Gender gaps in high-stakes teacher assessments

Konstantina Maragkou

University of Cambridge

Conference on Research on Economic Theory and Econometrics

10-14 July 2023, Naxos

Motivation I

- In England, as in most OECD countries, girls show superior performance in school examinations compared to boys (Machin and Pekkarinen, 2008)
- This is also reflected in higher rates of university participation
- Despite this, girls are far less likely to engage in STEM (Cavaglia et al., 2021)
- Because STEM is associated with higher earnings, differential participation has important implications for the gender pay gap (e.g. Blau and Khan, 2017; Francesconi and Parey, 2018)
- One driver of the gender gap in STEM could be exposure to stereotypical teacher assessments (e.g. Lavy and Megalokonomou, 2019; Carlana, 2019; Terrier 2020)

Motivation II

- In this paper, we look at gender gaps in teacher assessments in STEM and non-STEM subjects in a context where these are high-stakes:
  - Students apply to university using teacher predicted grades
  - It is, therefore, predicted rather than awarded grades than inform how ambitiously students apply and whether they receive offers in response
- We know little about how and why the "prediction accuracy" might vary across students and schools
- Does it matter?
  - Probably: If some groups are predicted more or less optimistically than others, this can distort the applications, offers, and acceptances with important implications for widening participation and social mobility
  - Evidence that more optimistic predictions are associated with applying to more selective university courses (Murphy and Wyness, 2020)

This paper

- Studies gender differences in the probability of under- and over- prediction
- Focuses on ten facilitating and popular A-level subjects (STEM and non-STEM)
- Uses a newly linked administrative dataset
- Finds that girls are predicted more optimistically in both STEM and non-STEM subjects
- But accounting for gender differences in past performance reveals gender gaps in favour of **boys in STEM** and in favour of **girls in non-STEM** $=>$ The intuition is that boys have a higher than expected growth in achievement and therefore receive less optimistic predictions

Contributions

- Add to literature on the importance of external assessments and explaining gender gaps in STEM
- New administrative dataset that allows access to detailed information on students' grades and the university application and admission process at an individual basis
- First empirical evidence on gender gaps in teacher assessments in a context where these are high-stakes
- Discuss and account for a wide range of alternative explanations for the patterns in differential under- and over- prediction

Institutional Background

- University admissions in England are controlled by a centralised body (UCAS)
- Students apply for up to five courses in each application round
- The application process is complex => students required to decide on which courses to apply to almost a year in advance of entry
- Importantly, they apply on the basis of teacher predictions rather than their actual achieved grades
- Universities offer places based on assessment of the predicted grades and a supplementary personal statement
- Students then choose first and back-up course from the available options
- Only after this process is completed the students go on to take actual exams

Predicted Grades

- Limited guidance provided to teachers on how to predict students
- Schools free to apply their own general regulations
- UCAS states that: 'a predicted grade is the grade of qualification a student's school or college believes they're likely to achieve in positive circumstances'
- Predictions should be 'aspirational' => motivate students to apply optimistically and achieve higher grades
- But also, 'achievable' and 'data-driven'
- Combination of three components:
  - GCSE exams => external high-stakes exams taken 14-17 months before predictions in about 10 subjects
  - Any internal assessments (e.g. coursework, tests, etc)
  - Teachers' personal judgement

Dataset

- GRading and Admissions Data for England (GRADE)
- Joint open data initiative initiated between UCAS, Ofqual, and the Department for Education
- First student-level linked data made available on A-level results and university applications
- Cohorts: 2016/17 to 2019/20
- Information on socio-demographics, school characteristics, and prior attainment including age 16 high-stakes exams (GCSEs)
- No information on international or mature students
- Also, no information on non-A-level qualifications

Analytical sample

- University applicants via the A-level route
- About 80% of university applicants study A-levels in three subjects
- Facilitating and "popular" subjects
  - STEM: Biology, Chemistry, Economics, Maths, Physics
  - Non-STEM: English Literature, Geography, History, Psychology, Sociology
- 90% of A-level students study for at least one of these subjects
- UCAS application rounds: 2017-2019
- 56% girls and 44% boys

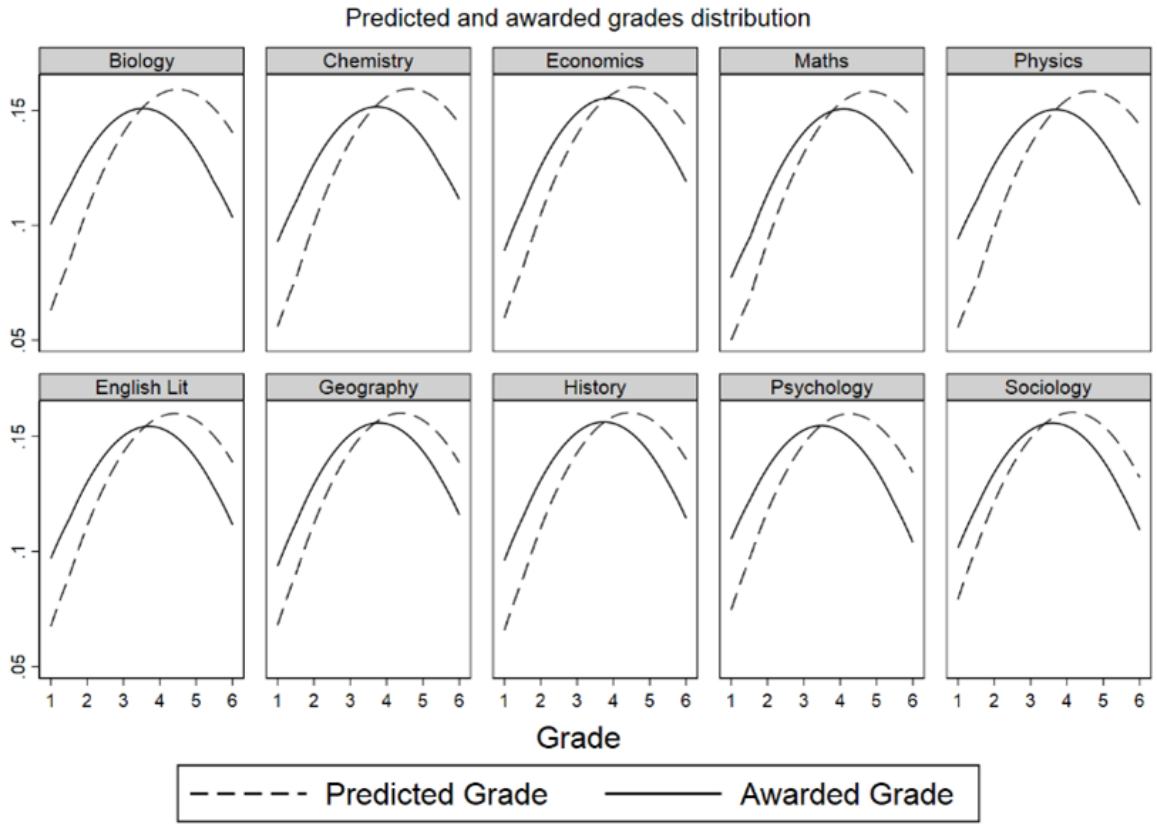
Descriptive statistics I

Table 1: Student prior attainment, socio-demographics and subject-taking

	(1) Biol	(2) Chem	(3) Econ	(4) Math	(5) Phys	(6) Eng Lit	(7) Geog	(8) Hist	(9) Psych	(10) Soc
Girls										
Mean GCSE standardised	0.52	0.74	0.37	0.76	0.89	-0.04	0.18	0.12	-0.19	-0.68
Managerial & prof occ	0.29	0.30	0.32	0.33	0.36	0.26	0.31	0.28	0.23	0.17
Routine occ	0.05	0.05	0.04	0.04	0.04	0.05	0.04	0.05	0.06	0.08
White	0.67	0.59	0.58	0.66	0.72	0.75	0.85	0.78	0.73	0.67
Applied to selective uni	0.84	0.90	0.83	0.88	0.92	0.72	0.75	0.78	0.65	0.56
Offer from selective uni	0.73	0.80	0.77	0.82	0.87	0.65	0.69	0.72	0.55	0.45
N	79068	56555	17697	58140	15253	66339	36366	57484	93517	52405
% of girls	64	53	31	40	22	78	54	58	77	80
Boys										
Mean GCSE standardised	0.24	0.46	0.06	0.39	0.40	-0.15	-0.14	-0.14	-0.47	-0.93
Managerial & prof occ	0.30	0.32	0.33	0.33	0.34	0.29	0.33	0.31	0.24	0.20
Routine occ	0.04	0.05	0.04	0.04	0.04	0.04	0.03	0.04	0.06	0.07
White	0.65	0.61	0.65	0.69	0.74	0.78	0.85	0.82	0.72	0.67
Applied to selective uni	0.83	0.90	0.81	0.88	0.88	0.74	0.71	0.75	0.63	0.53
Offer from selective uni	0.71	0.81	0.73	0.80	0.82	0.66	0.64	0.68	0.51	0.42
N	44643	49899	39250	87831	55122	18831	31399	41991	27569	13450
% of boys	36	47	69	60	78	22	46	42	23	20

Descriptive statistics II

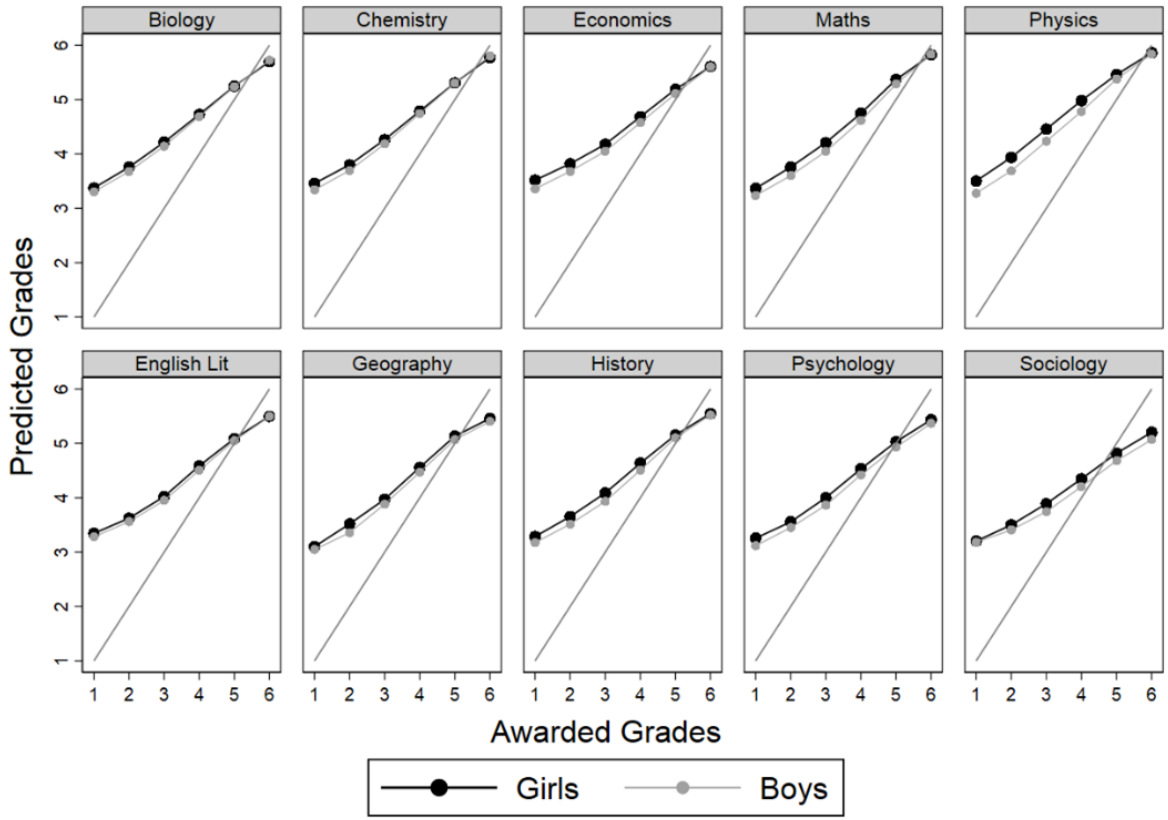
Figure 1: Distribution of predicted and awarded grades



Notes: Predicted and awarded grades are scored between A* to E and transformed to the following scores per grade: A*=6, A=5, B=4, C=3, D=2, E=1.

Descriptive statistics III

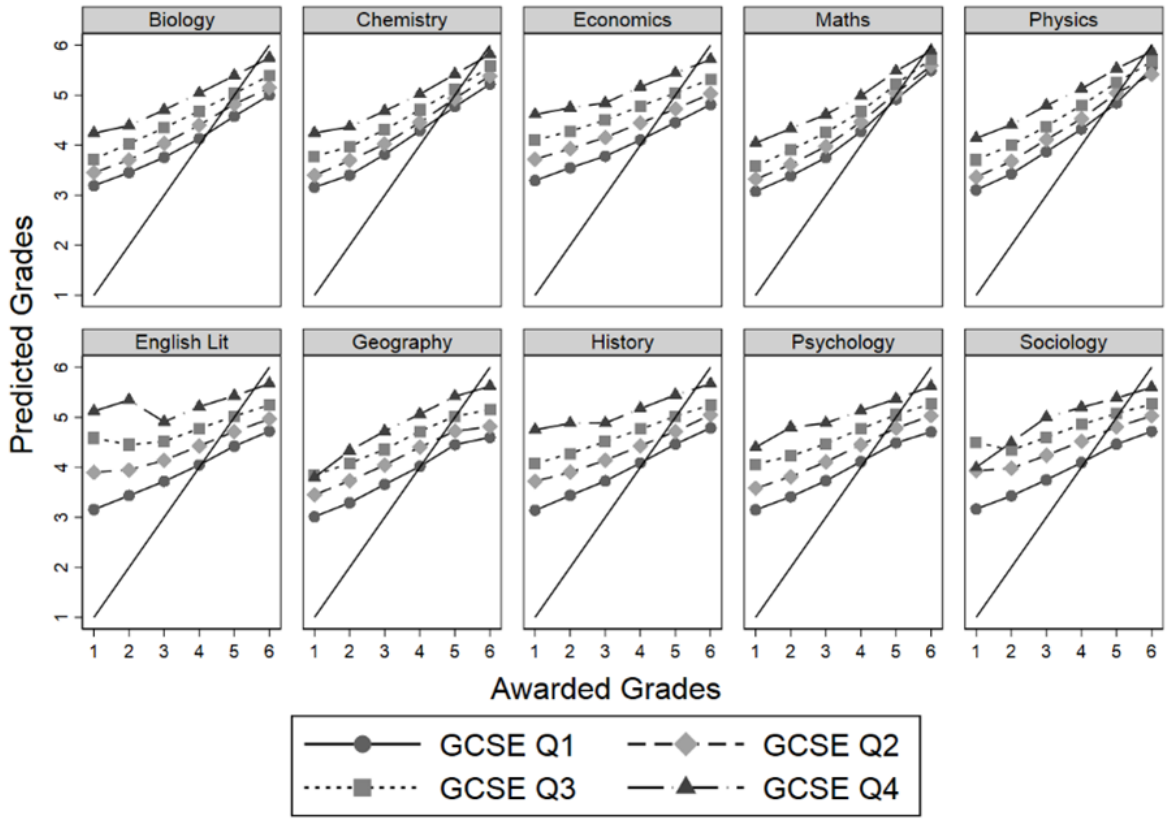
Figure 2: Prediction accuracy



Notes: The 45° line represents perfect accuracy in prediction. Predicted and awarded grades are scored between A* to E and transformed to the following scores per grade: A*=6, A=5, B=4, C=3, D=2, E=1.

Descriptive statistics IV

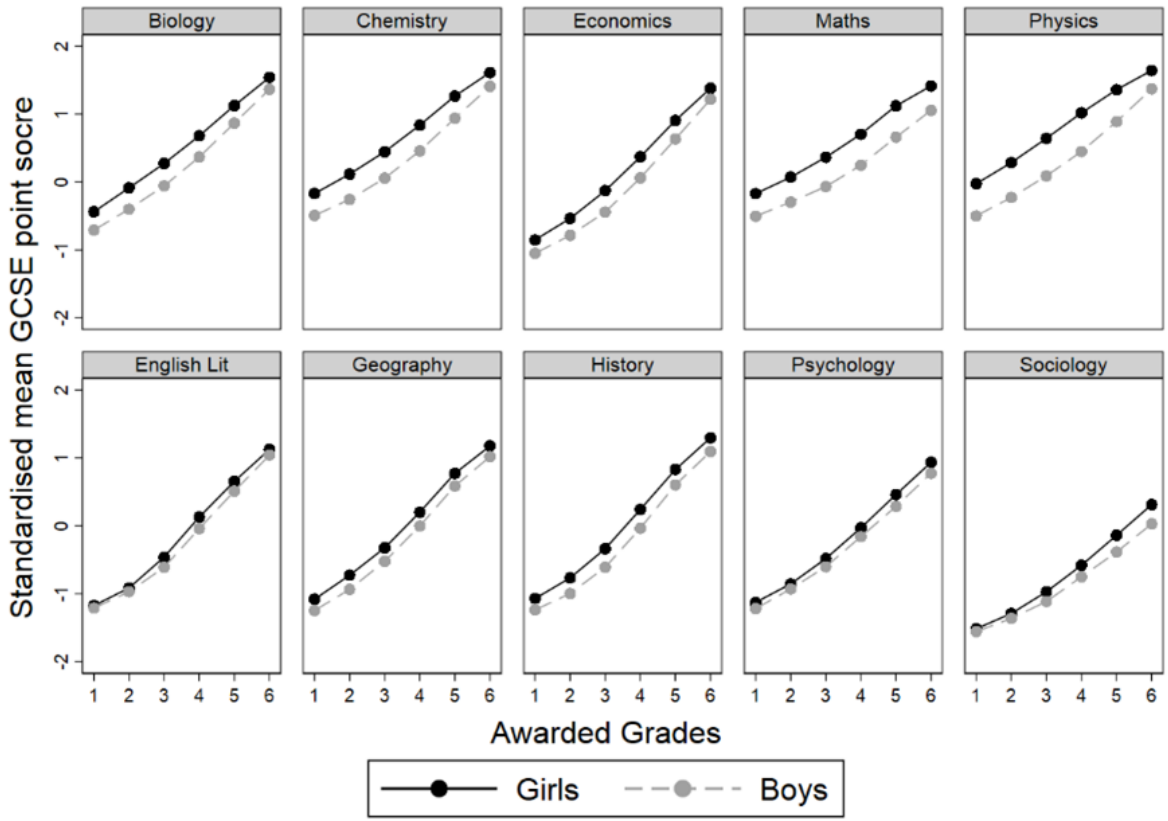
Figure 3: Prediction accuracy and prior attainment



Notes: The 45° line represents perfect accuracy in prediction. Predicted and awarded grades are scored between A* to E and transformed to the following scores per grade: A*=6, A=5, B=4, C=3, D=2, E=1. Prior attainment is the mean GCSE point score (total point score divided by number of subjects taken), standardised for university applicants and divided in quartiles of attainment.

Descriptive statistics V

Figure 4: A-level awarded grade and prior attainment by gender



Notes: The 45° line represents perfect accuracy in prediction. Predicted and awarded grades are scored between A* to E and transformed to the following scores per grade: A*=6, A=5, B=4, C=3, D=2, E=1. The mean GCSE point score is the total point score divided by the number of subjects taken and standardised for university applicants.

Empirical specification

LPM (similar to Burgess and Greaves, 2013 and Shi and Zhu, 2023) for the probability of under- or over- prediction in subject j:

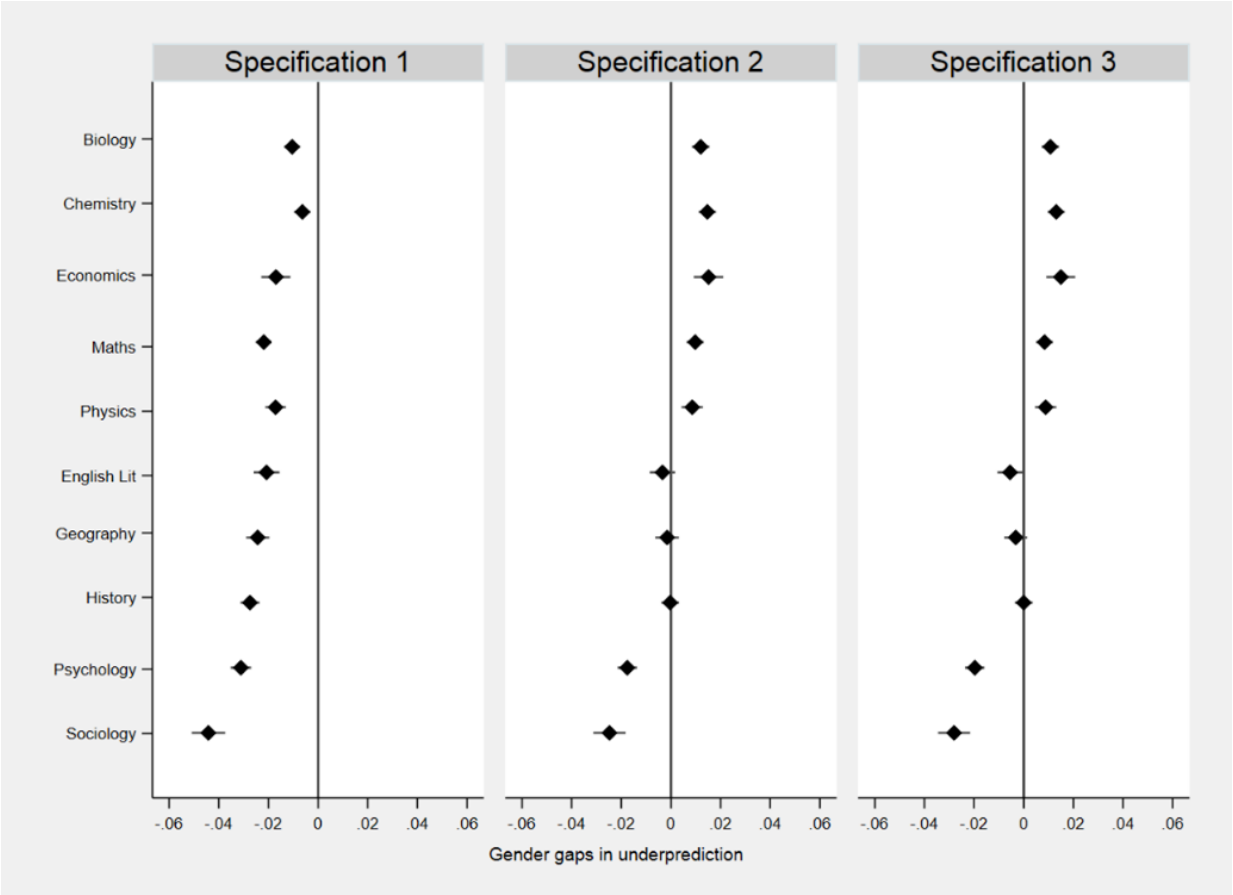
$$P_{i,s} = \gamma_g + \eta_s + \tau_t + \beta G_{i,s} + \delta \overline{A}_{i,s} + \lambda' X'_{i,s} + \epsilon_{i.s}$$

where:

- $P_{i,s}$ is the probability of under- (AG>PG) or over- (AG<PG) prediction
- γ_g is the awarded grade fixed effect
- η_s is the school fixed effect
- τ_t is the application round fixed effect
- $G_{i,s}$ is gender (=1 if girl)
- $\overline{A}_{i,s}$ is prior attainment (standardised mean GCSE point score)
- $'X'_{i,s}$ is a vector of individual socio-demographic, application and offer characteristics
- $\epsilon_{i.s}$ is the standard error clustered at school level

Results I

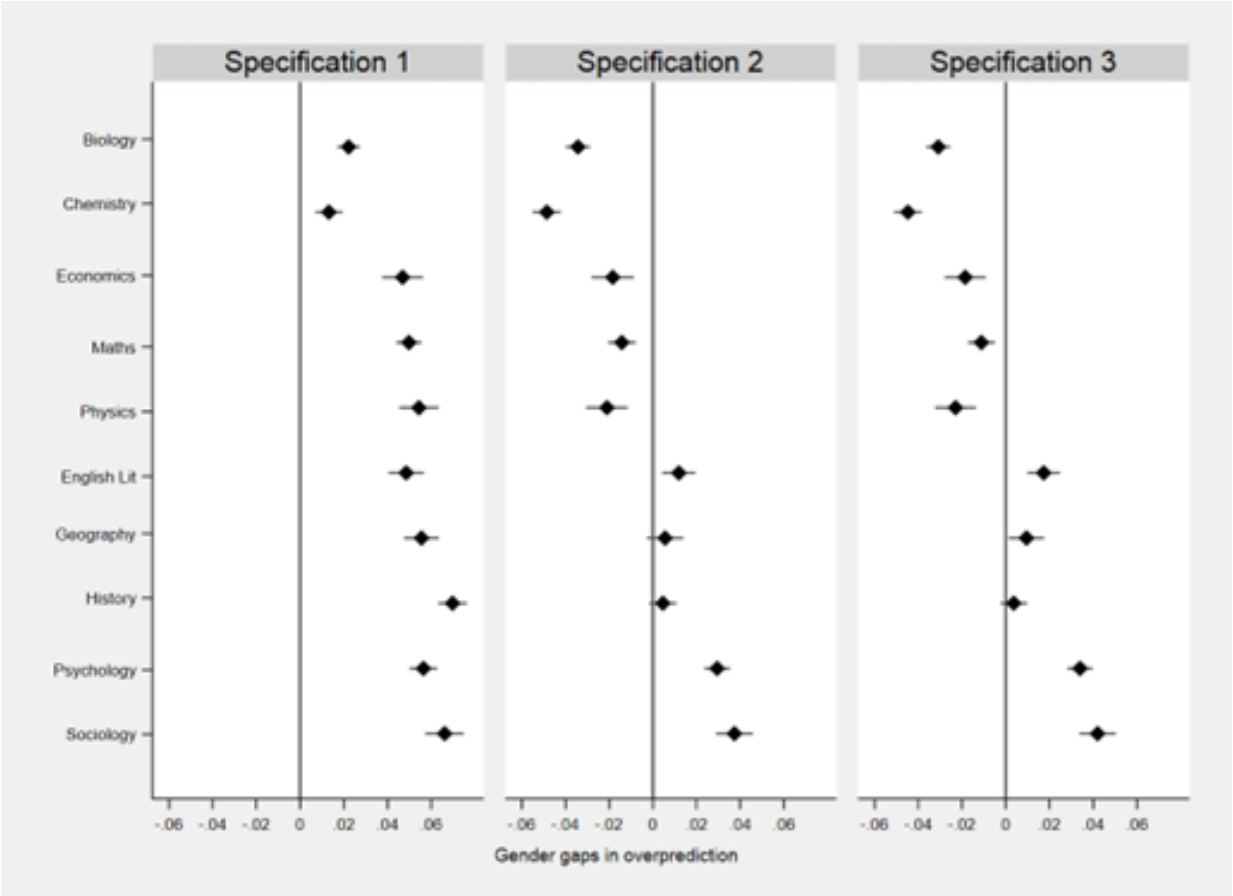
Figure 5: Gender gaps in the probability of under- prediction



Notes: The dependent variable is the probability of under-prediction. The standard errors are clustered by school level. We present 95% CI. Sample excludes students if AG=E.
Specification 1: gender + application round, school and awarded grade fixed effects
Specification 2: + mean standardised GCSE point score
Specification 3: + socio-demographics, application and offer characteristics

Results II

Figure 6: Gender gaps in the probability of over- prediction



Notes: The dependent variable is the probability of over-prediction. The standard errors are clustered by school level. We present 95% CI. Sample excludes students if AG=A*.
Specification 1: gender + application round, school and awarded grade fixed effects
Specification 2: + mean standardised GCSE point score
Specification 3: + socio-demographics, application and offer characteristics

Results III

Table 2: Effect size

	(1) Biol	(2) Chem	(3) Econ	(4) Math	(5) Phys	(6) Eng Lit	(7) Geog	(8) Hist	(9) Psych	(10) Soc
Probability of under-prediction										
Girl	0.011***	0.013***	0.015***	0.008***	0.009***	-0.006**	-0.003	-0.000	-0.020***	-0.028***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Mean under-prediction	0.063	0.052	0.099	0.077	0.047	0.102	0.095	0.083	0.094	0.147
%	17	25	15	10	19	6	3	0	21	19
Observations	113150	98926	55119	138176	64794	82216	65800	96617	113945	63283
Clusters	2603	2580	1701	2690	2484	2489	2376	2525	2446	1824
Probability of over-prediction										
Girl	-0.031***	-0.045***	-0.019***	-0.011***	-0.023***	0.017***	0.009**	0.004	0.034***	0.042***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Mean over-prediction	0.712	0.699	0.608	0.646	0.737	0.639	0.585	0.623	0.636	0.550
%	4	6	3	2	3	3	2	1	5	8
Observations	113045	96816	52367	119967	62838	77143	63410	93490	114702	62047
Clusters	2629	2603	1716	2702	2512	2494	2385	2533	2453	1825
Awarded Grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
School FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Application round FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Prior attainment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Socio-demographics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Application	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Offer	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The standard errors are clustered by school level. Sample in under-prediction model excludes students if AG=E. Sample in over-prediction model excludes students if AG=A*.

Concluding remarks I

- Predicted grades are optimistic
- However, they are deferentially optimistic across student characteristics
- Particularly penalising boys who go on to achieve better than girls given prior attainment => Consistent with the stereotype that boys work less steadily than girls over their A-level studies but cram effectively just before their exams (Times, 2017).
- Adjusting for gender differences in prior performance, predicted grades favouring boys in STEM and girls in non-STEM => consistent with the explanation of unconscious gender bias by teachers

Concluding remarks II

- We account for other possible explanations (incl. ceiling and floor effects, differential ambitiousness and motivation)
- Robustness checks for unobserved student behaviour and subject-specific prior attainment Individual FE
- The differences in under- and over- prediction across gender are sufficiently large to suggest we shouldn't rely on teacher assessments
- Particularly true in a context where these could distort university applications, offers, acceptances at prestigious courses, and exaggerating existing gender gaps in STEM
- The policy implication is moving to a post-qualification application system

Thank you

km844@cam.ac.uk
@K_Maragkou

Table 3: Individual fixed effects

	(1) Pr(AG>PG)	(2) Pr(AG>PG)	(3) Pr(AG<PG)	(4) Pr(AG<PG)
STEM	0.003	0.004	-0.038***	-0.039***
	(0.005)	(0.005)	(0.007)	(0.007)
STEM*Girl	0.021***	0.019***	-0.051***	-0.047***
	(0.003)	(0.003)	(0.005)	(0.005)
	169,202	162,139	166,328	159,991
	2558	2517	2560	2519
Subject FE	✓	✓	✓	✓
A level Awarded Grade FE	✓	✓	✓	✓
GCSE Grade FE		✓		✓

back