and Falch and Naper (2013) suggest that students benefit from having a same-sex teacher. Hinnerich, Höglin, and Johannesson (2011) find no evidence of discrimination using blind- and non-blind grading of the same exam, while Lavy (2008), Terrier (2016), and Berg, Palmgren, and Tyrefors (2019) find that boys face discrimination in teacher grading.

[7]All students follow the same curriculum in compulsory school, while there is a range of high school programs, both vocational and academic. The vocational tracks include academic subjects (such as mathematics, English, and Swedish) granting access to some university programs.

allocated based on high school GPA (HSGPA) and SAT scores through a centralized process. In the case of oversubscribed programs, Swedish universities are legally required to fill at least one third of slots based on a GPA ranking, and at least one third of slots based on a SAT ranking.

The SAT is voluntary, but nevertheless a high stakes test given the admission process. For instance, all medical programs in Sweden typically require the top score on the SAT or a HSGPA exceeding the mean by more than two standard deviations. Aside from medical programs, the top score on the SAT is a sufficient condition for admission to nearly all university programs in Sweden. In general, a higher SAT expands the set of programs that a student has access to, especially for students with a relatively low HSGPA (Graetz, Öckert, and Skans, 2018).

The assessment formats producing CSGPA, HSGPA, and SAT scores differ along several dimensions, as summarized in Table 1. Apart from natural differences in purpose, participation requirements, and timing, it is noteworthy that both CSGPA and HSGPA are based on more than a dozen separate written and oral assessments occurring during periods of several years. The SAT score, in contrast, is determined in a single one-day exam involving 120-150 multiple-choice questions.[8] In addition to written exams, school grades can be based on several other test formats. For the majority of cohorts studied in this paper, the grading system is a criterion-referenced grade scale. According to the Swedish National Agency for Education (Skolverket), the grades should reflect the students' acquired skills and knowledge based on a holistic assessment of written examinations, lab reports, in-class discussions, oral presentations etc. Thus, teachers have considerable discretion in setting questions for tests underlying the GPAs as well as in setting the course grades that go towards the final GPA. The SAT is instead a centralized test administered by the Swedish Council for Higher Education, with the same questions faced by all students on a given test date. Unlike in the case of GPAs, grading of the SAT is done blindly and graders do not have any discretion, given the multiple-choice format. There is no negative marking in the SAT: each correct answer is rewarded with one point, and wrong answers yield zero points. All assessment formats that we consider have in common that they test for knowledge in various areas. GPAs are based on subjects tests including math, Swedish, and English. The SAT has two parts, one testing for language, and one testing for numerical skills.[9]

3 Data

Our data come from population-wide individual-level administrative registers. The data include year and country of birth, gender, parents' country of birth and educational attainment, grade point averages (GPAs) from compulsory school and high school, as well as SAT scores. For the purpose of documenting gender gaps in test scores, we focus on the compulsory school

[8]Students may repeat the SAT, however, and only the best score counts in admission (the test takes place twice each year).

[9]The Swedish SAT is designed based on the American SAT (SOU, 2004), but differs from the latter in that it does not contain an essay component.

Table 1: Overview of assessments

	CSGPA	HSGPA	SAT
Purpose	Progression to HS	Progression to university	
Participation	Compulsory	Voluntary	Voluntary
Timing	Years 7-9	Years 10-12	Usually year 12+
Number of tests	15+	15+	1, may repeat
Format	Written	Written	Multiple choice
Content	Mixed	Mixed	Mixed
Teacher discretion in choosing content	✓	✓	✗
Blind grading	✗	✗	✓
Teacher discretion in grading	✓	✓	✗

graduation cohorts of 1993-2012, corresponding to birth cohorts 1977-1996 as students typically graduate compulsory school at age 16. For high school and SAT, we focus on the years 1996-2015, corresponding to the same birth cohorts given the typical high school graduation age of 19, and given that most students take the SAT around the time of high school graduation.[10] Among the students who have ever taken the SAT, about half have taken the test more than once (Graetz, Öckert, and Skans, 2018). For repeaters, we use the results from the first test throughout. We have checked that our results are robust to using the highest life-time score instead.

Our main sample, however, consists of individuals born in 1992, for whom we have data on cognitive skills in grade 6 from the Evaluation Through Follow-up (ETF) study[11] conducted by the Department of Education at Göteborg University. The cognitive tests measure inductive (number sequences), spatial (plate folding), and verbal (synonyms and opposites) skills. We use these detailed measures as well as a composite index obtained by principal component analysis. In addition to cognitive tests, the ETF administers a comprehensive questionnaire to the test-taking students to elicit their motivation and time spent on homework. Using principal component analysis we create three measures of motivation: a general one that captures students' motivation to work towards getting admitted to a high-quality university program, achieving higher pay, becoming a productive member in society, etc.; a school-specific one capturing students' interest and motivation to learn in school; and a composite index combining the

[10]In our final sample, the maximum time between graduating from compulsory school and taking the SAT is seven years, and most individuals take the SAT within four years from compulsory school graduation.

[11]In Swedish: Utvärdering Genom Uppföljning (UGU).

two. Our measure of effort exerted is the time spent on homework that the students report.[12]

The ETF data cover a 10-percent stratified random sample of students born in 1992, which corresponds to some 10,000 individuals.[13] Due to non-response in the survey, our final sample consists of roughly 4,300 individuals. The ETF data further include sampling weights to allow nationally representative statistics. We adjust these weights to make the final sample representative in terms of gender, immigrant status, and compulsory school GPA decile. The cognitive and non-cognitive measures from the ETF survey are standardized to have zero mean and unit variance within the final estimation sample. GPAs and SAT scores are standardized within each cohort.

4 Results

4.1 Documenting gender gaps across assessment forms

Figure 1 reports standardized test scores and grades over time, by gender.[14] Females typically outperform males on both compulsory school GPA (CSGPA) and high school GPA (HSGPA) by about a third of a standard deviation (average gaps of 0.34 and 0.37 for CSGPA and HSGPA, respectively). But the reverse is true for the SAT, where females *under*-perform by a third of a standard deviation (an average gap of -0.32). These gaps are largely stable over time, with the exception of the narrowing of the HSGPA gap in 2011 and the widening of the SAT gap also in 2011.[15] The gaps are present across the score distributions, in the sense that across deciles of CSGPA and HSGPA the fraction of females (and hence the chance that a female student scores in a given decile) increases nearly monotonically, but decreases monotonically across deciles of the SAT; and the average SAT score of males is higher than that of females at all deciles of CSGPA and HSGPA (see Figures A1 and A2). It is also worth noting that despite the flipping gender gaps, the individual-level correlations between GPAs and SAT scores are

[12]Similar measures of cognitive skills, as well as non-cognitive (psychological) ability, are available for a large fraction of Swedish males born between 1955-1985, as these cohorts were subject to military conscription and underwent extensive enlistment examinations (Lindqvist and Vestman, 2011). These data include only a small number of female volunteers, and thus are less useful in our context as we seek to explain gender gaps in test scores. However, in results available on request, we document that GPAs correlate strongly and positively with a student's father's cognitive as well as non-cognitive skills as measured at the enlistment exams. In contrast, and consistent with our findings from the ETF data, SAT scores only correlate with cognitive skills, not non-cognitive skills.

[13]The ETF study performs a two-stage stratified cluster sampling, where municipalities are drawn at random in the first stage, and catchment areas within municipalities in the second stage. All students in the relevant cohort of the included catchment areas are covered in the sample. The ETF data contain 10 percent random samples of nine cohorts born between 1948 and 1998. While the cognitive tests are identical across the samples, the survey questions do not overlap. Therefore, we focus here on one cohort for which we have relevant data on traits, and for which we also have data on GPAs and SAT scores.

[14]The means plotted in the figure, along with their standard errors, are listed in Table A1.

[15]Although these shifts, which both favor males, coincide in timing, they were likely caused by two separate changes: a grading reform affecting the HSGPA gap, and an expansion of the quantitative section of the SAT. It is beyond the scope of this paper to investigate this issue.

quite similar across the genders (CSGPA, HSGPA, and SAT scores are all strongly positively correlated, although the correlation between HSGPA and SAT scores is somewhat weaker than that between CSGPA and SAT scores—see Table A2).

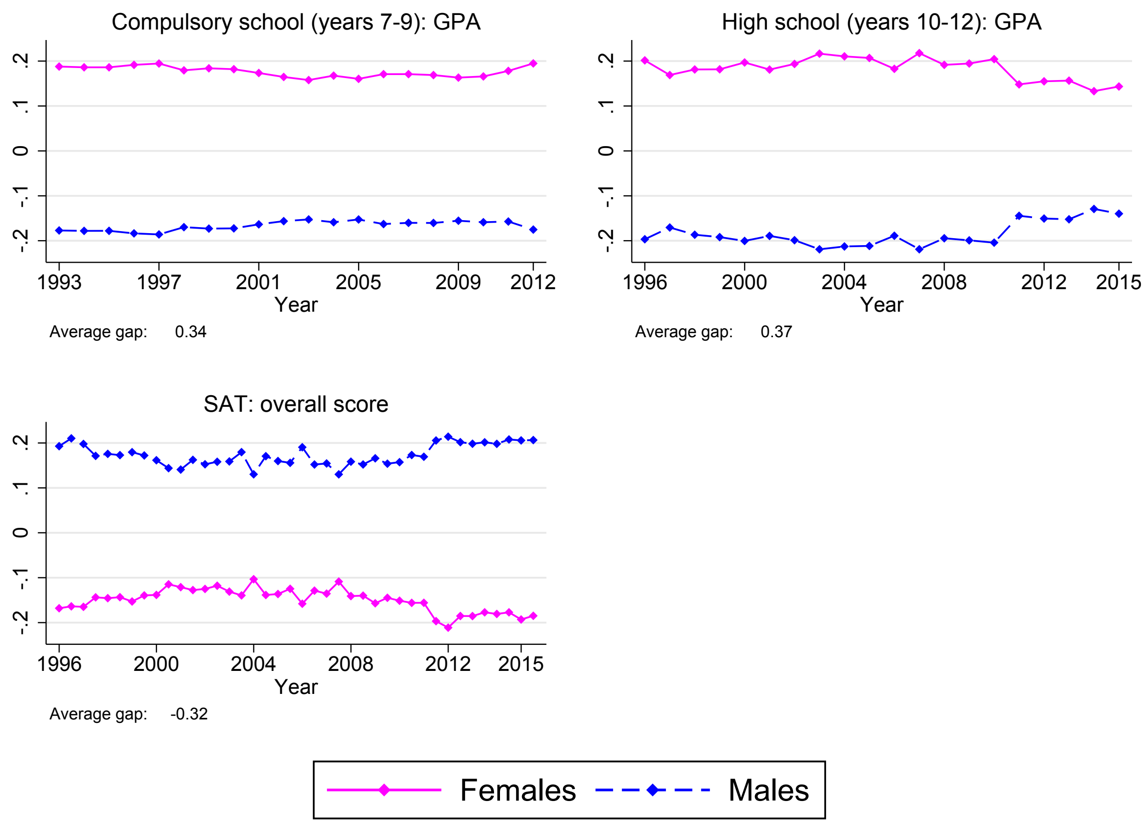


Figure 1: Standardized scores over time, by assessment form and gender

Gender gaps flip also within subject areas between the CSGPA and SAT score. In CS mathematics, girls outperform boys by 0.09, while in the quantitative part of the SAT, females under-perform by 0.53.[16] On a standardized score averaging CS grades in Swedish and English, girls outperform boys by 0.42, while in the verbal part of the SAT, females under-perform by 0.14. Between CS grades and SAT scores, the decreases in the gender gaps are very similar across overall (-0.66), mathematical content (-0.62), and language content (-0.56). In other words, the genders' relative advantages are largely constant across tests: Subtracting females' overall score from their average subject scores, we obtain -0.26 for CS math and -0.21 for the quantitative part of the SAT; and we obtain 0.08 for CS Swedish and English, and 0.19 for the verbal part of the SAT. These results are shown in Figure A3. There is some convergence in the genders' relative advantages over time. Gender gaps also flip, both overall and within subject areas, between CSGPA and SAT score when using a matched sample of individuals: with ob-

[16] All figures mentioned in the text are in units of standard deviations, unless noted otherwise.

servations on CSGPA; non-missing HSGPA; and who took the SAT 1–6 years after graduating compulsory school. These results are shown in Figures A4–A5.

Finally, there is also a sizeable gender gap in within-individual score differences. For each student in the ETF sample who took the SAT at least once, we calculate the difference between their first SAT score and their CSGPA, and between the SAT score and the HSGPA (see Table A3).[17] This difference is 0.5 less on average for females than males, in both the SAT-CSGPA and SAT-HSGPA comparisons. Put differently, a female student's position in the score distribution deteriorates by half a standard deviation, relative to males, when moving from school grades to the SAT.

In the following sections, we test for two potential explanations for the flipping gender gap across GPAs and SAT scores. First, we investigate whether GPAs reflect different individual attributes than SAT scores, and if so, whether gender differences in the endowments of these attributes are large enough to account for meaningful portions of the gaps. Second, we ask whether gender differences in attributes among SAT takers arise due to non-random selection into taking the test, and calculate what the gender gap in SAT scores would be in the population, if all individuals took the test.

Before moving on, we consider a potential explanation for the flipping gender gap that is related to strategic concerns. Since the SAT score and HSGPA are substitutable when applying to university, those with a good HSGPA—who are more likely to be female—may be less motivated to study hard for the SAT. We find this an unlikely explanation for two reasons. First, it is not clear that the premise is correct. There is an incentive to do well on the SAT even for students with a good HSGPA. For instance, it would typically take a HSGPA of two standard deviations above the mean for the top score on the SAT to not expand one's available choices (Graetz, Öckert, and Skans, 2018). And second, males have substantially higher SAT scores than females also conditional on HSGPA, as noted above (Figure A2).

4.2 Different assessment forms test for different attributes, some of which are unequally distributed between the genders

Figure 2 plots standardized indices of cognitive skills and motivation, measured at age 13 against deciles of various scholastic assessments at older ages for our ETF sample, which is representative of the 1992 birth cohort. We see that both sets of attributes are strongly positively related to CSGPA and HSGPA: There is a two standard deviation difference in cognitive skills, and a one standard deviation difference in motivation, between the bottom and top decile of the CSGPA. For the HSGPA, the differences between bottom and top are 1.5 for cognitive skills, and 0.5 for motivation. SAT scores are even more strongly informative about cognitive skills, with a greater than two standard deviation gap between bottom and top deciles. How-

[17] We take the SAT score from the first test, but the results are largely unchanged when taking the best SAT score instead. The results are also largely unchanged when using all cohorts.

ever, motivation shows no correlation with SAT scores whatsoever. (The fact that motivation was measured about six years prior to taking the SAT is not a likely explanation for this zero-correlation, since motivation does have predictive power for the high school GPA, and most students take the SAT around the time of high school graduation.) It is also worth noting that CSGPA and HSGPA reflect the sub-components of cognitive skills (as well as motivation and effort) in a relatively uniform way. In contrast, SAT scores reflect verbal skills to a greater extent than inductive and spatial skills (see Figure A7).

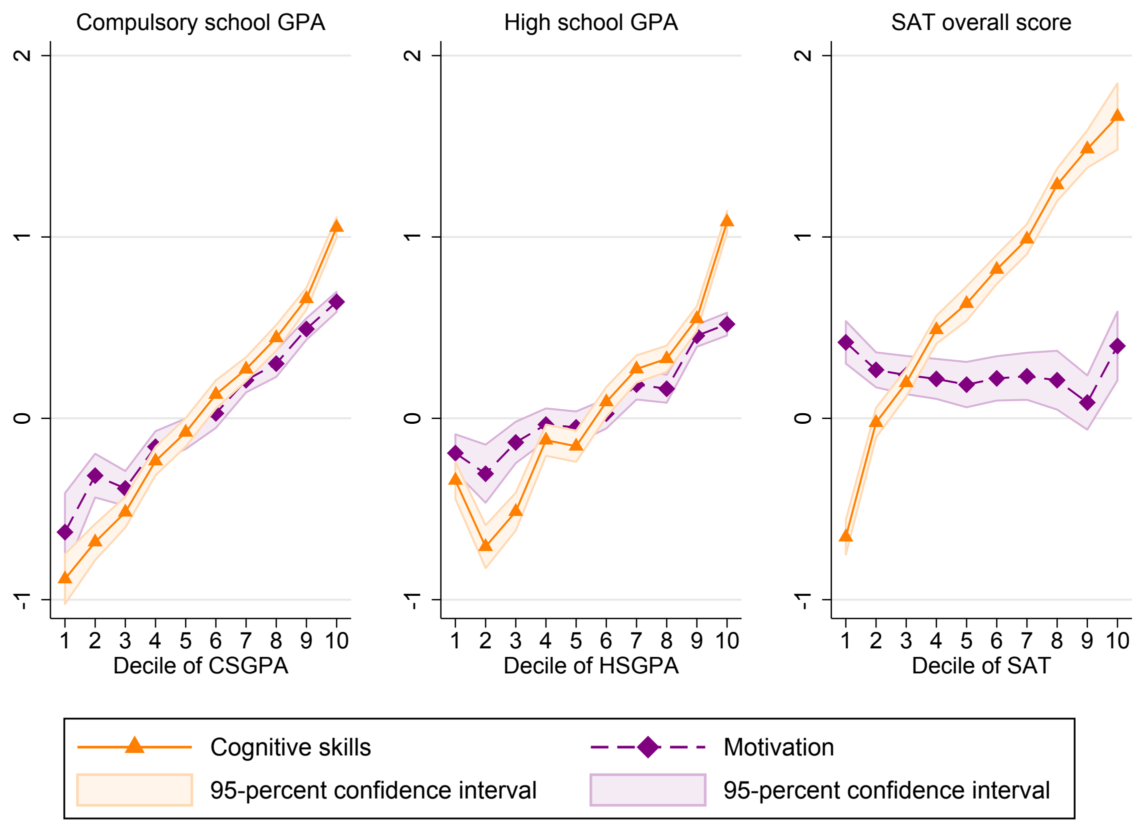


Figure 2: Test scores, cognitive skills, and motivation

The patterns in Figure 2 thus suggest that school-level assessments test for different attributes than the SAT, in particular, GPAs are highly informative of students' motivation, while SAT scores are not.[18] Table 2 shows that females exhibit higher motivation than males (and they also report spending more time on homework). The gender differences in cognitive skills are less systematic in the representative sample, although they do seem to favor males in the sub-sample of SAT takers. Taken together, this suggests that differences in the attributes that are tested for, together with gender differences in endowments, may go some way towards explaining the flipping gender gap. We investigate this issue in the next section.

[18]The relationships between attributes and scores shown in Figure 2 do not vary by gender, see Figure A6.

Table 2: Descriptive statistics

	All				SAT-takers			
	Female	Male	Diff.	t-stat	Female	Male	Diff.	t-stat
Cognitive skills, comp. index	0.024	-0.023	0.047	1.54	0.35	0.46	-0.11	-2.63
Inductive skills	-0.046	0.044	-0.089	-2.95	0.23	0.51	-0.28	-6.67
Spatial skills	0.13	-0.12	0.25	8.47	0.34	0.14	0.19	4.50
Synonyms skills	-0.068	0.065	-0.13	-4.40	0.21	0.44	-0.23	-5.38
Verbal opposites skills	0.058	-0.055	0.11	3.72	0.31	0.34	-0.030	-0.67
Motivation, comp. index	0.17	-0.16	0.34	11.2	0.41	0.076	0.34	8.42
Motivation (general)	0.15	-0.14	0.30	9.85	0.38	0.072	0.31	7.62
Motivation (school)	0.23	-0.22	0.45	15.4	0.44	0.064	0.38	9.80
Time spent on homework	0.21	-0.20	0.41	13.8	0.47	0.011	0.46	9.94
Mother graduated university	0.22	0.23	-0.0069	-0.55	0.29	0.34	-0.046	-2.17
Father graduated university	0.16	0.15	0.0024	0.22	0.22	0.25	-0.025	-1.31
Took SAT	0.41	0.33	0.082	5.64	1	1	0	
Observations	4,351				1,940			

4.3 Gender differences in cognitive skills, motivation, and effort help explain gender gap variation across assessment forms

We first regress GPAs and SAT scores on only a female dummy, which shows that the gender gaps are very similar in this sample of 1992-born students to the gaps we see in the population and over time in Figure 1 (see columns (1), (5), and (9) of Table 3). Next, we add four components of cognitive skills on the right-hand side. In terms of standardized coefficients, inductive skills are most predictive of both CSGPA and HSGPA, with the other three components spatial, synonyms, and verbal opposites skills being less than half as important, although they still have predictive power (columns (2) and (6) of Table 3). Given the lack of systematic gender differences in these attributes, adding cognitive skills does not affect the gender gap in GPAs.

We now turn to motivation and effort, the latter being measured as time spent on homework. These variables are highly predictive of CSGPA and HSGPA, as shown in columns (3) and (7) of Table 3. This, together with the fact that females exhibit higher motivation and effort, implies that these measures account for a large fraction of the gender gaps: over 60 percent in the case of CSGPA, and one third in the case of HSGPA.

A summary index of cognitive skills suggests no substantial differences in their importance across GPAs and SAT scores (Figure 2). However, there are such differences when it comes to the sub-components. Similar to the unconditional relationships discussed in Section 4.2, the most predictive for SAT scores among them are synonyms skills and verbal opposites skills, followed by inductive skills; spatial skills are relatively unimportant. But among SAT takers in this sample, spatial skills is the only component of cognitive abilities where females outperform males (Table 2). Taken together, the four components of cognitive skills actually account for more than 40 percent of the gender gap in SAT scores (column (10) of Table 3). However, general motivation and effort are slightly negatively correlated with SAT scores, while school-specific motivation is slightly positively correlated. Overall, adding these variables has no effect on the gender gap in SAT scores (column (11) of Table 3).

The conclusions that motivation and effort help explain the gender gaps in GPAs, and cognitive skills help explain the gender gap in SAT scores do not change when we enter these sets of variables jointly, as shown in columns (4), (8), and (12) of Table 3.

We also explore to what extent subject-level gender gaps can be explained by gender differences in cognitive skills, motivation, and effort. Differences in motivation and effort account for all of the (modest) female advantage in CS math, and half of the female advantage in languages. Gender differences in cognitive skills among SAT takers account for 20 percent of the male advantage in the quantitative part, and for nearly all the male advantage in the verbal part. As with overall scores, cognitive skills do not help explain gender differences in CS subject grades, while motivation and effort do not help explain gender gaps in the two parts of the SAT (see Table A4).

Finally, we find that the gender gap in within-individual score differences is accounted for in part by cognitive skills, motivation, and effort. These variables are capable of explaining one third of the gender gap in the CSGPA-SAT comparison, and one fifth of the gap in the HSGPA-SAT comparison (Table A3).

Table 3: Gender gaps, cognitive skills, and motivation

	Compulsory school GPA				High school GPA				SAT			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	0.32	0.32	0.12	0.14	0.33	0.33	0.22	0.24	-0.29	-0.17	-0.29	-0.19
	(0.045)	(0.038)	(0.042)	(0.035)	(0.044)	(0.039)	(0.044)	(0.040)	(0.051)	(0.033)	(0.055)	(0.035)
Inductive skills		0.29		0.25		0.20		0.18		0.21		0.20
		(0.023)		(0.020)		(0.024)		(0.023)		(0.021)		(0.021)
Spatial skills		0.11		0.12		0.064		0.074		0.072		0.071
		(0.022)		(0.019)		(0.023)		(0.022)		(0.019)		(0.019)
Synonyms skills		0.13		0.12		0.088		0.083		0.29		0.29
		(0.033)		(0.028)		(0.032)		(0.031)		(0.027)		(0.027)
Verbal opposites skills		0.14		0.15		0.12		0.13		0.26		0.25
		(0.030)		(0.027)		(0.030)		(0.029)		(0.023)		(0.024)
Motivation (general)			0.21	0.24			0.12	0.14			-0.098	-0.010
			(0.029)	(0.021)			(0.029)	(0.025)			(0.038)	(0.020)
Motivation (school)			0.10	0.027			0.058	0.0050			0.17	0.065
			(0.028)	(0.023)			(0.029)	(0.025)			(0.033)	(0.023)
Time spent on homework			0.22	0.21			0.12	0.12			-0.072	-0.018
			(0.020)	(0.016)			(0.021)	(0.019)			(0.022)	(0.015)
R-squared	0.03	0.33	0.22	0.49	0.03	0.17	0.08	0.21	0.03	0.53	0.06	0.53
Observations (raw—weighted)	4,351—123,668				4,114—113,194				1,940—45,478			

Notes: The dependent variables are standardized test scores as indicated in the column headings. All right-hand side variables, except the female dummy, are standardized. Regressions are weighted using sampling weights. Robust standard errors in parentheses.

4.4 Non-random selection into taking the SAT accounts for a large part of the SAT gender gap

Table 2 shows that females are 8 percentage points more likely to take the SAT. Much of this gap is explained by gender differences in motivation and effort, both of which are positive predictors of SAT participation.[19] Moreover, the distribution of cognitive skills among test takers differs from that in the full sample, with male test takers exhibiting relatively higher skills; in contrast, the gender gap in measures of motivation is nearly identical among test takers to that in the full sample (Table 2). Here we ask how much of the SAT gender gap is due to selection on cognitive skills into taking the test.

Using the regression whose results are reported in column (12) of Table 3, but shutting off the female dummy, we predict the SAT score for all individuals in our representative sample. We then compare the predicted values across genders. Among test takers, the predicted gender gap is -0.10, the same as the difference between the gender gap in the regression without any covariates except the female dummy (column (9) of Table 3) and the gap in the regression including all the covariates (column (12) of Table 3).[20] But the predicted gender gap in the full sample is small and positive at 0.01 (robust standard error: 0.03). Thus, the explanatory power of cognitive skills for the SAT gender gap is entirely due to the fact that males select into taking the test according to the attributes that predict performance more systematically than females. If all individuals took the SAT, we would not predict a gender gap in scores based on differences in cognitive skills (and less importantly, motivation and effort). Of course, even when controlling for individual attributes the SAT gender gap is still sizeable at -0.19 (column (12) of Table 3). We cannot say whether this unexplained component would be different in the full sample.

The non-random selection into taking the SAT raises an additional concern about the lack of a positive relationship between SAT scores and measures of motivation: The absence of such a relationship may be due to the fact that test takers are positively selected on motivation and cognitive skills (recall though that the positive selection on motivation does not differ by gender), combined with non-linearities or limited support. One way to assess the relevance of this concern is to re-run our regressions of GPAs on skills, motivation, and effort within the sample of SAT takers. In fact, there are positive relationships between the GPAs and motivation also in this sub-sample, and the gender gaps conditional on motivation, effort, and cognitive skills are very similar in this sub-sample to the ones in the full sample (compare columns (3)-

[19]Table A5 shows that the gap drops from 8 percentage points to 2pp when controlling for motivation and effort, although it increases slightly to 3pp when also controlling for cognitive skills. Figure A8 indicates that the relationships between participation and composite indices of motivation and cognitive skills do not vary much by gender. We have also re-estimated the regressions reported in Table A5 interacting allowing the slopes of all right-hand side variables to differ by gender. The interaction terms were never statistically significant.

[20]This is not a coincidence. The explained component from a Oaxaca-Blinder decomposition is analytically identical to the difference between short and long regression coefficients, where the short regression gives the gap that is being decomposed, and the long regression gives the gap conditional on the covariates used in the decomposition.

(4) and (7)-(8) across Tables 3 and A6). What is different in the sample of SAT takers is that, when only controlling for cognitive skills, the gender gap in GPAs actually increases relative to the raw gap. This is expected given the differential-by-gender selection on cognitive skills into taking the SAT. The latter observation also aids the interpretation of our results on within-individual score differences discussed in Section 4.3 (Table A3).

5 Discussion

In this paper we document a flipped gender gap between school-level assessments and standardized achievement test scores using Swedish longitudinal administrative data. Females outperform males on cumulative compulsory school GPA and on high school GPA, by about a third of a standard deviation in both cases. At the same time females under-perform by about a third of a standard deviation in the Swedish SAT. Our results suggest that differences in the endowments of non-cognitive traits—in particular, motivation and effort—account for a sizeable portion of the female advantage in school performance. In contrast, motivation has no predictive power for SAT scores. Turning to cognitive skills, we account for 40 percent of the male advantage in SAT scores by observing gender gaps in the endowments of inductive, spatial, and verbal skills among SAT takers. The latter can, however, be fully explained by differential self-selection into taking the SAT across the genders. Taken together, our findings show that school-level assessments and standardized achievement tests capture different skills, and that differences in endowments of skills and selection effects go a long way towards explaining gender gaps in school- and test-performance.

Our findings have implications for the issues of measurement of university preparedness and design of university admission systems. Countries or institutions can choose between admissions being based on standardized test results, on school grades, or on a combination of the two. And in the latter case, they can choose between combining different scores into a composite measure or creating separate quotas for each measure. To inform the choice of which measures to emphasize, it is important to know which skills and traits they reflect, and hence what the potential distributional impacts of choosing different measures are.

How, then, would a change in admission criteria at Swedish universities affect the composition of students enrolled? This is a difficult question, because students may change how they allocate their effort across different assessments in response. If admissions were solely based on SAT scores, for instance, the distribution of gender, cognitive skills, and motivation conditional on SAT scores might look very different from what we find in this paper, because a much broader population would take the SAT. However, in the opposite scenario, where admissions are solely based on HSGPA, selection concerns are less severe, and our findings should be quite informative. This is because under the current system, all students must graduate high school to be eligible for university.[21] We thus conclude, based on our results, that abolishing SAT scores as an admission criterion would lead to an increase in the fraction of females, an increase in motivation, and a decrease in cognitive skills (because HSGPA does not reflect cognitive skills as strongly as SAT scores), among students enrolled in university.

To get a sense of the magnitudes involved, consider a stylized scenario where competitive programs are filled by drawing randomly from the top two deciles of the score distributions. Start with the case where HSGPA and SAT are weighted equally, as is the spirit of the current system. Among admitted students, the fraction female would be 0.55, and the indices of motivation and cognitive skills would be 0.34 and 1.18 on average, respectively. If HSGPA were the sole criterion, the fraction female would increase to 0.68, motivation would increase to 0.49, and cognitive skills would decrease to 0.8. These figures are suggestive of large effects of the choice of admission criterion on the characteristics of admitted students, at least in competitive, oversubscribed programs.

How would university preparedness be affected if either HSGPA or SAT scores were more strongly emphasized in admissions? When we regress the probability of having graduated from university by age 30 on GPAs and SAT scores, we find that the HSGPA is a much stronger predictor of graduation than SAT scores, with the ratio of standardized coefficients equaling three among females and four among males.[22] Based on this evidence, we tentatively conclude that putting more emphasis on HSGPA at the expense of SAT scores might lead to greater preparedness among admitted students.[23]

[21]One may still worry that some students allocate less effort towards achieving a good HSGPA if there is also the option of using SAT scores when applying to university. However, we have documented that the HSGPA reflects measures of cognitive skills, motivation, and effort in a very similar way to the CSGPA, where such strategic concerns could not play any role.

[22]Educational attainment is measured in 2014, and the sample includes the population of Swedish residents born between 1977-1984. These results are shown in Table A7.

[23]We also find that GPAs are much stronger predictors of earnings than SAT scores, see Table A7.

References

AKYOL, Ş. P., J. KEY, AND K. KRISHNA (2016): "Hit or Miss? Test Taking Behavior in Multiple Choice Exams," Discussion Paper 22401, National Bureau of Economic Research.

ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, AND T. KAUTZ (2011): "Personality psychology and economics," in *Handbook of the Economics of Education*, vol. 4, pp. 1–181. Elsevier.

BALDIGA, K. (2013): "Gender differences in willingness to guess," *Management Science*, 60(2), 434–448.

BERG, P., O. PALMGREN, AND B. TYREFORS (2019): "Gender Grading Bias in Junior High School Mathematics," Working Paper Series 1263, Research Institute of Industrial Economics.

BOLGER, N., AND T. KELLAGHAN (1990): "Method of measurement and gender differences in scholastic achievement," *Journal of Educational Measurement*, 27(2), 165–174.

CARD, D., AND A. A. PAYNE (2017): "High school choices and the gender gap in STEM," Discussion paper, National Bureau of Economic Research.

CORNWELL, C., D. B. MUSTARD, AND J. VAN PARYS (2013): "Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school," *Journal of Human Resources*, 48(1), 236–264.

DEE, T. S. (2005): "A teacher like me: Does race, ethnicity, or gender matter?," *American Economic Review*, 95(2), 158–165.

——— (2007): "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, 42(3), 528–554.

DIAMOND, R., AND P. PERSSON (2016): "The long-term consequences of teacher discretion in grading of high-stakes tests," Discussion paper, National Bureau of Economic Research.

DUCKWORTH, A. L., AND M. E. SELIGMAN (2006): "Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores.," *Journal of educational psychology*, 98(1), 198.

DYNARSKI, S. (2017): "Make Everyone Take the SAT or ACT. And Make It Free," *The New York Times*.

EDWARDS, D., H. COATES, AND T. FRIEDMAN (2012): "A survey of international practice in university admissions testing," *Higher Education Management and Policy*, 24(1), 1–18.

FALCH, T., AND L. R. NAPER (2013): "Educational evaluation schemes and gender gaps in student achievement," *Economics of Education Review*, 36, 12–25.

FORTIN, N. M., P. OREOPOULOS, AND S. PHIPPS (2015): "Leaving boys behind: Gender disparities in high academic achievement," *Journal of Human Resources*, 50(3), 549–579.

FRYER JR, R. G., AND S. D. LEVITT (2010): "An empirical analysis of the gender gap in mathematics," *American Economic Journal: Applied Economics*, 2(2), 210–40.

GEVREK, Z. E., C. NEUMEIER, AND D. GEVREK (2018): "Explaining the Gender Test Score Gap in Mathematics: The Role of Gender Inequality," Discussion Paper 11260, IZA Discussion Papers.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 118(3), 1049–1074.

GOLDIN, C., L. F. KATZ, AND I. KUZIEMKO (2006): "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic perspectives*, 20(4), 133–156.

GRAETZ, G., B. ÖCKERT, AND O. N. SKANS (2018): "College admission opportunities and educational outcomes," working paper.

GUISO, L., F. MONTE, P. SAPIENZA, AND L. ZINGALES (2008): "Culture, gender, and math," *Science*, 320(5880), 1164.

HINNERICH, B. T., E. HÖGLIN, AND M. JOHANNESSON (2011): "Are boys discriminated in Swedish high schools?," *Economics of Education review*, 30(4), 682–690.

HOLMLUND, H., AND K. SUND (2008): "Is the gender gap in school performance affected by the sex of the teacher?," *Labour Economics*, 15(1), 37–53.

JURAJDA, Š., AND D. MÜNICH (2011): "Gender gap in performance under competitive pressure: Admissions to Czech universities," *American Economic Review*, 101(3), 514–18.

KERSEY, A. J., E. J. BRAHAM, K. D. CSUMITTA, M. E. LIBERTUS, AND J. F. CANTLON (2018): "No intrinsic gender differences in children's earliest numerical abilities," *npj Science of Learning*, 3(1), 12.

LAVY, V. (2008): "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment," *Journal of Public Economics*, 92(10-11), 2083–2105.

LINDAHL, E. (2016): "Are teacher assessments biased?–Evidence from Sweden," *Education Economics*, 24(2), 224–238.

LINDQVIST, E., AND R. VESTMAN (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3(1), 101–128.

MACHIN, S., AND T. PEKKARINEN (2008): "Global sex differences in test score variability.," *Science*.

MURNANE, R. J. (2013): "US high school graduation rates: Patterns and explanations," *Journal of Economic Literature*, 51(2), 370–422.

NIEDERLE, M., AND L. VESTERLUND (2007): "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 122(3), 1067–1101.

OECD (2017): *Education at a Glance 2017: OECD Indicators*. OECD Publishing, Paris.

ORS, E., F. PALOMINO, AND E. PEYRACHE (2013): "Performance gender gap: does competition matter?," *Journal of Labor Economics*, 31(3), 443–499.

PEKKARINEN, T. (2015): "Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations," *Journal of Economic Behavior & Organization*, 115, 94–110.

POPE, D. G., AND J. R. SYDNOR (2010): "Geographic variation in the gender differences in test scores," *Journal of Economic Perspectives*, 24(2), 95–108.

PUHANI, P. A. (2018): "Do boys benefit from male teachers in elementary school? Evidence from administrative panel data," *Labour Economics*, 51(C), 340–354.

ROTHSTEIN, J. M. (2004): "College performance predictions and the SAT," *Journal of Econometrics*, 121(1-2), 297–317.

SCB (2017): "Young people left behind? The situation on the labour market for those born in the 1990s without completed upper secondary education," Discussion paper, Statistics Sweden.

SOU (2004): *Tre vägar till den öppna högskolan*, no. SOU: 2004:29 in Statens Offentliga Utredningar. Government of Sweden.

TERRIER, C. (2016): "Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement," IZA Discussion Papers 10343, Institute for the Study of Labor (IZA).

Appendix figures and tables

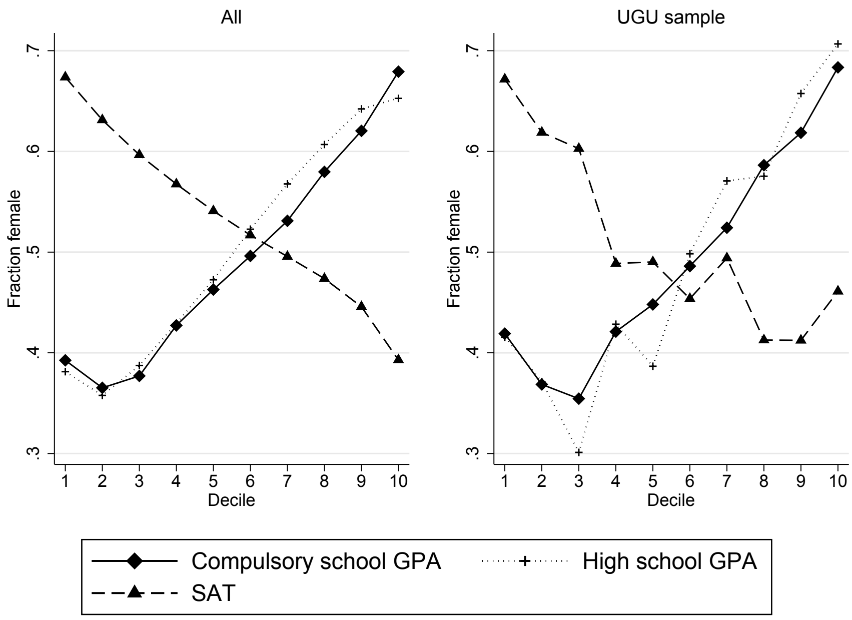


Figure A1: Fraction female by score decile, across assessment forms

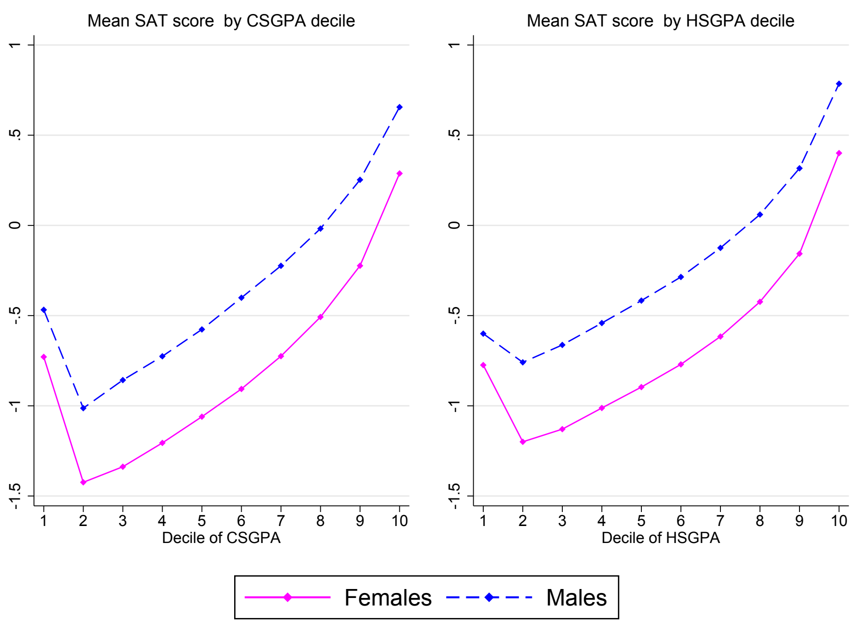
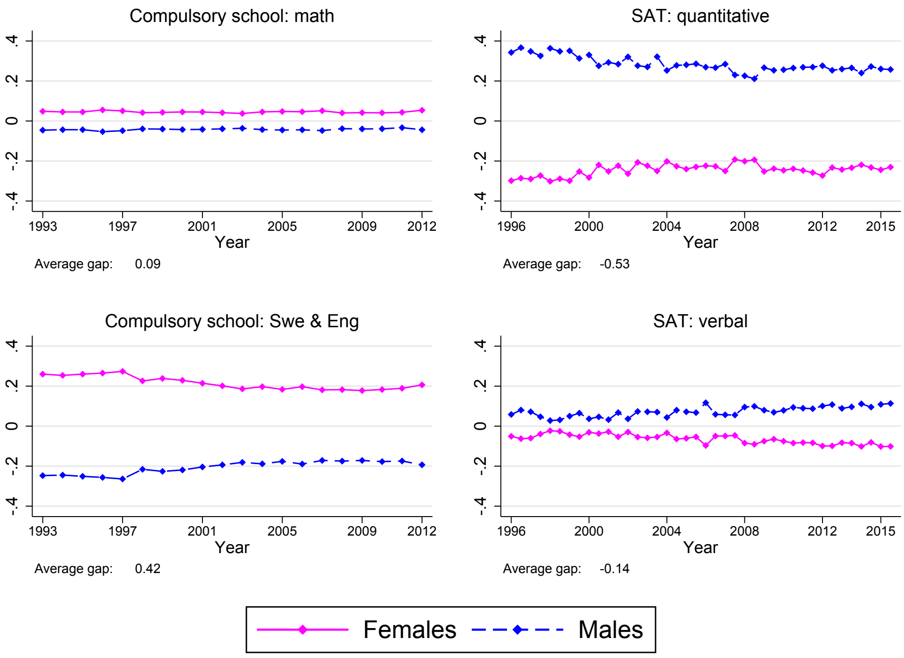
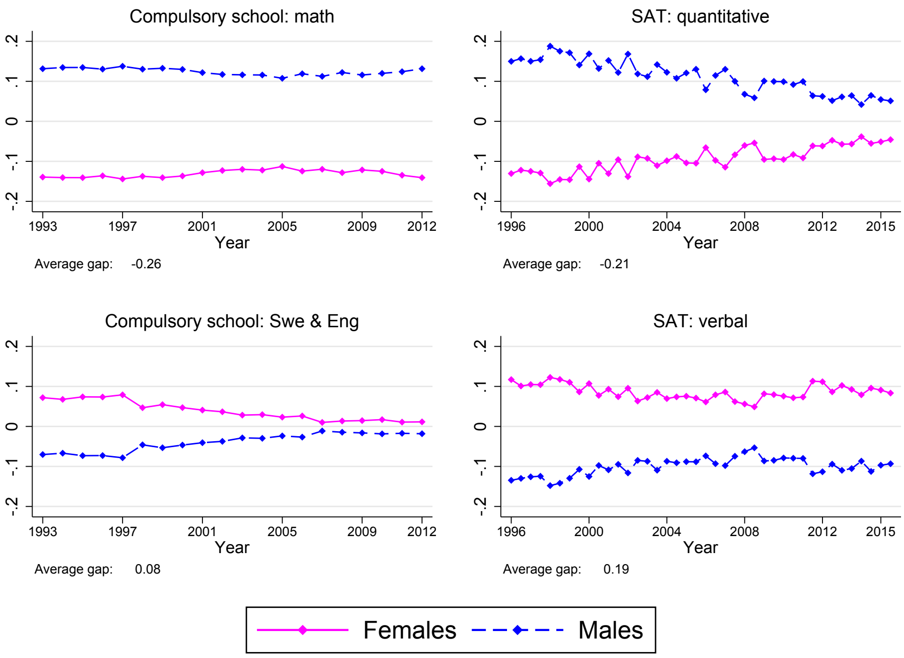


Figure A2: Average SAT score by gender and decile of GPA



(a) Standardized subject scores



(b) Relative standardized subject scores

Notes: Relative standardized subject scores are defined as the difference between the within-group mean subject score and the within-group mean overall score.

Figure A3: Standardized subject scores over time, by assessment form and gender

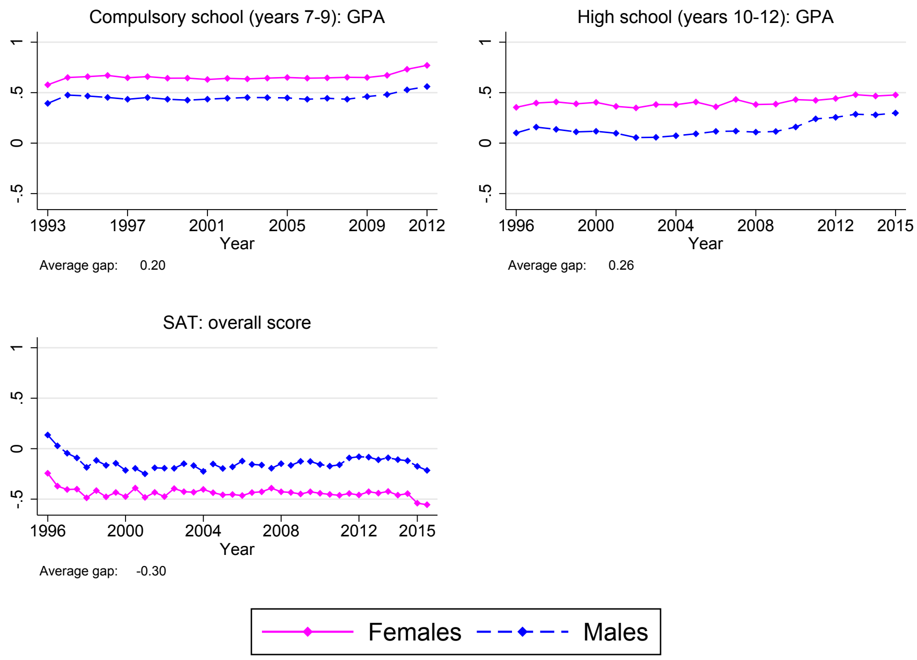
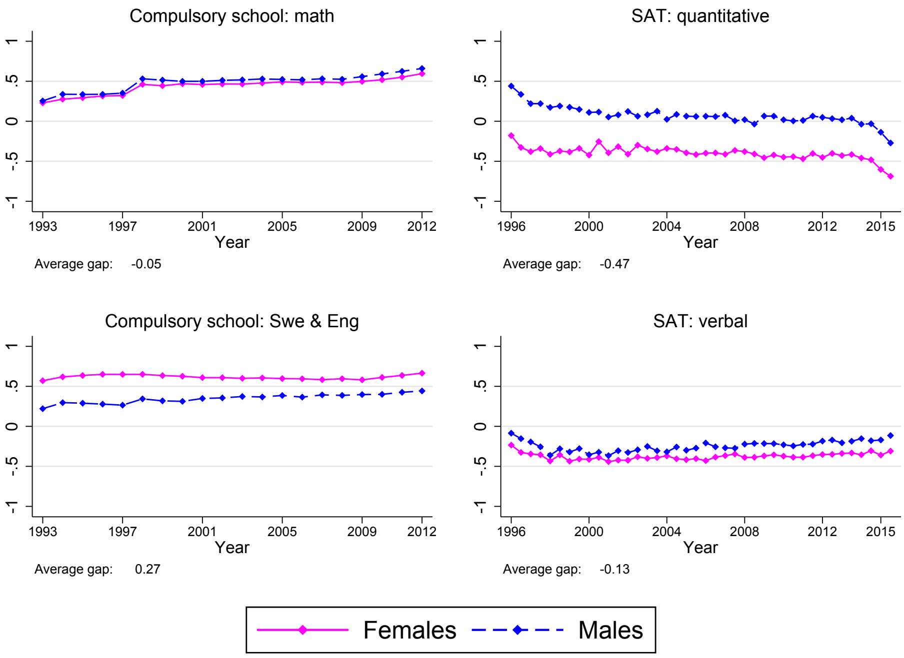
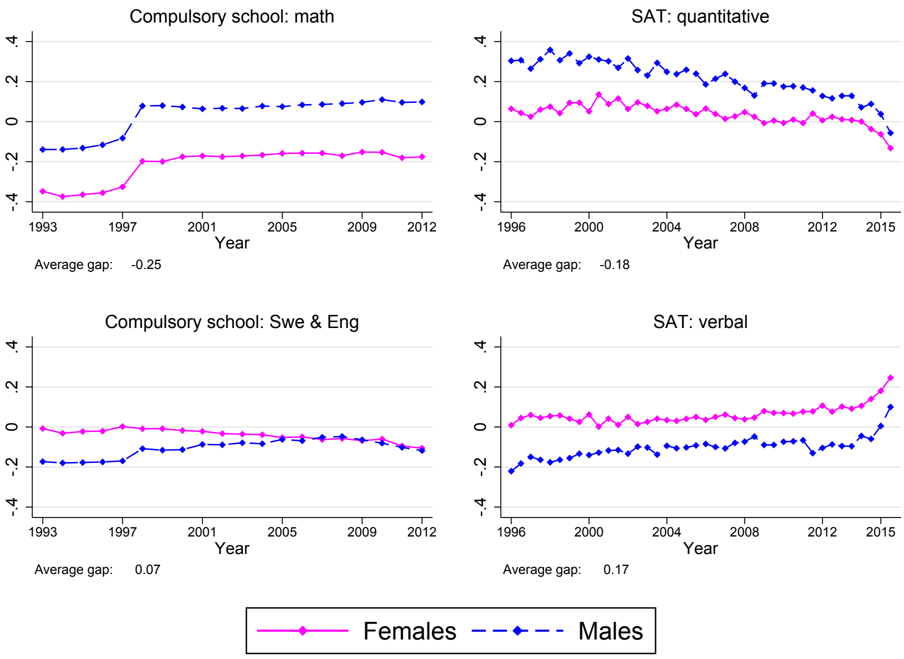


Figure A4: Standardized scores over time, by assessment form and gender—matched GPA-SAT sample



(a) Standardized subject scores



(b) Relative standardized subject scores

Notes: Relative standardized subject scores are defined as the difference between the within-group mean subject score and the within-group mean overall score.

Figure A5: Standardized subject scores over time, by assessment form and gender—matched GPA-SAT sample

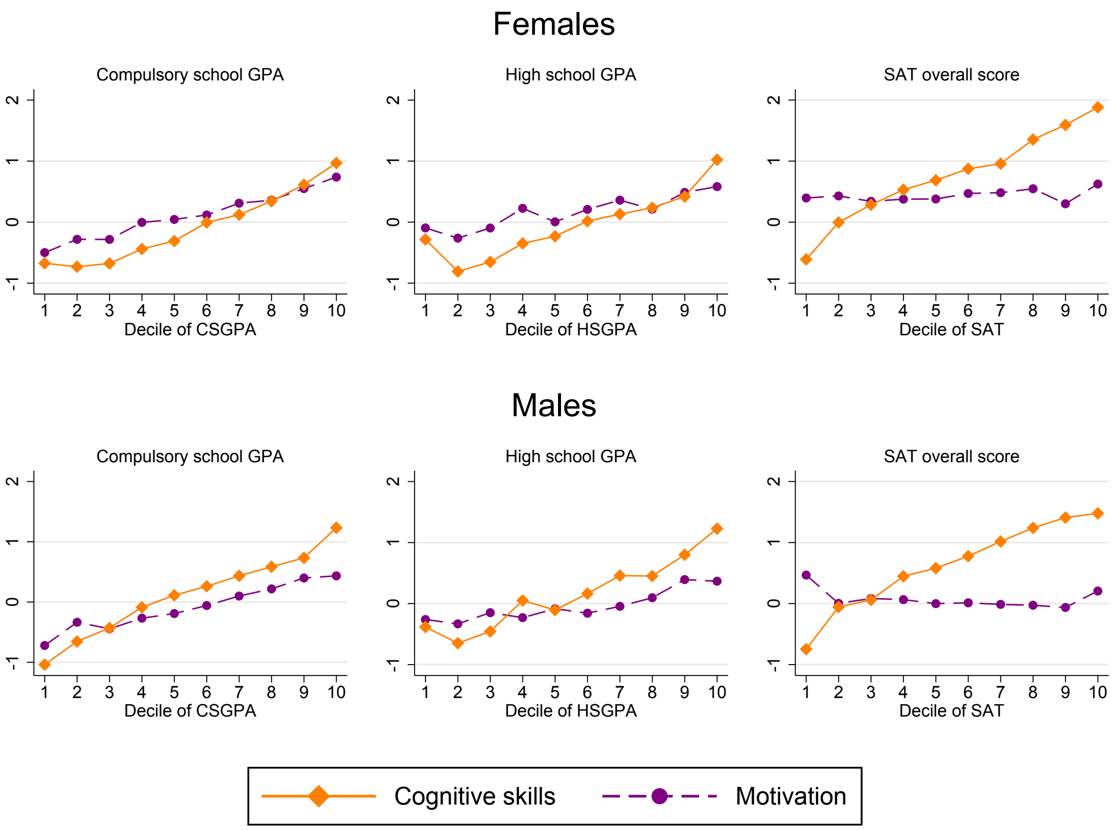


Figure A6: Test scores, cognitive skills, and motivation by gender

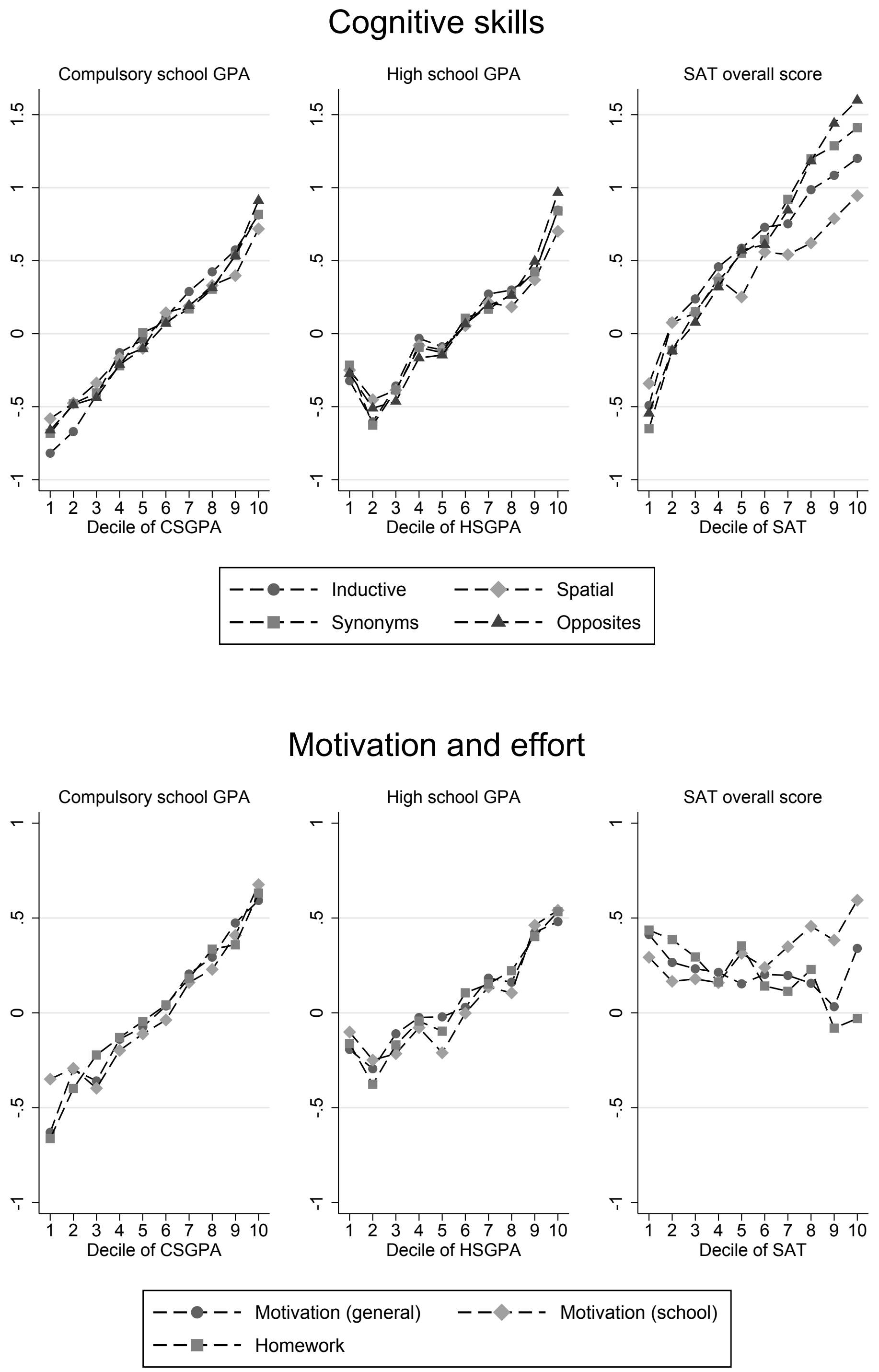


Figure A7: Test scores, detailed cognitive skills, and detailed measures of motivation and effort

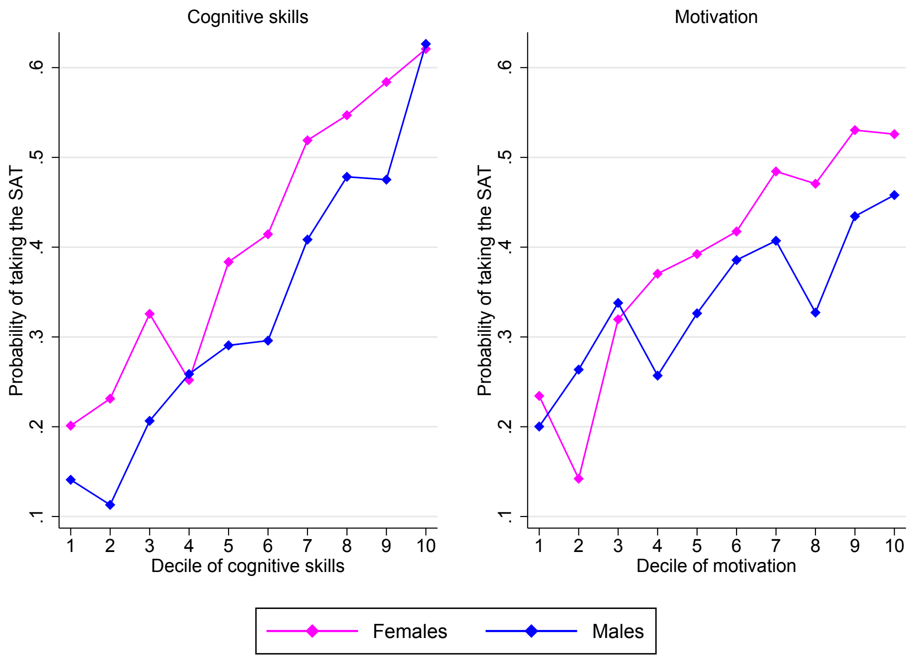


Figure A8: SAT participation, cognitive skills, and motivation

Table A1: Standardized scores over time

	CSGPA			HSGPA			SAT	
Year	Females	Males	Year	Females	Males	Year	Females	Males
1993	0.19 (0.004)	-0.18 (0.004)	1996	0.20 (0.005)	-0.20 (0.005)	1996	-0.17 (0.005)	0.19 (0.005)
1994	0.19 (0.005)	-0.18 (0.005)	1997	0.17 (0.005)	-0.17 (0.005)	1997	-0.16 (0.005)	0.20 (0.005)
1995	0.19 (0.004)	-0.18 (0.004)	1998	0.18 (0.005)	-0.19 (0.005)	1998	-0.15 (0.005)	0.18 (0.005)
1996	0.19 (0.004)	-0.18 (0.004)	1999	0.18 (0.005)	-0.19 (0.005)	1999	-0.15 (0.005)	0.18 (0.006)
1997	0.19 (0.004)	-0.19 (0.004)	2000	0.20 (0.005)	-0.20 (0.005)	2000	-0.14 (0.005)	0.16 (0.006)
1998	0.18 (0.005)	-0.17 (0.004)	2001	0.18 (0.005)	-0.19 (0.005)	2001	-0.12 (0.006)	0.14 (0.007)
1999	0.18 (0.005)	-0.17 (0.004)	2002	0.19 (0.005)	-0.20 (0.005)	2002	-0.12 (0.007)	0.15 (0.007)
2000	0.18 (0.004)	-0.17 (0.004)	2003	0.22 (0.005)	-0.22 (0.005)	2003	-0.13 (0.007)	0.16 (0.007)
2001	0.17 (0.004)	-0.16 (0.004)	2004	0.21 (0.005)	-0.21 (0.005)	2004	-0.10 (0.006)	0.13 (0.007)
2002	0.16 (0.004)	-0.16 (0.004)	2005	0.21 (0.005)	-0.21 (0.005)	2005	-0.14 (0.006)	0.16 (0.007)
2003	0.16 (0.004)	-0.15 (0.004)	2006	0.18 (0.005)	-0.19 (0.005)	2006	-0.16 (0.007)	0.19 (0.007)
2004	0.17 (0.004)	-0.16 (0.004)	2007	0.22 (0.005)	-0.22 (0.005)	2007	-0.14 (0.007)	0.15 (0.007)
2005	0.16 (0.004)	-0.15 (0.004)	2008	0.19 (0.005)	-0.19 (0.004)	2008	-0.14 (0.007)	0.16 (0.007)
2006	0.17 (0.004)	-0.16 (0.004)	2009	0.19 (0.004)	-0.20 (0.004)	2009	-0.16 (0.007)	0.17 (0.007)
2007	0.17 (0.004)	-0.16 (0.004)	2010	0.20 (0.004)	-0.20 (0.004)	2010	-0.15 (0.006)	0.16 (0.006)
2008	0.17 (0.004)	-0.16 (0.004)	2011	0.15 (0.004)	-0.14 (0.004)	2011	-0.16 (0.005)	0.17 (0.006)
2009	0.16 (0.004)	-0.16 (0.004)	2012	0.15 (0.004)	-0.15 (0.004)	2012	-0.21 (0.005)	0.21 (0.006)
2010	0.17 (0.004)	-0.16 (0.004)	2013	0.16 (0.004)	-0.15 (0.004)	2013	-0.19 (0.005)	0.20 (0.006)
2011	0.18 (0.004)	-0.16 (0.004)	2014	0.13 (0.005)	-0.13 (0.005)	2014	-0.18 (0.005)	0.20 (0.005)
2012	0.19 (0.004)	-0.18 (0.004)	2015	0.14 (0.005)	-0.14 (0.005)	2015	-0.19 (0.005)	0.21 (0.005)

Notes: For each year and gender, the mean of the indicated score is reported, with its standard error in parentheses.

Table A2: Correlations between scores

	All			Females			Males		
A1: Population of high-school graduates, graduation years 1996-2015 (1,726,166 observations)									
	CSGPA	HSGPA		CSGPA	HSGPA		CSGPA	HSGPA	
CSGPA	1			1			1		
HSGPA	0.64	1		0.63	1		0.63	1	
A2: Population of high-school graduates who took the SAT, graduation years 1996-2015 (633,126 observations)									
	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT
CSGPA	1			1			1		
HSGPA	0.57	1		0.55	1		0.58	1	
SAT	0.45	0.41	1	0.51	0.46	1	0.45	0.44	1
B1: Population of high-school graduates, 1992 birth cohort (110,223 observations)									
	CSGPA	HSGPA		CSGPA	HSGPA		CSGPA	HSGPA	
CSGPA	1			1			1		
HSGPA	0.56	1		0.54	1		0.56	1	
B2: Population of high-school graduates who took the SAT, 1992 birth cohort (40,021 observations)									
	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT
CSGPA	1			1			1		
HSGPA	0.45	1		0.42	1		0.48	1	
SAT	0.44	0.28	1	0.51	0.30	1	0.45	0.32	1
C1: UGU sample of high-school graduates, 1992 birth cohort (4,351 observations)									
	CSGPA	HSGPA		CSGPA	HSGPA		CSGPA	HSGPA	
CSGPA	1			1			1		
HSGPA	0.57	1		0.55	1		0.58	1	
C2: UGU sample of high-school graduates who took the SAT, 1992 birth cohort (1,940 observations)									
	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT	CSGPA	HSGPA	SAT
CSGPA	1			1			1		
HSGPA	0.48	1		0.45	1		0.50	1	
SAT	0.45	0.24	1	0.51	0.25	1	0.48	0.28	1

Notes: High school graduates whose compulsory school GPA is unknown were dropped from each sample.

Table A3: Gender gaps in within-individual score differences

	SAT minus CSGPA				SAT minus HSGPA			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.50	-0.45	-0.38	-0.33	-0.50	-0.43	-0.45	-0.39
	(0.043)	(0.045)	(0.043)	(0.049)	(0.052)	(0.052)	(0.054)	(0.054)
Inductive skills		0.022		0.019		0.071		0.068
		(0.028)		(0.031)		(0.032)		(0.032)
Spatial skills		-0.018		-0.034		0.036		0.026
		(0.023)		(0.022)		(0.028)		(0.028)
Synonyms skills		0.19		0.18		0.21		0.20
		(0.039)		(0.038)		(0.038)		(0.038)
Verbal opposites skills		0.14		0.13		0.18		0.18
		(0.033)		(0.033)		(0.033)		(0.033)
Motivation (general)			-0.21	-0.17			-0.14	-0.084
			(0.029)	(0.032)			(0.034)	(0.030)
Motivation (school)			0.038	-0.0051			0.10	0.042
			(0.029)	(0.027)			(0.033)	(0.032)
Time spent on homework			-0.15	-0.14			-0.10	-0.070
			(0.019)	(0.019)			(0.026)	(0.024)
R-squared	0.09	0.22	0.20	0.31	0.06	0.21	0.09	0.22
Observations (raw—weighted)	1,940—45,478							

Notes: The dependent variables are within-individual differences in standardized test scores as indicated in the column headings. All right-hand side variables, except the female dummy, are standardized. Regressions are weighted using sampling weights. Robust standard errors in parentheses.

Table A4: Gender gaps across assessments and subject areas

	Compulsory school: math				Compulsory school: Swe & Eng				SAT: quant				SAT: verbal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Female	0.14	0.14	-0.018	0.0064	0.39	0.41	0.22	0.26	-0.45	-0.35	-0.46	-0.38	-0.13	-0.025	-0.13	-0.028
	(0.047)	(0.038)	(0.046)	(0.038)	(0.045)	(0.037)	(0.043)	(0.036)	(0.053)	(0.037)	(0.056)	(0.039)	(0.049)	(0.036)	(0.053)	(0.038)
Inductive skills		0.38		0.35		0.25		0.21		0.36		0.36		0.055		0.052
		(0.023)		(0.022)		(0.024)		(0.023)		(0.023)		(0.023)		(0.023)		(0.023)
Spatial skills		0.15		0.16		0.076		0.083		0.16		0.17		-0.0038		-0.0056
		(0.021)		(0.020)		(0.022)		(0.020)		(0.021)		(0.021)		(0.020)		(0.020)
Synonyms skills		0.088		0.082		0.23		0.22		0.091		0.088		0.37		0.37
		(0.034)		(0.031)		(0.034)		(0.030)		(0.027)		(0.027)		(0.031)		(0.032)
Verbal opposites skills		0.13		0.14		0.19		0.20		0.17		0.17		0.27		0.27
		(0.030)		(0.028)		(0.032)		(0.030)		(0.025)		(0.025)		(0.026)		(0.026)
Motivation (general)			0.12	0.15			0.17	0.18			-0.084	-0.019			-0.093	-0.0053
			(0.033)	(0.024)			(0.029)	(0.022)			(0.040)	(0.023)			(0.034)	(0.022)
Motivation (school)			0.13	0.043			0.15	0.054			0.16	0.077			0.14	0.042
			(0.032)	(0.026)			(0.029)	(0.024)			(0.034)	(0.025)			(0.032)	(0.024)
Time spent on homework			0.17	0.14			0.14	0.13			-0.053	0.0038			-0.067	-0.026
			(0.023)	(0.018)			(0.023)	(0.018)			(0.023)	(0.018)			(0.023)	(0.017)
R-squared	0.01	0.36	0.10	0.43	0.04	0.39	0.16	0.48	0.06	0.45	0.08	0.45	0.01	0.46	0.03	0.46
Observations		4,351—123,668				4,270—118,395				1,940—45,478				1,940—45,478		

Notes: The dependent variables are standardized subject test scores as indicated in the column headings. All right-hand side variables, except the female dummy, are standardized. Regressions are weighted using sampling weights. Robust standard errors in parentheses.

Table A5: Gender gaps in SAT participation

	(1)	(2)	(3)	(4)
Female	0.082	0.086	0.019	0.030
	(0.017)	(0.017)	(0.017)	(0.017)
Inductive skills		0.088		0.076
		(0.011)		(0.010)
Spatial skills		0.018		0.022
		(0.0090)		(0.0087)
Synonyms skills		0.040		0.038
		(0.012)		(0.011)
Verbal opposites skills		0.042		0.042
		(0.011)		(0.011)
Motivation (general)			0.025	0.032
			(0.011)	(0.010)
Motivation (school)			0.064	0.043
			(0.0099)	(0.0093)
Time spent on homework			0.065	0.060
			(0.0089)	(0.0084)
R-squared	0.01	0.11	0.07	0.15
Observations (raw—weighted)		4,351—123,668		

Notes: The dependent variable is an indicator for ever having taken the SAT. All right-hand side variables, except the female dummy, are standardized. Regressions are weighted using sampling weights. Robust standard errors in parentheses.

Table A6: The explanatory power of skills and motivation within the sample of SAT takers

	Compulsory school GPA				High school GPA			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.22	0.28	0.095	0.14	0.21	0.26	0.16	0.20
	(0.046)	(0.042)	(0.053)	(0.046)	(0.046)	(0.048)	(0.052)	(0.051)
Inductive skills		0.18		0.18		0.13		0.13
		(0.024)		(0.027)		(0.031)		(0.031)
Spatial skills		0.090		0.10		0.036		0.045
		(0.019)		(0.017)		(0.023)		(0.023)
Synonyms skills		0.097		0.11		0.081		0.085
		(0.044)		(0.044)		(0.037)		(0.037)
Verbal opposites skills		0.12		0.12		0.075		0.080
		(0.034)		(0.033)		(0.031)		(0.030)
Motivation (general)			0.11	0.16			0.044	0.078
			(0.045)	(0.029)			(0.030)	(0.026)
Motivation (school)			0.13	0.070			0.065	0.025
			(0.028)	(0.023)			(0.030)	(0.029)
Time spent on homework			0.082	0.12			0.029	0.054
			(0.018)	(0.015)			(0.022)	(0.021)
R-squared	0.02	0.28	0.12	0.40	0.02	0.11	0.03	0.13
Observations (raw—weighted)	1,940—45,478							

Notes: This table replicates columns (1)-(8) of Table 3 for the sample of SAT takers. All right-hand side variables, except the female dummy, are standardized. Regressions are weighted using sampling weights. Robust standard errors in parentheses.

Table A7: School GPAs and SAT scores as predictors for adult earnings and university graduation

	Graduated university						2014 annual earnings, '000SEK						Percentile rank in 2014 annual earnings					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
A. Females (all: 367,138 observations, restricted: 124,556 observations)																		
CSGPA	0.25	0.17			0.06		45.26	43.24			19.81		8.02	5.90			3.01	
	(0.00)	(0.00)			(0.00)		(0.29)	(0.84)			(1.10)		(0.05)	(0.12)			(0.16)	
CS, quant						0.01						4.18						0.54
						(0.00)						(0.65)						(0.11)
CS, verbal						0.02						3.73						0.35
						(0.00)						(0.81)						(0.13)
HSGPA			0.17		0.11	0.13			39.45		27.57	30.97			5.25		3.66	4.28
			(0.00)		(0.00)	(0.00)			(0.61)		(0.84)	(0.77)			(0.09)		(0.13)	(0.12)
SAT				0.11	0.04					23.24	2.70					2.92	0.06	
				(0.00)	(0.00)					(0.53)	(0.64)					(0.08)	(0.10)	
SAT, quant						0.03						16.73						2.31
						(0.00)						(0.70)						(0.12)
SAT, verbal						0.02						-8.55						-1.38
						(0.00)						(0.69)						(0.11)
R-squared	0.24	0.08	0.11	0.06	0.12	0.11	0.09	0.04	0.05	0.03	0.05	0.06	0.08	0.02	0.03	0.01	0.03	0.03
B. Males (all: 385,209 observations, restricted: 106,362 observations)																		
CSGPA	0.23	0.22			0.06		59.46	55.85			25.83		8.67	6.38			3.51	
	(0.00)	(0.00)			(0.00)		(0.40)	(1.46)			(1.82)		(0.05)	(0.14)			(0.18)	
CS, quant						0.01						2.32						0.48
						(0.00)						(1.00)						(0.12)
CS, verbal						0.01						3.73						-0.25
						(0.00)						(1.70)						(0.14)
HSGPA			0.21		0.16	0.17			51.93		48.97	53.69			5.72		5.48	6.28
			(0.00)		(0.00)	(0.00)			(1.05)		(1.43)	(1.23)			(0.10)		(0.13)	(0.13)
SAT				0.13	0.04					11.93	-18.15					0.77	-2.79	
				(0.00)	(0.00)					(0.99)	(1.36)					(0.09)	(0.11)	
SAT, quant						0.05						14.69						2.11
						(0.00)						(1.00)						(0.12)
SAT, verbal						0.01						-29.35						-4.21
						(0.00)						(1.66)						(0.12)
R-squared	0.24	0.09	0.15	0.07	0.16	0.16	0.09	0.04	0.05	0.02	0.06	0.06	0.09	0.02	0.03	0.00	0.04	0.05

Notes: The dependent variables are the indicated measures of annual labor earnings and university graduation. Columns (1), (7), and (8) include all individuals born 1977-1984 with non-missing compulsory school GPA, while all other columns restrict the sample to individuals with non-missing high school GPA and SAT scores. All regressions include cohort dummies. Robust standard errors in parentheses.