A longitudinal study of the gender gap in school grades via flexible Bayesian Beta regression

Laura Bondi[1], Beatrice Franzolini[2], and Marco Palma[3]

[1] Health Data Science Centre, Human Technopole, Milan, IT
[2] Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, IT
[3] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
marco.palma@mrc-bsu.cam.ac.uk

Abstract. The gender disparity in school grades, known as the grade gender gap, has become a focal point for researchers, educators, and policymakers. Despite this heightened interest, there remains a scarcity of longitudinal studies addressing this issue. This study seeks to address this gap by examining the grade gender gap in schools using longitudinal data from Rimini, Italy. A Bayesian Beta regression with splines is devised to detect gender differences across different cohorts of students and along the educational journey from primary to upper-secondary school. Our findings reveal a significant gap in favour of female students, which widens during secondary school years but diminishes in younger cohorts. These pages contain preliminary results of the extended study in [2].

Keywords: Bayesian distributional regression, gender differences, panel data, regression splines

1 Gender gap in school grades

The *grade gender gap* is the difference in school performance between male and female students across various subjects and educational levels and is often used as a key metric by researchers, educators, and policymakers in recent years [6, 8]. Quantification and investigation of this disparity may underscore potential imbalances within educational systems and their implications for individuals and society. While gender differences in school performance are a flourishing area of research, many studies are based on cross-sectional data and therefore do not capture changes over time. The aim of this study is to quantify and analyse the grade gender gap within primary and secondary schools over multiple years, analysing longitudinal data from the province of Rimini, Italy.

Recent studies conducted in Western nations indicate that, on average, girls tend to outperform boys academically, with narrower advantages in STEM subjects [see, for instance, 11, 10]. However, gender disparities in favour of female students in STEM are not universally consistent. For example, findings from the 2022 OECD PISA report [9] reveal slight yet significant differences favouring boys in mathematics and science, despite girls exhibiting superior reading abilities. Notably, the literature highlights the fluctuations of the gender gap across historical epochs and geographical regions. Such variability highlights the influence of cultural, social, and economic factors on the magnitude and direction of this gap, thereby guiding interventions and policies aimed at its mitigation. These nuances display the complexity of the grade gender gap and stress the need for further research and monitoring of the phenomenon.

2 Data description and RiminInRete database

The Italian pre-university education system comprises three main cycles of studies: primary school (PR, five years), lower-secondary school (S1, three years), and upper-secondary school (S2, five years). Prior to primary school, children often attend three years of non-compulsory nursery school, and after upper-secondary school around 60% of students enrol in university [1]. The curriculum is consistent across all primary and lower-secondary schools, while in upper-secondary schools there is a broader array of tracks available. These are typically categorized into high schools (i.e., *licei*), which offer a general education primarily aimed at preparing students for subsequent university studies, and vocational schools (i.e., *istituti tecnici* and *istituti professionali*), which supplement standard topics with specialized training in technical and professional fields. In the Italian grading system, student performance is evaluated on a scale from 0 to 10, with 10 being the highest achievable grade. Grades below 6 are considered insufficient (fail), while those above 6 indicate varying levels of proficiency and success in each discipline. In primary schools, this grading system (with the regulatory framework of OM 172/2020) recently underwent a substantial change from numerical grades to a judgement-based system.

The data analyzed in this work are a stratified random sample from the RiminiInRete (RIR) database (https://www.rimininrete.net/). RIR is an initiative established in 2015 by a network of schools in the province of Rimini, Italy, to create a comprehensive database of periodic and final assessments across all schools in the province, to assess the efficacy of the evaluation system and to monitor the influence of assessment methodologies on students' academic advancement. The longitudinal nature of the RIR database allows for the analysis of the grade gender gap and its comparison among different student cohorts and stages of the educational journey.

Table 1. Average GPA, standard deviation, and effect size (Cohen's d) by gender, academic cycle, and academic year in the stratified sample of 1920 students.

	Number of observations		**Mean**		**Stand. dev.**		**Cohen's d**
	Female	**Male**	**Female**	**Male**	**Female**	**Male**	
Primary (PR)	1165	1180	8.7618	8.4822	0.8477	0.8828	0.3231
Lower-secondary (S1)	784	775	7.9112	7.3964	1.0087	0.9320	0.5301
Upper-secondary (S2)	1202	1209	7.3721	6.9761	0.9565	0.9731	0.4103
a.y. 2014/15	512	501	7.9798	7.6038	1.1549	1.0783	0.3364
a.y. 2015/16	655	667	8.0326	7.6301	1.1112	1.1888	0.3498
a.y. 2016/17	631	645	8.0717	7.6413	1.1353	1.1759	0.3722
a.y. 2017/18	669	664	7.9861	7.6102	1.1026	1.1712	0.3305
a.y. 2018/19	684	687	8.0236	7.7069	1.0620	1.0958	0.2935

We select $n = 1920$ students from the RIR database using a stratified sample by gender (as recorded by the school from the ID documents) and year of birth (from 1997 to 2012) with $n_s = 60$ students per stratum. A summary of the stratified sample is reported in Table 1. We consider their grade point averages (GPAs) over five consecutive academic years from a.y. 2014/2015 to a.y. 2018/2019 for a total number of observations equal to 6351. It is worth noting that the years considered in our analysis are not affected by the recent change in the grading system.

3 Flexible Bayesian Beta regression model

Beta regression [7, 4] is a regression model tailored for response variables within the interval $(0,1)$. Unlike standard linear regression models (which work for unbounded continuous outcomes), Beta regression provides a coherent support for the response variable and appropriately handles skewness in errors. It can be easily generalized to responses in a bounded interval (a,b) via data transformations [5].

Denoting with $y_{i,t}$ the GPA of student i for the academic year t rescaled by 10.05 to fit into the interval $(0,1)$, we use the following Bayesian model:

$$\begin{aligned} y_{i,t} \mid \mu_{i,t}, \phi_i &\sim \text{Beta}(\mu_{i,t}\phi_i, (1-\mu_{i,t})\phi_i) \\ \text{logit}(\mu_{i,t}) &= \alpha^{(\mu)} + \beta^{(\mu)} x_i + \gamma_{x_i} \times t + g_{x_i}(z_{i,t}) + b_i^{(\mu)} \\ \log(\phi_i) &= \alpha^{(\phi)} + \beta^{(\phi)} x_i + b_i^{(\phi)} \end{aligned} \tag{1}$$

with

$$\begin{bmatrix} b_i^{(\mu)} \\ b_i^{(\phi)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^{(\mu)^2} & 0 \\ 0 & \tau^{(\phi)^2} \end{bmatrix} \right) \tag{2}$$

so that $\mathbb{E}[y_{i,t}] = \mu_{i,t} \in (0,1)$ and $\text{Var}[y_{i,t}] = \frac{\mu_{i,t}(1-\mu_{i,t})}{\phi_i+1}$. The regressors consider for the mean-submodel are the *gender* (x_i) of the student, the *academic year (t)* of the recorded GPA - whose regression coefficient (γ_{x_i}) is allowed to change based on the gender of the student - and a non linear effect g of the school level $z_{i,t}$ attended by the student at time t (from the first year of primary school to the fifth year of upper-secondary school) which is also allowed to change based on gender. The smooth function g is estimated via thin plate regression splines [12]. We assume the positive precision parameter ϕ_i to be time-independent and gender-dependent (through the regression coefficient $\beta^{(\phi)}$). Finally, we incorporate two student-specific random effects - one for the mean, $b_i^{(\mu)}$, and another for the precision, $b_i^{(\phi)}$ - to take into account unobserved individual-specific characteristics.

Priors for population-level linear effects are non-informative improper flat prior over the real numbers. Priors for the standard deviations of the random effects and splines are weakly-informative Half Student-t prior with 3 degrees of freedom. See [3] for more details on the prior choice.

The model was fitted using the R package `brms` [3] for Bayesian multilevel distributional regression. We run 2 chains of 3000 iterations each (of which 2000 warm-up).

4 Results

Table 2 and Figure 1 summarise the results. Our analysis reveals several notable trends in the academic performance. Firstly, females demonstrate a consistent advantage, outperforming males on average throughout the entirety of their school careers. In particular, for the a.y. 2014/2015, the expected GPA for a female student is 0.4503 points higher than for her male peers. Nonetheless, the signs of the gender-dependent coefficients γ_F and γ_M associated with time suggest a gradual reduction on average in the gender gap in more recent years. This led to an average grade gender gap of 0.3288 points in the year 2018/2019. Refer to Table 3 and Figure 1 for details on the temporal trend across the considered academic years.

Interestingly, we identify a significant positive coefficient for the scale parameter associated with male students, indicating that GPA tends to be less variable among male students compared to their female peers.

Figure 1, which includes both linear and non linear effects, showcases the estimated dynamic of the grade gender gap (represented in Figure 1 by the distance between the two curves) for different school grade levels. While changes in the gender gap across different school grade levels are relatively limited for any fixed academic year, the gap appears to undergo a slight increase during lower and upper secondary schools. For both males and females, we also note that grades exhibit a relevant drop from the end of one school cycle to the start of the next, which is followed by a small rebound (possibly stronger for females on average) both at lower and upper secondary school.

Table 2. Posterior estimates for the coefficients in the Beta regression model in 1. l-95% CI and u-95% CI indicate the lower and upper bound of the 95% credible interval for each coefficient, respectively.

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
μ submodel (logit scale)					
Intercept ($\alpha^{(\mu)}$)	1.4789	0.0239	1.4320	1.5262	1.01
SexM ($\beta^{(\mu)}$)	-0.2729	0.0348	-0.3391	-0.2000	1.01
SexF:a.y (γ_F)	-0.0073	0.0059	-0.0191	0.0041	1.01
SexM:a.y (γ_M)	0.0109	0.0054	0.0007	0.0213	1.01
ϕ submodel (log scale)					
ϕ_Intercept ($\alpha^{(\phi)}$)	4.4809	0.0537	4.3746	4.5870	1.00
ϕ_SexM ($\beta^{(\phi)}$)	0.2409	0.0737	0.0993	0.3861	1.00
SD of random effects					
SD(Intercept) ($\tau^{(\mu)}$)	0.5484	0.0125	0.5242	0.5734	1.01
SD(ϕ_Intercept) ($\tau^{(\phi)}$)	1.0504	0.0388	0.9754	1.1276	1.00

Table 3. Grade gender gap estimates by academic year. Top row: posterior mean. Bottom row: 95% credible interval.

a.y. 2014/2015	a.y. 2015/2016	a.y. 2016/2017	a.y 2017/2018	a.y. 2018/2019
0.4503	0.4199	0.3894	0.3591	0.3288
[0.3299, 0.5576]	[0.3104, 0.5129]	[0.2868, 0.4755]	[0.2641, 0.4453]	[0.2297, 0.4199]

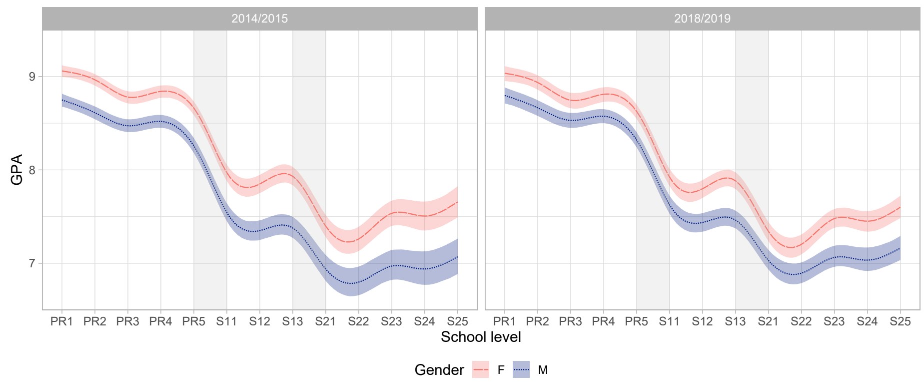

Fig. 1. Conditional effect of the school level by gender on GPA for two selected academic years. The areas in gray represent the change from one school cycle to the next.

5 Conclusions and further developments

This work highlights the existence of differences along the entire academic trajectories in a random sample of male and female students enrolled in schools of the province of Rimini between the academic years 2014/2015 and 2018/2019. Females outperform males on average across all the school levels, but the gender gap decreased on average in more recent years. To further investigate this gender gap, future work may focus on quantifying the discrepancy between performances within specific disciplines, e.g. gender gap in humanities or maths, with particular focus on gender-related changes from one school cycle to the next.

Another research direction is aimed at evaluating the impact on school grades of the COVID-19 pandemic and the subsequent disruption of in-person school attendance. To this aim, the plan is to extend the analysis to include data collected from academic years 2019/2020 to 2022/2023, which are now available from the RIR database.

Acknowledgements

We would like to acknowledge RiminInRete https://www.rimininrete.net/ for providing the data. Franzolini was supported by the National Recovery and Resilience Plan of Italy (PE1 FAIR - CUP B43C22000800006). Palma was supported by the UK Medical Research Council (MR/V020595/1).

Bibliography

[1] AlmaLaurea: Rapporto 2020 sulla condizione occupazionale e formativa dei diplomati (2020)

[2] Bondi, L., Franzolini, B., Palma, M.: A Bayesian longitudinal study of the grade gender gap. Working paper (2024)

[3] Bürkner, P.C.: brms: An R package for Bayesian multilevel models using Stan. Journal of statistical software **80**, 1–28 (2017)

[4] Ferrari, S., Cribari-Neto, F.: Beta regression for modelling rates and proportions. Journal of applied statistics **31**(7), 799–815 (2004)

[5] Geissinger, E.A., Khoo, C.L., Richmond, I.C., Faulkner, S.J., Schneider, D.C.: A case for beta regression in the natural sciences. Ecosphere **13**(2), e3940 (2022)

[6] Goodier, M., Aguilar García, C., Kirk, A., Scruton, P.: Gender gap shrinks and regional gap widens: 2023's key GCSE trends in england. The Guardian (2023)

[7] Kieschnick, R., McCullough, B.D.: Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. Statistical modelling **3**(3), 193–213 (2003)

[8] Malespina, A., Singh, C.: Gender gaps in grades versus grade penalties: why grade anomalies may be more detrimental for women aspiring for careers in biological sciences. International Journal of STEM Education **10**(1), 13 (2023)

[9] OECD: PISA 2022 results (Volume I): The state of learning and equity in education (2023)

[10] O'Dea, R.E., Lagisz, M., Jennions, M.D., Nakagawa, S.: Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. Nature communications **9**(1), 3777 (2018)

[11] Voyer, D., Voyer, S.D.: Gender differences in scholastic achievement: a meta-analysis. Psychological bulletin **140**(4), 1174 (2014)

[12] Wood, S.N.: Thin plate regression splines. Journal of the Royal Statistical Society Series B: Statistical Methodology **65**(1), 95–114 (2003)