Gender differences in admission scores and first-year university achievement

Linn Karlsson* Magnus Wikström†‡

September, 2021

Abstract

This study explores female underprediction in first-year university achievement by using data from 8,971 Swedish university entrants in the fall semester of 2012. The Swedish admissions system selects students by two instruments: upper secondary school GPA or scores from a scholastic aptitude test (SweSAT). Nearest-neighbour matching allows us to compare students with similar admission scores and estimate achievement differences between male and female students. The results show that admission scores underpredict achievement for women relative to men in both admissions groups and more so for the SweSAT. As we condition on field of education, achievement differences tend to vary over fields and tend to become smaller, indicating that part of the differences is related to the male-female composition of students in the different fields.

JEL Codes: I21, I23, I24

Keywords: Swedish admissions test, grade point average, gender, female underprediction, higher education

1 Introduction

Notable gender differences in educational outcomes appear in elementary school (Figlio et al., 2016). This gender gap is also evident at the university level, where women tend to earn higher grades and also have a lower probability of dropping out (Conger & Long, 2010;

*Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, E-mail: linn.karlsson@umu.se

†Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, E-mail: magnus.wikstrom@umu.se

‡The authors would like to thank Niklas Hanes and David Granlund for helpful comments and suggestions. Research grants from the Swedish Research Council (ref. 2014-01990) and Handelsbanken (ref. P2016-0299) are gratefully acknowledged.

Jacob, 2002). Even when students are admitted with the same admission score, the gender gap in achievement persists. A sizeable body of literature addresses if and why admission scores underpredict female achievement in higher education. Fischer et al. (2013) provides a meta-analysis summarising the results from 42 different studies. The general finding is that women's academic achievement is underpredicted relative to men, but the size appears to be rather modest. In line with Duckworth and Seligman (2006), Keiser et al. (2016) and Mattern et al. (2017) suggest that part of the differential prediction between genders can be explained by an omitted variable problem, where unobserved non-cognitive abilities such as grit and self-consciousness also explain achievement. Keiser et al. (2016) also suggests that differences in course composition between men and women can explain the differential prediction.

Not all studies find a differential prediction to the advantage of men. Looking at admissions at a British elite university, Bhattacharya et al. (2017) found that men faced a higher threshold for admissions than women. Similarly, Bhattacharya and Rabovic (2020) found higher admission standards for men in STEM fields and economics but not in medicine and law. The admission practices in the British elite universities include in-house assessments with interviews and further testing, and one explanation for the observed higher admission standards for men might be related to an objective of balancing merit and diversity. They did not find any evidence that pre-admission standardised test scores are any worse for female applicants or that the relative scarcity of female students had any negative effect on achievement. Hanson (2017) also analysed admissions but focused on the response from admissions counsellors. That study found that female applicants received more positive and polite content in their responses.

The purpose of this study is to determine whether female achievement is underpredicted relative to men's according to their admissions score. We use data on 8,971 university students to analyse gender differences in admission scores and first-year achievement at university. The Swedish admission system is mainly based on two instruments that are fully observable to the researcher. Grade point average from upper secondary school (GPA) is the main instrument used in admissions. The upper secondary GPA is based on teacher assessments made throughout upper secondary school. Approximately 60% of the students entering university are admitted by the upper secondary GPA. A student can also apply with the result from a scholastic aptitude test, the SweSAT, a 160 question multiple-choice test administered twice a year and taken voluntarily. The test result is valid for five years. Admission by the SweSAT forms a second admission group, and a university must allocate at least one-third of the study positions using this instrument. We use detailed data on admissions scores and first-year higher educational achievement to analyse gender differences in academic achievement.

Previous literature has mainly analysed differential prediction using linear regression analysis. While this approach is suitable for studying gender differences on average, it is less useful to study if differential prediction exists in some particular part of the admission score distribution. We use nearest-neighbour matching, where female achievement is compared with a counterfactual male, and then perform kernel-weighted local polynomial regression on the observed achievement differences. This approach allows us to study if women's academic achievement is underpredicted and if the relationship is persistent over the whole score distribution. Using data from four different fields of specialisation, we study if underprediction is related to the gender composition of students in different fields. As we have information on admission scores in the two different admission groups, we can also relate underprediction to the different instruments. Because upper secondary GPA is based on assessments made throughout upper secondary school, one can conjecture that GPA is a more robust measure of the students' underlying abilities than the SweSAT. Additionally, if grades also to some extent reflect non-cognitive abilities and if women are better equipped in these attributes, we should observe less underprediction using GPA than the SweSAT scores.

The rest of the paper is organised as follows: Section 2 describes the Swedish admission system and the data used in the empirical analysis. Section 3 sets out the empirical strategy, Section 4 presents the results, and Section 5 discusses the findings and relates them to the previous literature.

2 Institutional Setting and Data

In the Swedish admissions system, selection to higher education is mainly based on two instruments, the upper secondary grade point average (GPA) and the Swedish scholastic aptitude test (SweSAT). GPA is a weighted average of the grades from courses in the upper secondary school, where not pass (IG) = 0, pass (G) = 10, pass with distinction (VG) = 15 and pass with particular distinction (MVG) = 20. The weights applied are calculated based on the course length relative to the total programme length. The eligibility criteria for higher education require an upper secondary diploma, meaning that students must have passing grades in basic courses, such as Swedish, English, and mathematics, and pass at least 2250 credits of 2500. Grades are criterion referenced and determined by the teacher. The GPA can vary from 0 to 20, but the studied sample of students have at least a pass grade in all courses. The implication is that the upper secondary GPA varies between 10 and 20 in our sample.[1] The Swedish Scholastic Aptitude Test (SweSAT) is a voluntary test that can be taken for a small fee[2], to improve the chances of being admitted to higher

[1]All students are required to have at least a pass grade in basic courses such as Swedish, English and mathematics for general eligibility to higher education. Depending on the educational track, there are also specific eligibility requirements.

[2]SEK 450 $\approx$ EUR 44

education. The test consists of two parts, a verbal part and a quantitative part, and is administered twice a year. A SweSAT score is valid for five years. Anyone can take the test as many times as they wish, and the best valid score counts in the admission process. SweSAT is a normed referenced test reported on a scale of 0.0 to 2.0 but is scaled from 0 to 20 in this paper for the sake of convenience.

In the application to the university, students rank their choices in a web-administered application. The students are admitted by the instrument in which they have the most competitive score. There is a positive selection into the SweSAT, meaning that high-achieving students with ambitions of being admitted to selective programmes are more likely to take the test than other students (Törnkvist & Henriksson, 2004). There is also a tendency for students who did not earn high enough grades in upper secondary school to use the SweSAT as a second chance to obtain admission into higher education.

University programmes are obligated to select at least one-third of the students based on GPA and at least one-third based on the SweSAT. This means that universities are free to choose which of the two instruments is used to select the final third. This also means that the competition can be high in the GPA group if the number of students applying with SweSAT is very low. This can, for example, be observed in social work programmes. The educational fields studied in this paper use only GPA and SweSAT as the basis for admission, meaning that all factors that influence admission are observable to the researcher.

University course grades in Sweden usually follow a scale fail, pass, and pass with distinction. Some programmes only have two levels (fail and pass), and some have more than three levels. A GPA is not calculated in the Swedish system. A degree certificate contains the programme's courses and the pass grades. Grades from university courses cannot be observed in our dataset. Instead, completed credits are used as a measure of academic achievement. We construct a measure that considers that some students study part-time by dividing each student's completed credits after one year by the enrolled credits. For example, if a student plans to enrol 60 credits (full time) but fails to complete 7.5 credits, the corresponding achievement for this student is 87.5%.

Data

The sample consists of 8,971 first-year students in the academic year 2012–2013. These students are spread across four different fields of education (engineering, business and economics, social work and law) that vary in available seats, educational content and popularity. For example, engineering, the largest field, can be taken at several universities, while law is only offered at a handful of traditional universities. Table A1 report detailed information about field of study and institutions included in the study. The majority of university students are women[3], but women are underrepresented in our sample (44%). This is mainly

[3]In 2012, 57% of the first time students were women. Source: Statistics Sweden

explained by the engineering programmes, which are dominated by men and contain a large share of the observations (N=4,619). In contrast, 85% of the social work students are women, which can be seen in Table 1. The number of men and women is more equal in the two remaining programmes (business and economics and law).

Table 1: Field of education by gender (%)

Field of education	Men	Women
Engineering	72	28
Business and Economics	50	50
Social work	15	85
Law	46	54
Pooled sample	56	44

The main variables used in the empirical analysis is the upper secondary GPA, the (best) valid SweSAT score, and first-year achievement at university. As discussed above, GPA is not calculated at the university level. The available information is if students have completed courses with (at least) a pass grade. Full-time university studies usually mean 60 credits (ECTS) per year. However, students can register for less than (or sometimes more than) 60 credits.[4] To make students achievement comparable, the share of course credits completed out of the credits registered for are calculated. This way, the first-year achievement measure is restricted to a range from zero to one. One potential problem in the Swedish data is that it is not possible to observe drop-outs because universities were not required to unregister students as they leave. Therefore, those with zero observed credits could either be unsuccessful students or those that changed programme or left early during the academic year. Therefore, we restrict the analysis to those students who completed at least one credit. Definitions of all the variables used in the empirical analysis are given in Table A2 in the appendix.

Figure 1 show the GPA and SweSAT distribution for women and men in the sample. In line with previous studies (Wikström and Wikström (2017) and Duckworth and Seligman (2006)), women have higher GPAs, but men perform better on the SweSAT.

[4] 1,536 students were registered for more than 60 credits and 1,399 were registered for less than 60 credits. A majority of these studied engineering. Excluding these students from the analysis did not change the main results.

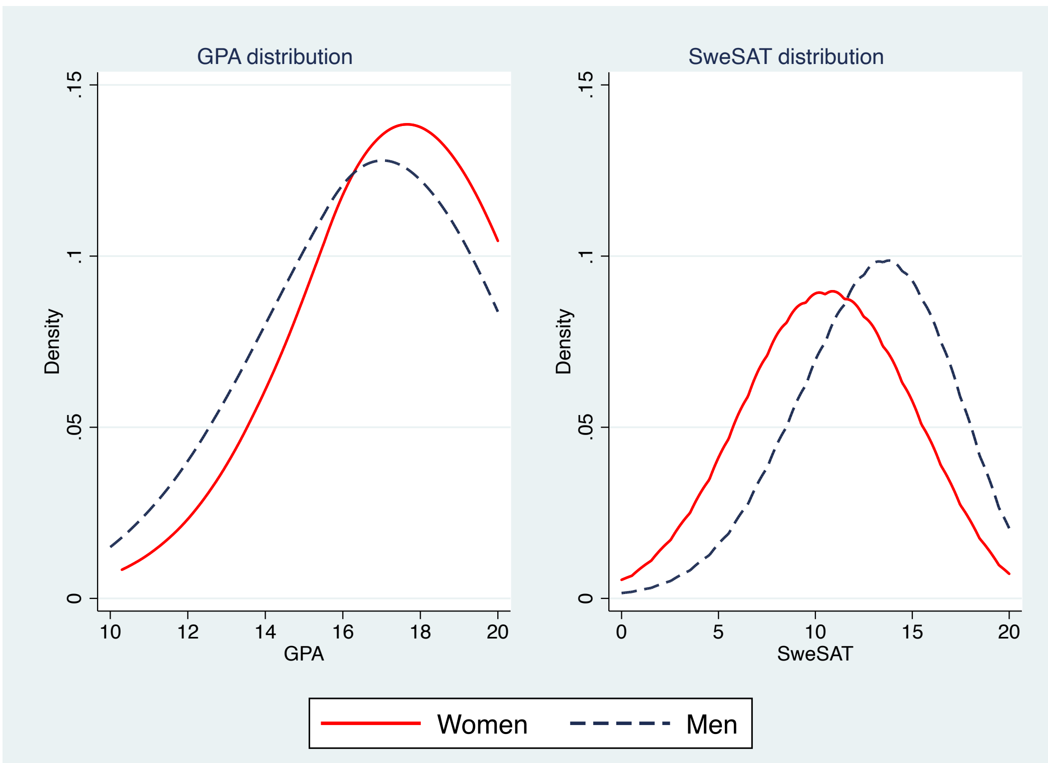

Figure 1: Kernel density estimates for GPA and SweSAT in pooled sample.

Table 2 shows more detailed information on GPA, SweSAT and achievement in the total sample and for each field of study, separately. The average achievement in the total sample is 77%, the GPA is 17, and the SweSAT is 12 on average. If we instead look at men and women separately, gender differences appear. Female students have on average 11 percentage points higher achievement than men. As seen in Figure 1, women have a higher GPA than men, but male students tend to perform better on the SweSAT. Given that 10 is the mean SweSAT score in the population, the average SweSAT score for both men and women, 12.9 and 10.5 respectively, is above the average in the population. Table 2 also displays means conditional on educational field. Women have on average higher achievement, higher GPA, and lower SweSAT scores than men in all of the four fields contained in the data.

We also note that the share with a valid SweSAT is larger among men. This holds especially in the social work programmes (where only 65% of the women has taken the SweSAT compared to 84% of the men). The difference in test-taking behaviour is smaller in engineering, where the difference is only 1%. The observed gap may reflect that women have higher grades, and their GPA is high enough for their choice of education, and thus they do not need to take the SweSAT. Previous studies have found that men and women apply to different programmes. Men apply to STEM fields while women tend to apply for programs in the social sciences or humanities (Brenøe & Zölitz, 2020; Mouganie & Wang, 2020). The skills measured by SweSAT might also seem more important for STEM educations than, for example, the social work curriculum.

Table 2: Descriptive statistics by gender

	Men		Women	
Pooled sample	Mean	sd	Mean	sd
Achievement	0.73	0.28	0.84	0.30
GPA	16.71	2.29	17.44	1.98
SweSAT	12.87	3.50	10.50	3.79
Valid SweSAT	0.82	0.39	0.72	0.45
	N=5,067		N=3,904	
Engineering	Mean	sd	Mean	sd
Achievement	0.70	0.28	0.76	0.25
GPA	16.89	2.14	17.74	1.78
SweSAT	13.02	3.35	11.67	3.50
Valid SweSAT	0.82	0.39	0.81	0.39
	N=3,337		N=1,282	
Business and Economics	Mean	sd	Mean	sd
Achievement	0.77	0.26	0.85	0.22
GPA	16.23	2.43	17.35	1.98
SweSAT	11.99	3.86	9.66	3.69
Valid SweSAT	0.82	0.39	0.69	0.46
	N=1,156		N=1,154	
Social work	Mean	sd	Mean	sd
Achievement	0.78	0.26	0.90	0.18
GPA	14.65	2.36	16.48	2.00
SweSAT	11.13	2.97	8.86	3.49
Valid SweSAT	0.84	0.37	0.65	0.48
	N=179		N=997	
Law	Mean	sd	Mean	sd
Achievement	0.85	0.25	0.89	0.23
GPA	17.47	2.37	18.73	1.47
SweSAT	14.90	2.69	12.14	3.64
Valid SweSAT	0.82	0.39	0.68	0.47
	N=395		N=471	

Note: *Achievement* shows the accomplished credits relative to enrolled and is measured in percentage. *GPA* is a weighted average of upper secondary grades and ranges from 10–20. *SweSAT* is reported on a scale from 0–20. *Valid SweSAT* is a dummy variable taking the value of one if the student has a valid SweSAT.

3 Empirical Strategy

Essentially, we study the achievement differences by estimating the average treatment effect. Formally, the average treatment effect can be described as $ATE = E[Y_{1i} - Y_{0i}]$, where Y_{1i} is the achievement of individual i if treated, and Y_{0i} is the counterfactual outcome in the absence of treatment (Rubin, 1973). This framework is often applied to analyse the effect of a policy intervention, but in this study, we are interested whether men and women with the same admission score perform equally well at university. Therefore gender is here defined as being the 'treatment'. The treatment variable takes the value of one if the student is female, meaning the central measure is then the female-male difference in first-year achievement, which we simply denote the achievement difference.

The problem in analysis based on empirical data is that both outcomes, Y_{1i} and Y_{0i}, cannot be observed because an individual cannot be treated and untreated. To estimate the ATE, one can use matching methods to construct the counterfactual outcomes. Because our baseline model is only conditional on the score from one admission instrument, we use a simple nearest-neighbour matching to estimate differences in female-male achievement.[5]

Moreover, we apply four-to-one matching, meaning that we match the four closest observations in one treatment group to each observation in the other.[6] The number of matches per individual is a trade-off between bias and the sample variability, where more matches per observation, that is, a larger k, increase bias but decrease variability (Rosenbaum, 2020). The gains in variability vanish at k larger than 5. We use matching with replacement, meaning that each 'control' can serve as a potential match for several treated units. Allowing matching with replacement is especially useful when the sample size and control group are small because it ensures that the potential matches are similar in terms of the covariates X (Dehejia & Wahba, 1999; Stuart, 2010).

Finally, a kernel-weighted local polynomial regression is applied to the estimated achievement differences to visualise achievement differences for different parts of the score distributions. This approach allows us to analyse if achievement differences are more or less pronounced in the different parts of the distributions. The bandwidth decides the smoothness of the polynomial regressions, and we used the default bandwidth provided by Stata.[7]

GPA or SweSAT are the only matching variables in our baseline models, and we anal-

[5] Other common methods to produce matches are propensity score matching or coarsened exact matching. The best matching algorithm ultimately depends on the available data and the research question. Propensity score matching matches treated and untreated by their probability of being treated, given a set of covariates to mimic the characteristics of a randomised design (Austin, 2010). However, the matched units from propensity score matching do not necessarily have similar covariate values. Thus, King and Nielsen (2019) suggest using coarsened exact matching instead to ensure that the treated and matched comparisons are similar. These approaches are useful when the matching algorithm include many covariates because simpler models would result in poor or few matches.

[6] We also applied one-to-one matching, reported in Figure B1 and B2 in the Appendix. The number of matches did not alter the estimated achievement differences.

[7] The bandwidth varied around 0.5–1.5 with the default setting in Stata.

yse achievement differences by each instrument separately. Then, a second matching is performed on the selection instruments as well as field of education. This makes it possible to detect if achievement differences depend on the gender composition in different fields, for example, if women tend to register in programmes and courses in which it is easier to obtain a pass grade. Matching on educational fields also provides the opportunity to study if there are differences between the fields.

However, we do not condition the matching on admission group, meaning that students admitted from both instruments are represented in each match. Students could be matched on their GPA scores even if they were admitted in the SweSAT group and vice versa. Students in the lower parts of the distributions were most likely admitted by the other instrument but this should not be a problem if the matched pairs of men and women are admitted by the same admission instrument. A simple tabulation on gender and admission instrument, not reported here, shows no systematic differences at the bottom of the distributions; that is, the matched pairs are often in the same admission group even if we do not explicitly condition the matching on admission instrument.

Finally, in addition to this, we perform additional analyses including background characteristics such as age, parental highest level of education and foreign background because socioeconomic background has been found to explain at least part of the predictive power of admissions instruments (e.g. Rothstein, 2004; Wikström & Wikström, 2017) and academic success (Björklund & Salvanes, 2011, and the references therein).

4 Results

In the baseline model, matching generates 8,603 and 6,930 pairs when matching on GPA and SweSAT respectively.[8] The average achievement differences are presented in Table 3. A positive value of the achievement difference means that women are underpredicted relative to men. The table displays that women are on average underpredicted by 8 percentage points in GPA and 10 percentage points in the SweSAT scores.

Table 3: Average achievement difference in baseline model

	GPA	SweSAT
Achievement difference	0.081***	0.109***
	(0.005)	(0.007)
No. matches	8,603	6,930

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Figure 2 presents local polynomial regressions to illustrate gender differences in predic-

[8]The 368 individuals who had no GPA were dropped from the GPA matching, and the corresponding number for the SweSAT is 2,041.

tion over the score distributions. From that, we see that women perform up to 25 percentage points better than men in the lower part of the GPA distribution (see Figure 2). The achievement difference shrinks when the GPA increases but is nonetheless around 5 percentage points when GPA is 18 or more. The result indicates that women are underpredicted by their GPA score in relation to their academic performance over the entire distribution of GPA. Women are also underpredicted by their SweSAT score, even if the difference is not as large in the lower part of the SweSAT distribution. However, the downward trend is less pronounced compared to the GPA, and women perform around 7 to 18 percentage points better than men.

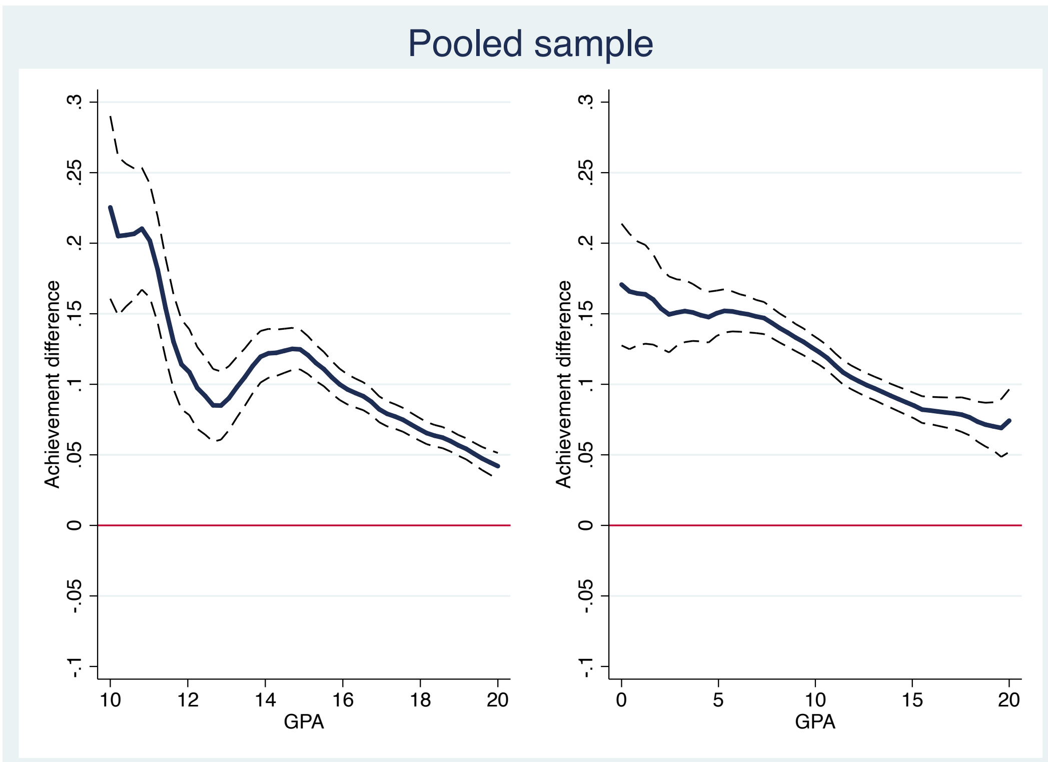

Figure 2: Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching. The areas between the dashed lines show the associated 95% confidence intervals.

Matching on educational fields

Hitherto, we have presented achievement differences in the pooled sample when the matches were only based on GPAs or SweSAT scores. This implies that individuals can potentially be matched across different fields of study. For example, a woman in the business and economics programme can be matched with a male engineer if they have similar GPAs or SweSAT scores. Thus, we also match the sample on field of education.

Table 4 presents the average achievement differences for the sample as a whole and for the educational fields separately. Starting with the sample as a whole, denoted as *'All Fields'*, we find that the average differences are lower when conditioning on educational fields. The main average difference is 2.8 percentage points for GPA and 7.4 percentage points for the SweSAT. A conclusion we can draw based upon a comparison with the results in the previous section, therefore, is that part of the achievement differences can be explained by composition effects. We will return to the field-particular achievement differences when presenting the result from the local polynomial regressions below.

Table 4: Average achievement difference by field of education

	All fields	Engineering	Business and Economics	Social work	Law
GPA	0.028***	0.016*	0.037***	0.069***	0.009
	(0.006)	(0.009)	(0.010)	(0.026)	(0.020)
No. matches	8,603	4,510	2,225	1,048	820
SweSAT	0.074***	0.073***	0.081***	0.102***	0.023
	(0.007)	(0.010)	(0.012)	(0.025)	(0.020)
No. matches	6,930	3,756	1,730	800	644

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

To investigate the impact of potential composition effects further, we perform matching where we exclude each field separately. Table A3 in the Appendix shows the estimated average achievement differences for these matches, and we note that excluding engineering or social work, which have the most uneven gender composition, changes the estimated differences the most. These results strengthen our hypothesis that the relatively large effect reported in the baseline model partially stems from the gender compositions in the included fields. In contrast, the estimated average achievement is only marginally affected by the exclusion of the other fields.

Figure 3 presents the local polynomial regression for the sample as a whole. It also shows that the estimated achievement differences tend to become smaller as we include field of education, again indicating that part of the difference is related to the male-female composition of students in the different fields. The achievement difference between women and men is relatively larger for SweSAT than GPA, indicating that SweSAT tends to under-predict female achievement more than GPA. A difference in comparison with the baseline model is that the downward slope of the regression lines that was found for the baseline model is not at all as pronounced in Figure 3. The achievement difference is approximately constant over the GPA distribution. For the SweSAT, the relation is less stable, but there is no systematic trend.

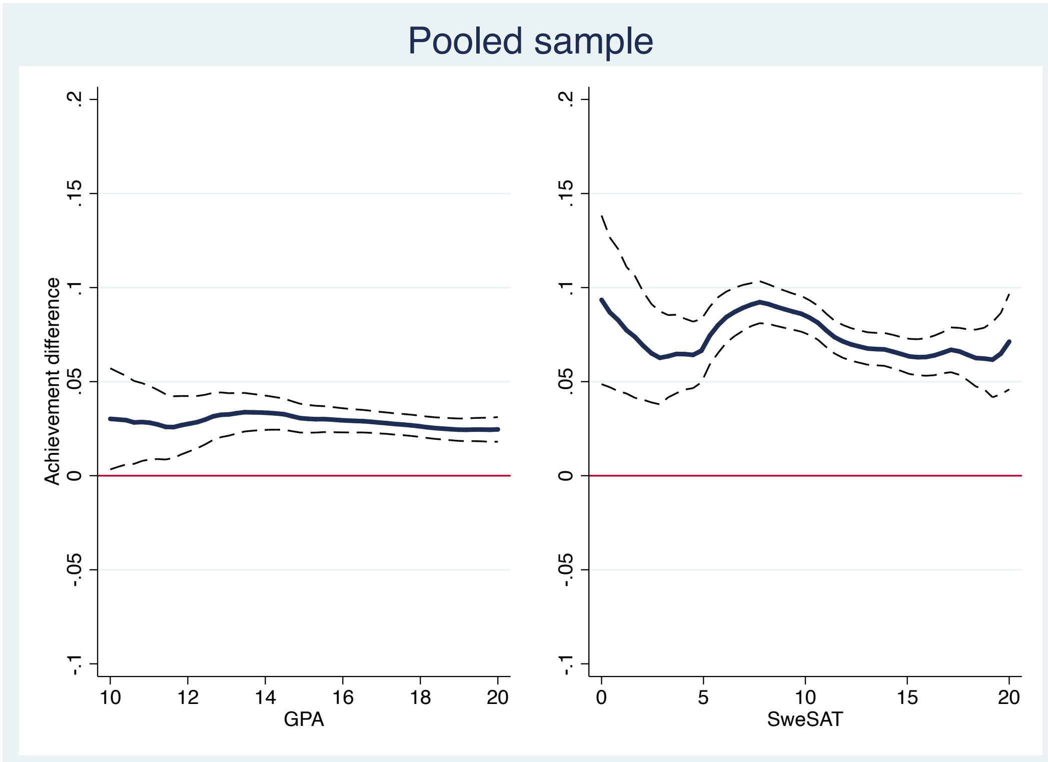

Figure 3: Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching conditional on field of education. The areas between the dashed lines show the associated 95% confidence intervals.

Figures 4 to 7 shows local polynomial regressions for the different educational fields separately. We note differences in gender composition across studied fields of education. Engineering, the largest field, has a large overrepresentation by men. The average achievement difference is 1.6 percentage points for GPA and about 7 percentage points for SweSAT (see Table 4). Inspecting Figure 4 one can see that women with low GPA, between 10–12, have a non significant overprediction of achievement, but the difference becomes smaller for the students with higher grades. The number of students is fewer in the lower part of the GPA and SweSAT distribution, resulting in larger standard errors as the number of matching pairs decreases. In the field of business and economics, more in line with the pooled results, we observe a positive achievement difference on average (see Table 4), and Figure 5 supports the conclusion that the differences are positive for most parts of the score distributions. Notice that the average differences, reported by Table 4, are large for the SweSAT also in this case.

Figure 4: Engineering

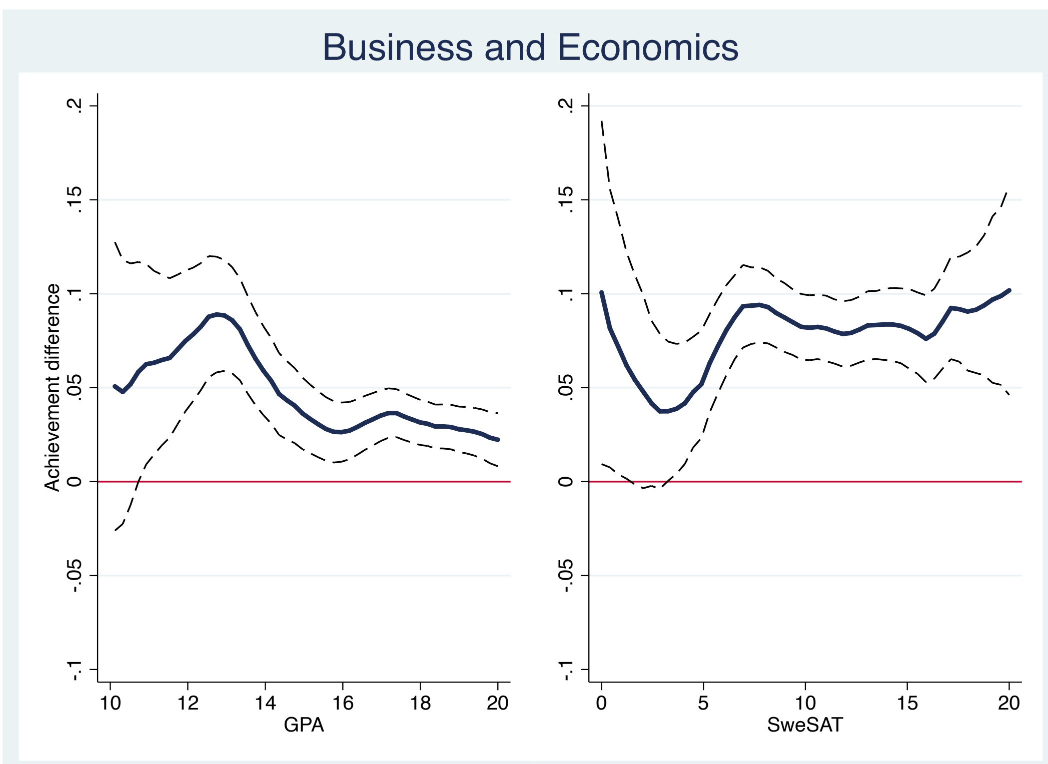

Figure 5: Business and Economics

Note: Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching. The areas between the dashed lines show the associated 95% confidence intervals.

Looking at the fields of social work and law, we note that the results are less stable than the other two fields. The number of matches becomes smaller in these fields, which may explain why the achievement difference tends to fluctuate more over the score distributions. Focusing on social work and estimations based on GPA, the average difference is 6.9 percentage points but varies largely across the score distribution and is negative for parts of the distribution. The achievement difference is largest for the highest scores in SweSAT. Part of this is most likely explained by the overrepresentation of women in the social work programmes, where only 15% of the students are men. Finding good matches in the lower and upper parts of the distributions is probably difficult. In addition, the SweSAT is overall lower in social work than other programmes, resulting in very few matches in the upper end of the SweSAT score distribution. The distance between the observations were nevertheless rarely larger than one, but for example, at SweSAT scores lower than three, women were matched with men that had SweSAT scores between three to five.

Law is one of the most popular fields, and the admissions scores in both GPA and SweSAT are consequently high. We note a large overprediction of female achievement for GPA lower than 16. Only 14.4% of the law students had grades below 16, and men are heavily overrepresented (72%) in this part of the GPA distribution. At grades lower than 14, the gender imbalance increase further as there were only 9 women and 50 men. Thus, the overprediction of female achievement is probably related to the lack of observations for lower grades. Nevertheless, the average achievement differences in law are close to zero and statistically insignificant, implying that we find a tendency to underprediction in the field of law.

As previously mentioned, we use the default bandwidth provided by Stata in our polynomial regressions. The bandwidth size is a trade-off between bias and variability, whereas larger bandwidth comes at the cost of larger bias and smaller bandwidths yield more precise smoothed values with high variability. Because the estimated achievement differences diverged from the main result in the social work and law programmes, we elaborated with different bandwidths to ensure that our result is not driven by choice of bandwidth. However, the additional analysis did not alter the overall results. There was, for example, still a large overprediction of female achievement for low GPA in law programmes even when we increased the bandwidth from 0.54 to 1.

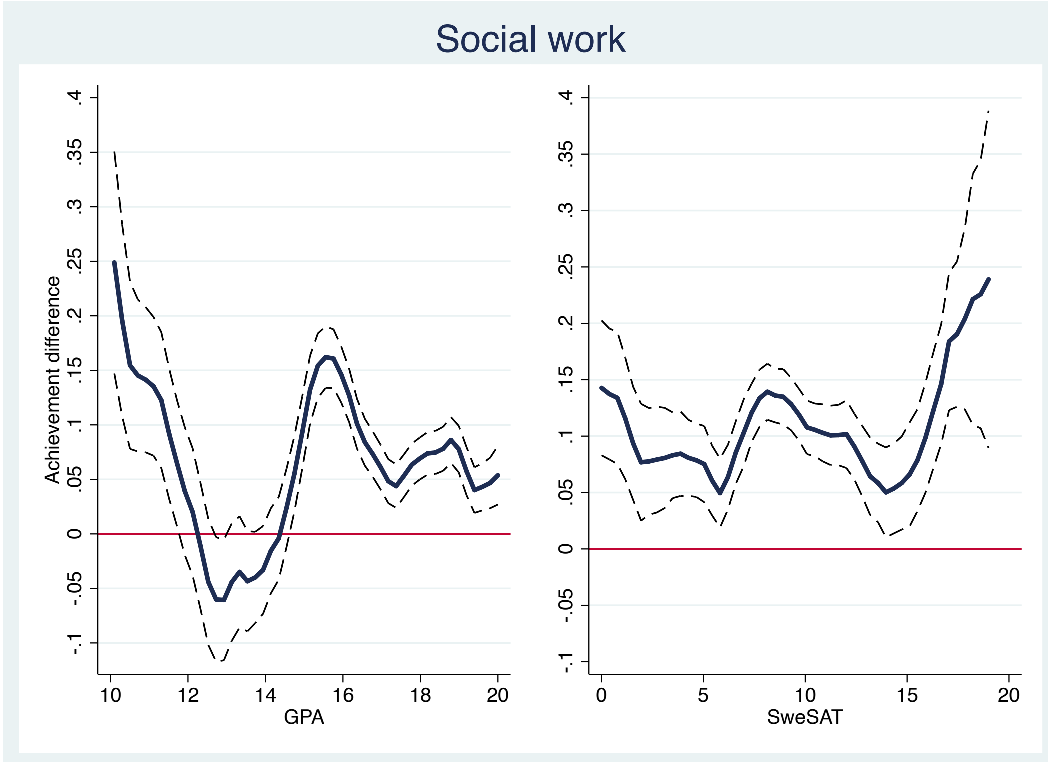

Figure 6: Social work

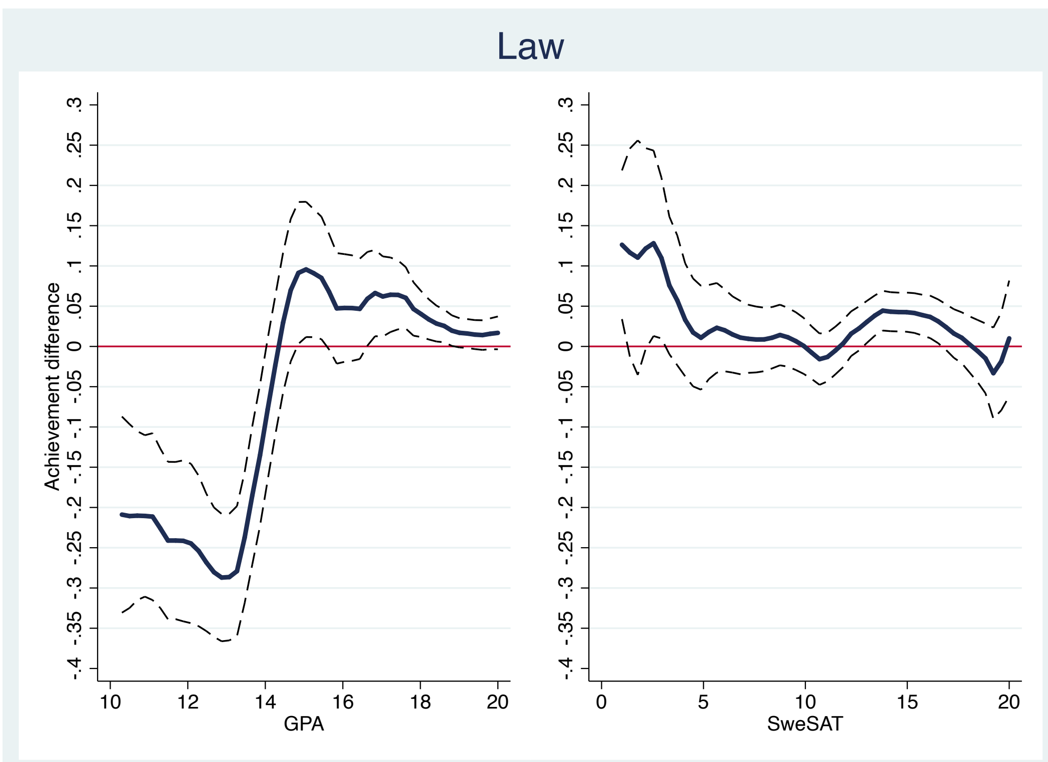

Figure 7: Law

Note: Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching. The areas between the dashed lines show the associated 95% confidence intervals.

Finally, we also report on estimates where matching has been performed on background characteristics as well as the selection instruments and the fields of education. Although we do not expect that background characteristics should vary much between men and women, conditioning on them functions as a control that this is not a problem that causes misinterpretation. We choose to match on age, parental highest level of education and foreign background (see Table A1 in the appendix for variable definitions). Doing so reduces the number of potential matches, so we only report the results for the sample as a whole. The polynomial regression line is displayed in Figure 8. Compared to Figure 3, the estimates from the GPA matches become less stable and have larger standard errors, especially for lower grades. The estimates on SweSAT scores are more or less unaffected by the included background controls. The average achievement difference for women relative to men is 3.3% (SE=0.007) for GPA and 7.7% (SE=0.007) for SweSAT, indicating that the inclusion of background controls does not substantially change the previously estimated relationships.

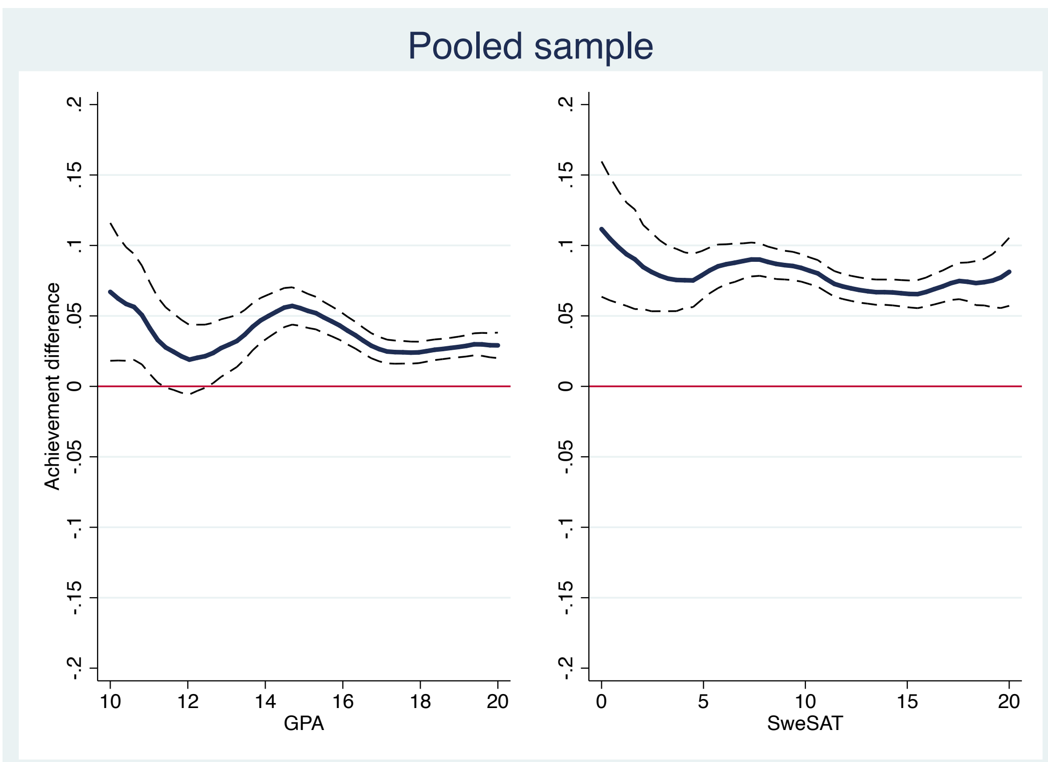

Figure 8: Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching on pooled sample conditional on age, parental highest education, foreign background and field of education. The areas between the dashed lines show the associated 95% confidence intervals.

5 Conclusion

This study analyses gender differences in admissions scores and achievement using detailed data from 8,971 first-time students in four different fields of specialisation in Sweden. Using nearest-neighbour matching conditional on upper secondary school GPA or SweSAT scores, we find notable achievement differences for women relative to men. Both instruments underpredict female achievement, but the estimated achievement differences are more distinct for SweSAT compared to the GPA.

The baseline model, where matching is based on GPA or SweSAT, shows that female students achieve on average 8 and 10 percentage points more credits than men conditional on GPA and SweSAT, respectively. Applying local polynomial regressions that allow us to observe achievement differences for both low and high scores shows that the differential prediction is persistent throughout the score distributions. However, when we include field of education in the matching algorithm, the average achievement difference decreases; GPA underpredicts female achievement relative to men by approximately 3 percentage points and SweSAT by 7.5 percentage points on average. The differential predictions are reasonably stable across the score distributions. This means that the results from the baseline model are, to some extent, explained by gender-composition effects. As a third step, we estimate the achievement differences for each field separately. For engineering and business and economics, which contain a relatively large number of students, the estimated local polynomials are stable over the distribution. At the five per cent level, there is no statistically significant female underprediction in the engineering field with respect to GPA, while we observe underprediction for the SweSAT. In business and economics, underprediction persists for both admission instruments. The method used works less well in the fields where the number of students is small. Law is a competitive field; the number of students admitted with a low GPA is small, and it is difficult to find good matches. The tentative conclusion we draw is that there is no evidence of differential prediction in law, while women are, on average, substantially underpredicted in social work.

The results of this study are in accordance with the previous literature that also found female underprediction in tests as well as GPAs (Fischer et al., 2013). Keiser et al. (2016) investigated course-taking patterns and found that this can partly explain the differential prediction, which can explain why the achievement differences tend to become smaller once we include educational field in the matching algorithm. One of the most consistent results is that differential prediction is substantially larger for the SweSAT than for the upper secondary GPA. Why is it that we observe this pattern? One explanation that has been pointed at previously is that there is an omitted variable problem in the sense that the instruments capture different abilities to a different degree (Duckworth & Seligman, 2006; Keiser et al., 2016; Mattern et al., 2017). However, it is also known that males perform better on the SweSAT and on tests in general (Wikström & Wikström, 2017), so another form of omitted variable problem would be gender differences that have to do with test taking. Previous research has found that test anxiety tends to be more severe among women than among men, especially among low achievers (e.g. Cassady & Johnson, 2002; Stenlund et al., 2017).

Also, descriptive results show that women complete more credits than men, which might reflect that women are more motivated and ambitious than men, and this may contribute to explain our results. Therefore, the cause of the observed differences between the SweSAT and GPA is still an open question. Note, finally, that our data only contain first-year achievement. To the extent that men tend to become more motivated over the course of studies, they may catch up as university studies continue. Thus, an interesting continuation on this topic would be to extend the studied period by year two or three to see if the differential prediction persists.

References

Austin, P. C. (2010). Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. *American journal of epidemiology*, 172(9), 1092–1097.

Bhattacharya, D., Kanaya, S. & Stevens, M. (2017). Are university admissions academically fair? *Review of Economics and Statistics*, 99(3), 449–464.

Bhattacharya, D. & Rabovic, R. (2020). Do Elite Universities Practise Meritocratic Admissions? Evidence from Cambridge. *Cambridge Working Papers in Economics: 2056 (June 22, 2020)*.

Björklund, A. & Salvanes, K. G. (2011). "Education and family background: Mechanisms and policies". *Handbook of the Economics of Education*. Vol. 3. Elsevier, pp. 201–247.

Brenøe, A. A. & Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4), 1009–1054.

Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology*, 27(2), 270–295.

Conger, D. & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science*, 627(1), 184–214.

Dehejia, R. H. & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448), 1053–1062.

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

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., et al. (2016). School quality and the gender gap in educational achievement. *American Economic Review*, 106(5), 289–95.

Fischer, F. T., Schult, J. & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology*, 105(2), 478.

Hanson, A. (2017). Do college admissions counselors discriminate? Evidence from a correspondence-based field experiment. *Economics of Education Review*, 60, 86–96.

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, 21(6), 589–598.

Keiser, H. N., Sackett, P. R., Kuncel, N. R. & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, 101(4), 569.

King, G. & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454.

Mattern, K., Sanchez, E. & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice*, 36(1), 47–57.

Mouganie, P. & Wang, Y. (2020). High-performing peers and female stem choices in school. *Journal of Labor Economics*, 38(3), 805–841.

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, 7, 143–176.

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

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.

Stenlund, T., Eklöf, H. & Lyrén, P.-E. (2017). Group differences in test-taking behaviour: An example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice*, 24(1), 4–20.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1), 1.

Törnkvist, B. & Henriksson, W. (2004). Repeated test taking: Differences between social groups. Institutionen för beteendevetenskapliga mätningar, Umeå.

Wikström, C. & Wikström, M. (2017). "Group differences in Student performance in the Selection to Higher Education: tests vs Grades". *Frontiers in Education*. Vol. 2. Frontiers, p. 45.

A Appendix

Table A1: Tabular of institution and field of study

Institution	Engineering	Business and Economics	Social work	Law	Pooled sample
Umeå University	204	201	146	74	625
Luleå University of Technology	411	56	0	0	467
Uppsala University	423	120	52	210	805
University of Gävle	0	113	58	0	171
Dalarna University	0	32	34	0	66
Mälardalen University	120	116	70	0	306
Örebro University	0	62	65	103	230
Stockholm University	0	0	126	172	298
KTH Royal Institute of Technology	823	0	0	0	823
Linköping University	761	221	81	0	1,063
Jönköping University	0	0	44	0	44
University of Gothenburg	0	236	88	176	500
Chalmers University of Technology	719	0	0	0	719
Karlstad University	114	119	35	0	268
University of Skövde	0	40	0	0	40
University of Borås	0	84	0	0	84
Lund University	871	269	132	131	1,403
Halmstad University	0	64	0	0	64
Stockholm School of Economics	0	154	0	0	154
Blekinge Institute of Technology	134	0	0	0	134
University West	0	73	0	0	73
Mid Sweden University	39	97	45	0	181
Gotland University	0	39	0	0	39
Malmö University	0	0	97	0	97
Ersta Sköndal Bräcke University College	0	0	18	0	18
Linnaeus University	0	214	85	0	299
N	4,619	2,310	1,176	866	8,971

Table A2: List of variables

Variable	Definition	Source
Gender (woman)	A dummy variable that takes the value 1 if the student is female.	The national education records, 2012
GPA	Student's upper secondary GPA.	The national education records, 2012
SweSAT	Student's SweSAT score.	The national education records, 2012
Completed credits	Student's completed credits in the 2012/2013 academic year. Students with zero completed credits are excluded.	The national education records, 2012
Enrolled credits	Student's enrolled credits in the 2012/2013 academic year.	The national education records, 2012
Achievement	Achievement is defined as the fraction of completed credits divided by enrolled credits.	
Age	Student's age in 2012.	The national education records, 2012
Foreign background	A dummy variable taking the value one if foreign-born or if both parents are foreign-born.	The national education records, 2012
Parental highest level of education	Three dummy variables. One dummy takes the value of 1 if the parent with highest education has 9 years of schooling; the other takes the value of 1 if the education level is at the upper-secondary level, and the last is taking the value of 1 if the parent with the highest education level has studied at the post-secondary level.	The national education records, 2012

Table A3: Average achievement differences when excluding each field separately

	Baseline model	Engineering excluded	Business and Economics excluded	Social work excluded	Law excluded
GPA	0.081***	0.057***	0.089***	0.048***	0.088***
	(0.005)	(0.008)	(0.007)	(0.006)	(0.006)
No. matches	8,603	4,093	6,378	7,555	7,783
SweSAT	0.109***	0.091***	0.117***	0.093***	0.109***
	(0.007)	(0.010)	(0.008)	(0.007)	(0.007)
No. matches	6,930	3,174	5,200	6,130	6,286

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

B Appendix

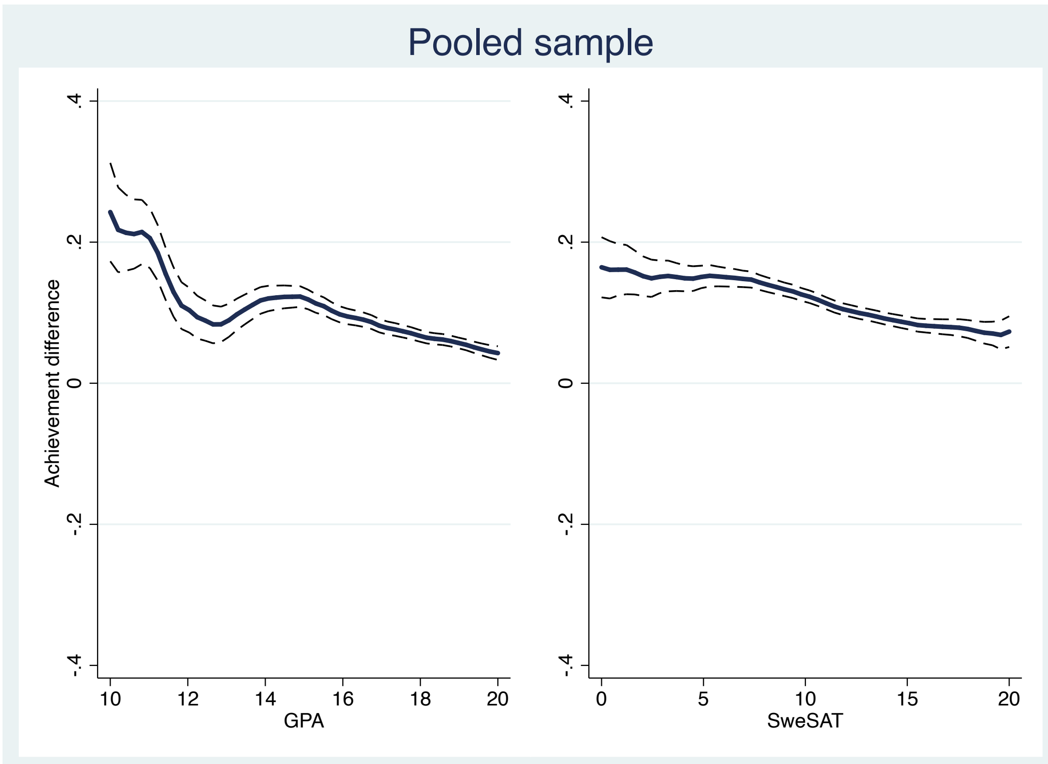

Figure B1: Poole sample using 1:1 matching. Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching. The areas between the dashed lines show the associated 95% confidence intervals. Note that this matching is not conditional on field of study.

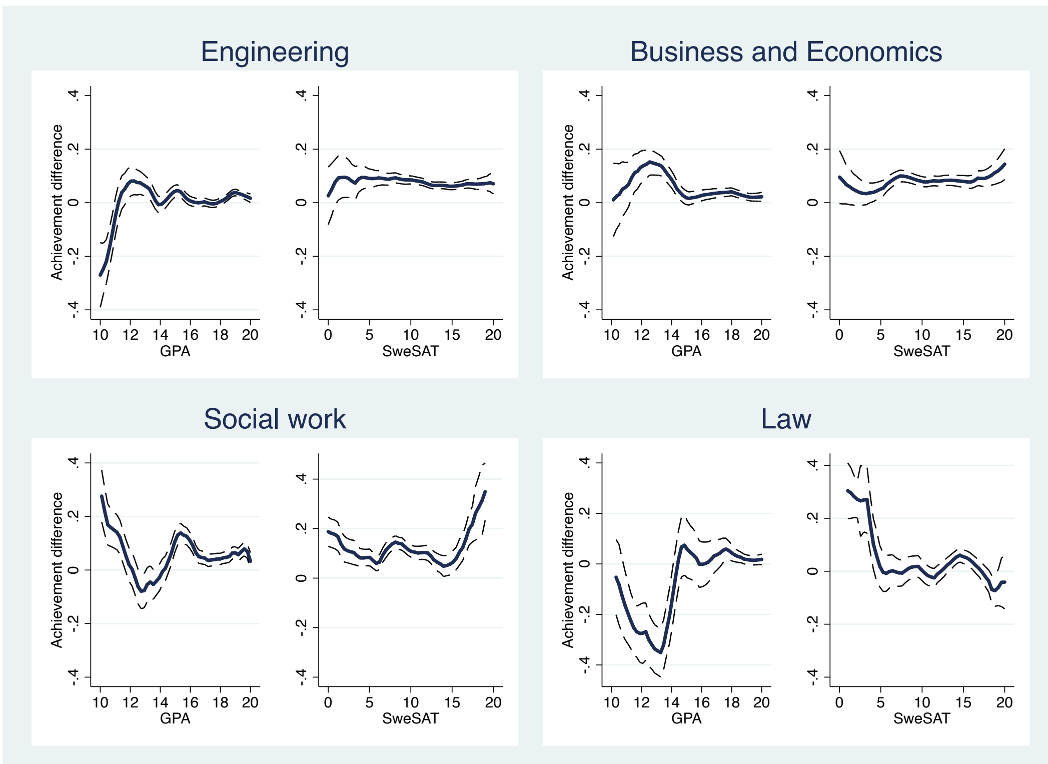

Figure B2: Field-specific matching using 1:1 matching. Estimated achievement differences, in percentages, for women relative to men from nearest-neighbour matching. The areas between the dashed lines show the associated 95% confidence intervals.