Does Gender-grade-gap Exist? Evidence from Administrative Data from Afghanistan

Shishir Shakya* ⓡ Ahmad Shah Mobariz† ⓡ

*Assistant Professor, Department of Economics, Shippensburg University of Pennsylvania, Email: sshakya@ship.edu

†Doctoral Candidate, Department of Economics, University of Arkansas, Email: asmobari@uark.edu

Abstract

This paper studies the gender-grade-gaps using unique administrative data that comprises the universe of students taking the 2017 university entrance examination in Afghanistan. Using individual student-level information, we construct a household-level data set and report the gender-grade-gaps among different sibling structures within the households. We exploit the randomness in the variation of twin sibling's birth order and use the difference-in-difference (DD) strategy to compare the average grade between the second-born female (within different-sex twins siblings) with the second-born male (within same-sex male twins siblings). In other words, we estimate grade differences for a second-born female twin had she been born as a male child. We find gender-grade-gap exists. However, when we account for Afghanistan's cultural rectitude of sending siblings to gender-specific schools using the triple difference-in-difference (DDD) strategy, we find the gender-grade-gap dilutes significantly. Our results indicate that the cultural stigma drives the gender-related performance heterogeneity and *"not"* the inherent biological differences in the abilities even for a highly conservative country like Afghanistan.

Keywords: Gender, Grade, Gap, Culture, DDD, Twins, Afghanistan, Exam
JEL: J16, I20, I21, R20

1 Introduction

The literature on the gender performance gap identifies several determinants of the gender-grade-gap. Studies on the United States show that the socioeconomic inequalities drive the gender-related performance heterogeneity and *"not"* the inherent biological differences in the abilities (Pope and Sydnor, 2010). Other than socioeconomic inequalities, gender-related cultural differences affect performance disparity. For example, Nollenberger et al. (2016) show that cultural peculiarities about women's role in society are an important cause of gender disparity, and it explains nearly two-thirds of math grade variation. Meanwhile studies exhibit other factors play a role to inducing gender-grade-gap, for example, the subject matter (Le and Nguyen, 2018), examination environment (Niederle and Vesterlund, 2010), competition Jurajda and Münich (2011); Cai et al. (2019), parent's belief about children's ability (Jacobs, 1991), and neighborhood effects Chetty et al. (2016); Entwisle et al. (1994). However, literature within this domain does not entirely study the heterogeneity within households regarding gender-grade-gaps.

This paper studies the gender-grade-gaps using unique administrative data that comprise the universe of students taking the 2017 university entrance examination in Afghanistan. Using individual student-level information, we construct a household-level data set and report the gender-grade-gaps among different sibling structures within the household.

Furthermore, motivated by Jayachandran and Pande (2017) on the importance of birth order in developing countries, we exploit the randomness in the variation of twin sibling's birth order and use the difference-in-difference (DD) strategy to compare the grade between the second-born female (within different-sex twins) with the second-born male (within same-sex twins). This birth-order variable substitutes the time variable in the conventional DD framework. Our identification strategy's central idea is to develop a counterfactual grade for the second-born female, from the different-sex twins' siblings, by comparing the second-born male's grade from the same-sex twins' siblings. In other words, we estimate grade differences for a second-born female twin had she been born as a male child. We consider subgroups of different sex and same-sex twins siblings and randomize their birth orders to generate a placebo distribution of the gender-grade-gap by simulating 999 runs of DD strategies.

In our analysis, we also consider school-level competition and school-related cultural stigma or rectitude. We use school's latitudes and longitudes to develop a "school-choice" variable. This variable counts the number of schools within 10 kilometers of diameter. It provides an insight into the competition that each school faces, freedom of school choice for household, and possibly unobserved heterogeneities associated with rural-urban settings and income-related socioeconomic factors. In all our regression, we allow "school-choice" fixed-effect along with school-level fixed-effects. We also report heteroskedasticity robust standard errors that are two-way clustered at "school-choice" and school-level to capture intra-correlation of grades within these clusters.

School choice is another important factor that may produce a heterogeneous effect by gender (Hastings et al., 2006). Male and female students in Afghanistan go to single-sex schools. Suppose the school lacks infrastructure or cannot afford to maintain single-sex classes. In that case, it shifts to separate male and female students. In rural areas, however, if there are not sufficient teachers, they might have mixed classes. In large cities, parents have more school choices. They can choose to send children to exclusively boys or exclusively girls school. We account for such cultural dimension within the triple difference-in-differences (DDD) framework to exhibit the gender-grade-gap among twins who go to different school versus who go to the same school. We perform these DDD within the same identification strategy and regression adjustment, as explain earlier.

Our results show that overall there is a significant gender-grade-gap. Without accounting for the order of birth, we see a substantial difference between male and female twins. Female students grade consistently lower across specifications. The only instance in which the results are not significant is when we compare different gender twins who go to different schools. The DD estimates with placebo siblings match our first difference estimates. The triple difference coefficient is economically in the same direction as in the first difference, but it is not statistically significant. The results imply that the gender gap disappears when different sex siblings go to different schools.

We contribute to the literature on the gender-grade-gap in two aspects. First, our data is related to a conservative conflict-affected country where gender disparity is the highest in the world (UNDP, 2020). We show that despite accounting for society-level heterogeneity in highly conservative societies, the gender gap still exists among siblings. Second, we provide a unique identification strategy. Earlier research is anonymous on rejecting an inherent difference in natural ability but fails to control the differences that may arise from the household structure. We compare female and male twins that remove numerous biases such as age, birth order, and household income. By limiting our sample size to twin siblings, we remove such unobserved sources of bias.

Section 2 provides details on data and explains methodology. Section 3 exhibits our results. Finally, in section 4 concludes our study.

2 Data and methods

2.1 Data

We use Afghanistan's University Entrance Exam data for the year 2017, which comprises test grade data for the universe of students that participated in the university entrance in 2017. Data contains the individual's university entrance grade and their father's and grandfather's names, gender, province, district, and school-level information. We match the father's name, grandfather's name, and location and identify if each observation has siblings coming from the same family or not. Then we subset the data for the student who graduated and took university entrance examinations in the same year, 2017. Furthermore, we use each school's latitude and longitude and identify the number of schools within a 5-kilometers radius to capture the school-choice heterogeneity structure. The school-choice variable potentially allows us to capture unobserved heterogeneity associated with rural-urban settings and income-related socioeconomic factors.

2.2 Gender-grade-gap on different sub-samples

Using such individual data, we begin our analysis by comparing university entrance examination grades of female with a simple regression framework as:

$$grade_i = \alpha_0 + \alpha_1 gender_i + \nu_s + \mu_c + \epsilon_i \tag{1}$$

where *gender* is binary, 1 represents female, and 0 represents the male. β shows the effect size of gender on the grade. ν_s and μ_c are additive school-level and school choice-level fixed effects, which allows capturing the unobserved heterogeneities associated with school- and school choice-level. We run this specification on several sub-sample of the dataset.

Furthermore, we employ a simple difference-in-differences framework to understand the impact of gender on the university entrance grade. We compare female's grade considering two-types of family sibling structure —twin siblings where first-born is male while second-born is female (treatment group, $treat_i = 1$), and twins male sibling (comparison group $treat_i = 0$). Assuming the second-born male's grade as counterfactual for the second-born female's grade, we then study the grade differential for the second-born female would she had been a male child. We can analyze this with the following difference-in-differences framework.

$$grade_{it} = \beta_0 treat_i + \beta_1 post_t + \beta_3 (treat_i \times post_t) + \nu_s + \mu_c + \epsilon_{it} \tag{2}$$

where, $post_t = 1$ indicates second-born sibling and $post_t = 0$ indicates first-born sibling. ν_s and μ_c are additive school-level and school choice-level fixed effects.

In Afghanistan, male and female siblings are more likely to go to a different school than same-gender siblings. To account for this cultural stigma or rectitude into our analysis, we divide the treatment and comparison into two groups —those twins who go to a different school or $school = 1$, and those who go to the same school or $school = 0$. Then we perform the triple difference-in-differences (DDD) analysis as the following framework.

$$\begin{aligned} grade_{itg} =& \gamma_0 + \gamma_1 treat_t + \gamma_2 post_t + \gamma_3 school_g + \\ & \gamma_4 (treat_i \times post_t) + \gamma_5 (post_t \times school_g) + \gamma_7 (treat_i \times school_g) + \\ & \lambda (treat_i \times post_t \times school_g) + \nu_s + \mu_c + \epsilon_{itg} \end{aligned} \tag{3}$$

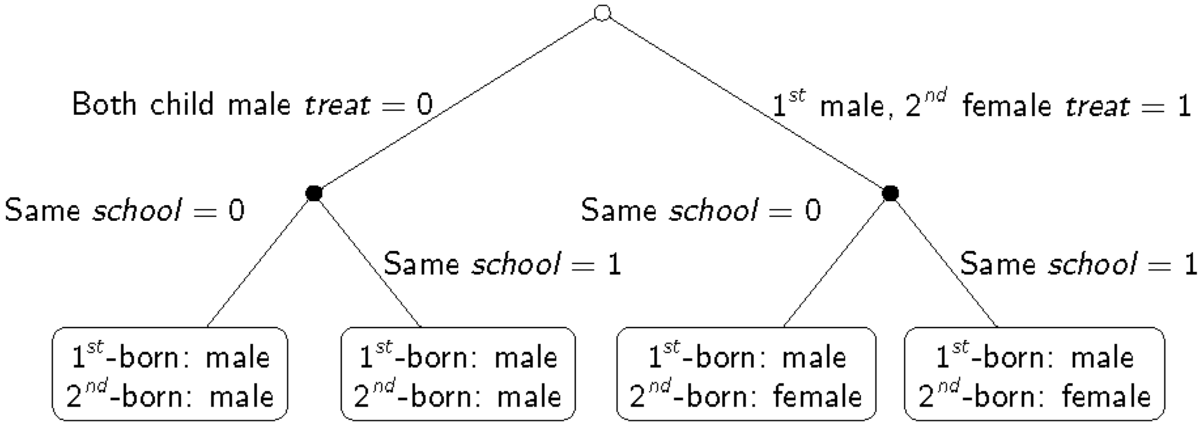
where, i, t, ν_s, μ_c, $post_t$, $treat_i$, $post_t \times treat_i$, have same interpretation as in equation 2. The coefficient λ is the triple difference-in-difference estimator for the effect of second-born female gender on the grade who study in different school than second-born female who study in same school.

$$\begin{aligned} \overline{\lambda} =& [(\overline{grade}_{treat=1,school=1,post=1} - \overline{grade}_{treat=1,school=1,post=0}) - \\ & (\overline{grade}_{treat=1,school=1,post=1} - \overline{grade}_{treat=0,school=1,post=0})] - \\ & (\overline{grade}_{treat=1,school=0,post=1} - \overline{grade}_{treat=1,school=0,post=0}) - \\ & (\overline{grade}_{treat=1,school=0,post=1} - \overline{grade}_{treat=0,school=0,post=0})] \end{aligned} \tag{4}$$

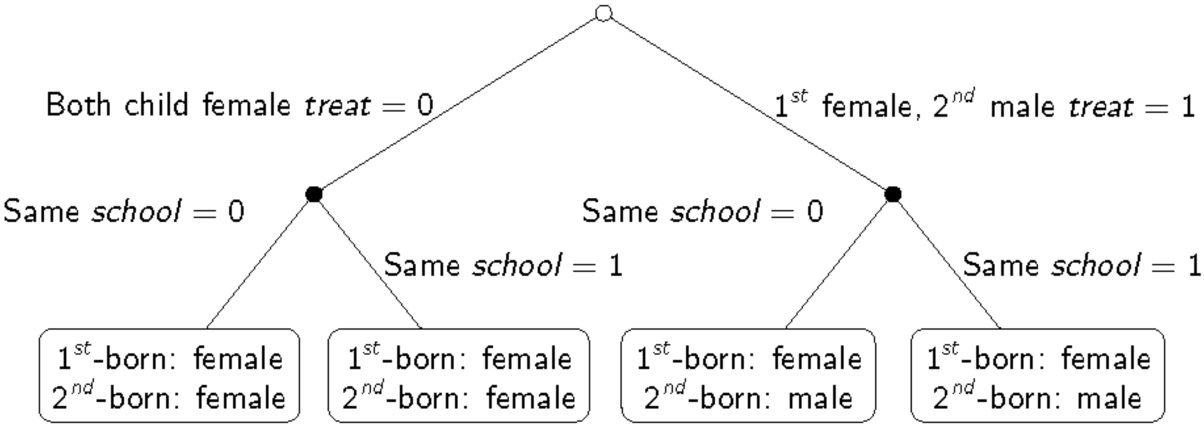
The triple difference estimator requires a parallel trend assumption for the estimated effect to have a causal interpretation. Even though the triple difference is the difference between two difference-in-differences, it does not need two parallel trend assumptions. Instead, it requires the relative outcome of siblings who go to a different school and who go to the same school to trend in the same way as the relative outcome of siblings who go to a different school and who go to the same school in the control state, in the absence of treatment. Figure 1, panel (a), provides a graphical depiction of triple difference-in-differences.

Similar to equation 2 and 3, we can run another analysis where we compare male's grade considering two-types of family sibling structure —twin siblings where first-born is female while second-born is male (treatment group, $treat_i = 1$), and twins female sibling (comparison group $treat_i = 0$). Assuming the second-born female's grade as

Figure 1: Graphical Depiction of Triple Difference-in-differences (DDD)



(a) Impacts on grade, when 2^{nd}-born is female



(b) Impacts on grade, when 2^{nd}-born is male

Notes: Same school = 0 represents the twin subgroups who go to different school.

counterfactual for the second-born male's grade, we can study what would be the grade had the second-born male would have been a female. Then considering cultural stigma or rectitude of male and female to go to a different school (more likely but not always), we can divide the treatment and comparison into two groups —those twins who go to a different school or *school* = 1, and those who go to the same school or *school* = 0. Then we perform triple difference-in-differences (DDD) analysis to estimate the effect of second-born male gender on the grade who study in a different school than the second-born male who studies in the same school female's grade. Figure 1, panel (b), provides a graphical depiction of triple difference-in-differences.

We assume that students' outcomes within the same school and the same household may be arbitrarily correlated within a school and within a household (or "cluster"). Hence we cluster the standard errors on school- and household-level and presented conservative standard errors robust to heteroskedasticity (Abadie et al., 2017).

3 Results

We begin our analysis by showing the gender-grade-gap estimates among different subsamples in Table 1. These results in column (1–8) are based on equation 1 from the section 2. However, we perform such regression for different subsamples.

In all these regressions, we implement fixed-effects to capture some level invariant heterogeneities. We indeed include "school-choice" and school-level fixed effects. The "school-choice" variable counts the numbers of schools within 10 kilometers of diameter. Including "school-choice" fixed-effects are likely to capture the competition that each school faces, freedom of school choice for household, and possibly unobserved heterogeneities associated with rural-urban settings and income-related socioeconomic factors. At the same time, the school-level fixed effects potentially absorb the school-level invariant unobserved heterogeneity. Furthermore, to allow the economic significance of estimates, we report highly conservative standard errors. Standard errors are heteroskedasticity robust with two-way clustered at "school-choice" and school-level to capture intra-correlation of grades within these clusters.

Table 1 column (1) shows males' average grade as 179.04, while the females' average grade, in column (2), is about 5 points less than males' average grade. These estimates are based on the full sample of 176,086 students who appeared in the 2017 university entrance examination in Afghanistan.

Table 1 column (1) and (2) provides a baseline. Still, these estimates are likely to be biased because these comparisons restrict to comparison average grades of homogeneous groups, who only differ in their gender or the simple terms of inference— apples-to-apples comparison. One way to mitigate such biases is to compare the grades of subsamples of twin siblings. Because it's more plausible that these twins are equally exposed to other unobservable factors, for example, equal parenting, locations, exposure to social/family constructs, among other things. We exhibit the results for twin siblings on Table 1 from column (3–8). Within the twin sibling subsample, in column (3), we find female's average grade is 2.82 points lower than male's average grade.

Table 1: University entrance grade: female versus male

	University entrance grade							
	Full	Full	Twins	Twins same school	Twins different school	Twins different gender	Twins different gender same school	Twins different gender different school
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Female=1		-5.08^{***}	-2.82^{***}	-3.23^{**}	-0.29	-5.12^{***}	-6.78^{***}	-1.16
		(0.33)	(1.06)	(1.33)	(2.10)	(1.80)	(1.91)	(4.15)
Male's grade	179.04^{***}							
	(0.08)							
Observations	176,086	176,086	19,548	11,068	8,480	4,868	966	3,902
R^2	-	0.254	0.298	0.348	0.320	0.362	0.453	0.363
$Adj - R^2$	-	0.235	0.206	0.228	0.168	0.190	0.231	0.185

Notes: ***, **, and * represent 10%, 5% and 1% levels of significance. Regressions includes "school-choice" and school-level fixed effects. Standard errors are heteroskedasticity robust with two-way clustered at "school-choice" and school-level.

In Afghanistan, male and female siblings are likely to attend different schools than same-gender siblings. To account for this cultural stigma in our analysis, we look into the gender-grade-gap of twins who went to the same school in column (4) and twins who attended different schools in column (5). The estimate of column (4) reveals twins who went to the same school, the female's average grade is 3.23 points lower than the male's average grade, but the column (5) shows twins who attended the different school, the gender-grade-gaps almost vanishes.

Compared to Table 1 estimate in column (2), the estimates in column (4) and (5) is likely to suppress bias drastically, we take one more step to absorb other unobservables linked with the twin sibling structure. So far, in columns (2–3), we only consider twins, but there can two categories of twins. First are same-sex twins: both twins are male, or both twins are female. Second are the different-sex twins: regardless of birth-order, one is female and the other male. We now consider these subgroups of different-sex twins within the same household and compare the gender-grade-gap. Column (6) exhibits the female's average grade is 5.12 points lower than the male's average grade, reflecting similar estimates as in column (2). However, suppose these different-sex twins attended the same school. In that case, the female's average grade is 6.78 points, substantially lower than the male's average grade. If they attended different schools, the gender-grade-gap perishes.

Our results are profound to deliver the insight that even in a very conservative country like Afghanistan, the gender-grade-gap is merely socioeconomic and cultural resultant rather than inherent biological differences in the abilities. Table 1 exhibits ranges of estimates regarding gender-grade-gap. However, to claim anything regarding "more likely causal" relationship, those estimates must undergo some stress testing. To tease out the causal relationship, we implement DD and DDD strategies and report our results in Table 2.

Our data is cross-sectional, and we also do not observe in the data if which of the twin sibling is first or second born. Hence performing conventional DD or the DDD estimations are not feasible. However, we can exploit the randomness in twin siblings' birth order variation to implement DD and DDD estimation creatively.

In Table 2 column (1) exhibits the DD results using the twins subsamples. We define birth-order variables as $post$, where $post = 1$ indicates second-born sibling and $post = 0$ indicates first-born sibling. We define treatment group or $treat = 1$ for twin siblings in which the first-born is male while the second-born is female, say MF. We define comparison group $treat = 0$ for twins male siblings, say MM. The interaction $treat \times post$ yields an interpretation of DD, which is the difference between two differences. First is the difference of average grade of first-born males in the treatment group compared with the average grade of the first-born males in the comparison group. Second is the difference in the average grade of second-born females in the treatment group compared to second-born males in the comparison group. In simpler words, the coefficient of $treat \times post$

Table 2: gender-grade-gap: placebo and counterfactual analysis

Variable	MM, MF		FF, FM	
	DD (1)	**DDD** (2)	**DD** (3)	**DDD** (4)
treat	1.36 (-0.36, 3.07)	1.79 (-1.72, 5.36)	-1.00 (-2.57, 0.63)	-2.44 (-6.02, 0.94)
post	-0.01 (-1.18, 1.15)	-3.73 (-5.32, -2.12)	-0.02 (-1.42, 1.33)	-0.96 (-3.46, 1.42)
treat $\times$ *post*	-4.64*** (-7.36, -1.86)	-6.82*** (-10.06, -3.38)	4.28*** (1.21, 7.16)	6.77*** (3.43, 10.24)
school		-0.02 (-1.19, 1.08)		-0.04 (-1.33, 1.21)
treat $\times$ *school*		1.43 (-2.88, 5.88)		2.51 (-2.17, 6.65)
school $\times$ *post*		0.02 (-3.13, 3.09)		0.09 (-4.55, 4.93)
treat $\times$ *school* $\times$ *post*		4.40 (-1.54, 10.76)		-5.88 (-13.41, 1.63)

Notes: ***, **, and * represent 10%, 5% and 1% levels of significance. Enclosed in parentheses are 95% confidence interval of mean estimates. Regressions includes "school-choice" and school-level fixed effects. Standard errors are heteroskedasticity robust with two-way clustered at "school-choice" and school-level.

shows the impacts on grade among second-born females, after accounting for the counterfactual —had these second-born females had born as a second-born male. We being female means a reduction of 5.12 points than the male's average grade.

In Table 2 column (2), we consider the same subsample and the same meaning for *post* and *treat* as in column (1). However, now we incorporate two types of the group —those twins who go to a different school or *school* = 1, and those who go to the same school or *school* = 0. The coefficient of interest is the coefficient of the interaction *post* $\times$ *treat* $\times$ *school*, which yields an interpretation of the treatment effect. The effect size or the difference between the grade of second-born females who went to a different school than their first-born male sibling and the grade of second-born males who went to a different school than the first-born male sibling. The distribution of such treatment effect centers at 4.40 but is likely not to be significantly away from the null effect, indicating that after accounting for the cultural norm of attending gender-specific schools, the gender-grade-gap dilutes.

Table 2 columns (3) and (4) are essentially replicas of columns (1) and (2), except we define treatment group slightly differently. In columns (3) and (4), we represent the treatment group or $treat = 1$ for twin siblings in which the first-born is female while the second-born is male, say FM. We define comparison group $treat = 0$ for twins female siblings, say FF. In column (3), the coefficient of interaction $treat \times post$ shows the impacts on grade among second-born males, after accounting for the counterfactual — had these second-born males had born as a second-born female. We find being male means additional 4.28 points grade than the female's average grade. Again, this estimate does not account for the cultural norm of attending gender-specific schools. Once we account for such in column (4), we find gender-grade-gap centers around 5.88 but is likely not to be significantly away from the null effect indicating the gender-grade-gap dilutes. In other words, the difference –between the grade of second-born males who went to a different school than their first-born male sibling and the grade of second-born females who went to a different school than the first-born female sibling– is statistically zero.

4 Conclusion

This paper presents causal estimates for the gender-grade-gap. We use unique student test score data with information about students' household identifiers to identify siblings who took Afghanistan's University Entrance Examination in 2017. Our estimates show that with strict bias-reduction measures, the gender score gap shrinks. On average, in the full sample, female students score approximately 5 points lesser after accounting for unobserved school heterogeneity. The average difference in the twin sample is approx. 3

points. However, the gap vanishes when different gender twins go to different schools. Hastings et al. (2006) also show that school choice, which may be driven by parental decisions, has a heterogeneous effect across gender.

In a narrower sample size, we identify twin siblings and the gender composition of twin births. The gender of twin birth is random by nature. To further account for unobserved heterogeneity in birth order and gender composition of twins, we construct a placebo in 999 simulations. In the placebo, we define a counterfactual for female twins by randomizing twins' birth order so that a second-born male in a twin birth will work as a counterfactual for a female twin in a same-sex twin structure. Using the placebo, we estimate a difference-in-difference model where the female in a twin structure is the treatment indicator. The order of birth is the second is the time dimension. The DD estimate in the placebo analysis confirms our first difference estimates.

Moreover, we test the effect of cultural particularities in our placebo sample. We interact with the same school variable with birth order and treatment. This triple difference-in-difference estimate gives the difference between a female twin whose birth order is second that goes to a different school than her twin brother compared to the same female twin that goes to the same school. This strategy also shows that with the introduction of separate schools for male and female students, the gender gap diminishes to zero. Our finding contrast with Doris et al. (2013) that suggest potentially widening gender gap in a single-sex education system. Our findings also differ from Brenoe and Zölitz (2020) who show women's likelihood of enrollment graduation in STEM fields reduces when the number of female peers increases.

We contribute to the literature by improving the identification. Figlio et al. (2016) estimates school choice effect by gender using siblings data from Florida. We use twins data to account for the birth order effect. This identification removes unobserved biases from an unmeasured family raised by Figlio et al. (2019), and birth order characteristics. Through the school choice as a measure of cultural norms, and data from a conservative conflict-affected country, we also contribute to the literature on culture norms, and gender gap such as Rodríguez-Planas and Nollenberger (2018); Zhang (2019); Friedman-Sokuler and Justman (2020).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.

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

Cai, X., Lu, Y., Pan, J., and Zhong, S. (2019). Gender gap under pressure: Evidence from china's national college entrance examination. *Review of Economics and Statistics*, 101(2):249–263.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., and Scuderi, B. (2016). Childhood environment and gender gaps in adulthood. *American Economic Review*, 106(5):282–88.

Doris, A., O'Neill, D., and Sweetman, O. (2013). Gender, single-sex schooling and maths achievement. *Economics of Education Review*, 35:104–119.

Entwisle, D. R., Alexander, K. L., and Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review*, pages 822–838.

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.

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., et al. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics*, 11(3):338–81.

Friedman-Sokuler, N. and Justman, M. (2020). Gender, culture and stem: Counter-intuitive patterns in arab society. *Economics of Education Review*, 74:101947.

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2006). Gender and performance: Evidence from school assignment by randomized lottery. *American Economic Review*, 96(2):232–236.

Jacobs, J. E. (1991). Influence of gender stereotypes on parent and child mathematics attitudes. *Journal of Educational Psychology*, 83(4):518.

Jayachandran, S. and Pande, R. (2017). Why are indian children so short? the role of birth order and son preference. *American Economic Review*, 107(9):2600–2629.

Jurajda, Š. and Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to czech universities. *American Economic Review*, 101(3):514–18.

Le, H. T. and Nguyen, H. T. (2018). The evolution of the gender test score gap through seventh grade: new insights from australia using unconditional quantile regression and decomposition. *IZA Journal of Labor Economics*, 7(1):2.

Niederle, M. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2):129–44.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

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

Rodríguez-Planas, N. and Nollenberger, N. (2018). Let the girls learn! it is not only about math… it's about gender social norms. *Economics of Education Review*, 62:230–253.

UNDP (2020). United Nations Development Programme, Human Development Reports.

Zhang, Y. J. (2019). Culture, institutions and the gender gap in competitive inclination: Evidence from the communist experiment in china. *The Economic Journal*, 129(617):509–552.