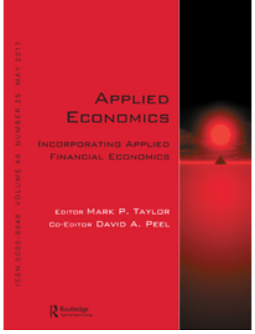

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

The negative consequences of school bullying on academic performance and mitigation through female teacher participation: evidence from Ghana

Shahriar Kibriya, Zhicheng Phil Xu & Yu Zhang

To cite this article: Shahriar Kibriya, Zhicheng Phil Xu & Yu Zhang (2017) The negative consequences of school bullying on academic performance and mitigation through female teacher participation: evidence from Ghana, Applied Economics, 49:25, 2480-2490, DOI: 10.1080/00036846.2016.1240350

To link to this article: https://doi.org/10.1080/00036846.2016.1240350

Published online: 12 Oct 2016.

APPLIED ECONOMICS, 2017
VOL. 49, NO. 25, 2480–2490
http://dx.doi.org/10.1080/00036846.2016.1240350

Routledge
Taylor & Francis Group

The negative consequences of school bullying on academic performance and mitigation through female teacher participation: evidence from Ghana

Shahriar Kibriya[a], Zhicheng Phil Xu [b] and Yu Zhang[c]

[a]Center on Conflict and Development, Texas A&M University, College Station, TX, USA; [b]School of Economics, Henan University, Kaifeng, China; [c]Department of Agricultural Economics, Texas A&M University, College Station, TX, USA

ABSTRACT
Exploiting data from Ghanaian schools' eighth grade students collected in 2011, we estimate the causal effects of school bullying on academic achievement and gender-based mitigating approaches by using propensity score matching (PSM) and doubly robust (DR) estimator approach. We find that students victimized by bullying score at least 0.22 standard deviation lower than their peers in a standardized mathematics examination. Meanwhile, we document that the effect of bullying is significantly attenuated in the presence of female teachers in the classroom. These results hold through a set of robustness checks including placebo regressions and matching quality test. We explain the results through gender difference in teaching paradigm and conclude that a feminine management approach in class is required to reduce the effect of bullying.

KEYWORDS
School bullying; academic performance; propensity score matching; doubly robust estimator; Ghana

JEL CLASSIFICATION
I21; I28; J13

If there's one goal of this conference, it's to dispel the myth that bullying is just a harmless rite of passage or an inevitable part of growing up.

US President Barack Obama, White House Anti Bullying Conference, 2011

I. Introduction and literature review

Bullying in educational institutions is a global phenomenon.[1] For instance, 50% students reported being bullied in an international survey conducted in 2011, consisting of more than 300,000 students from 48 countries (Mullis et al. 2012). Evidence of school persistent bullying has also been documented in Ghana as well as the United States and other European countries (Ammermueller 2012; Brown and Taylor 2008; Dunne et al. 2013; Nansel et al. 2001; Ponzo 2013). Due to the prevalence of school bullying, recently there has been a rising academic interest to precisely quantify the consequence of school bullying. Le et al. (2005) study a sample of twins chronologically in Australia and show that childhood conduct disorder can adversely affect an individual's academic attainment and competency in the labour market. Brown and Taylor (2008) explore the same question in Britain, and find similar results. Using a much broader data set including 11 European countries, Ammermueller (2012) finds that being bullied has a significantly negative impact on students' performance both in school and the labour market. Notwithstanding significant correlation between bullying and educational achievement in those studies, the causal direction remained unclear until recently. It is possible that a student has a lower academic performance because of being a victim, or the likelihood of a student being bullied is higher if he performs poorer. Furthermore, there could be omitted variables affecting both the likelihood of being victimized and academic performance, leading to biased estimates. Ponzo (2013) thereby overcomes these problems by measuring the effects of school bullying through a propensity score matching (PSM) approach. She concludes that school bullying decreases student performances in both fourth and eighth grades in Italy.

CONTACT Zhicheng Phil Xu phil.zhicheng.xu@gmail.com School of Economics, Henan University, 85 Minglun street, Kaifeng, Henan 475000, China

[1]Olweus pioneers in the systematical study on school bullying in the 1970s. Olweus (1993) defines a student being bullied at school 'when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other students'. These negative actions include to attack or discomfort someone physically or verbally, spreading rumours, and intentionally excluding someone from a group.

Research is yet to be conducted on bullying and academic performance for the low-income developing nations which may be significantly different from students from affluent countries due to differentiated access to facilities and economic conditions. Another glaring gap in both academia and policy remains on the proof of possible channels to attenuate the effect of bullying. In this article we attempt to contribute to the existing body of literature through three folds. First, we find further causal evidence of bullying impacting academic performance more severely in the context of a developing nation, Ghana. We choose Ghana as the research subject, since it is one of the poorest countries in the world, the score of which is also ranked at the bottom among 42 countries that assessed eighth grade students at the TIMSS 2011. Second, by examining the heterogeneous effects through PSM and doubly robust (DR) estimator we suggest that presence of a female teacher in the classroom will reduce the negative effects of bullying. Third, our results are validated through several robustness tests of PSM including overlap check, matching quality test, and placebo regression (Imbens and Wooldridge 2009).

Our heterogeneous analysis also contributes to the literature pertaining to the role of the teacher on student achievement (Fryer 2013; Metzler and Woessmann 2012). Conventional wisdom based on educational studies in developing countries has long held that the teacher's gender is associated with student performance (Saha 1983; Warwick and Jatoi 1994). That is, pupils with male teachers achieve better in mathematics and science than those with female teachers. Nevertheless, we find empirical evidence that females are more capable than their male counterparts in mentoring bullied students. We also discuss the underlying mechanism behind the findings, emphasizing the gender differences in teachers' teaching paradigms (Gray 1987). While victims of bullying are less willing to attend and engage in class (Dunne et al. 2013; Ripski and Gregory 2009), female teachers' mentoring methods may be more directed towards mitigating the effects of bullying. For instance, female teachers tend to use class discussion more frequently and promote more collaborative learning environments than male teachers (Singer 1996).

The remaining parts of this article proceed as follow. The next section describes the data. Section III presents the research methodology. Section IV delivers the results and interpretations, with special attention paid to the heterogeneity analysis. We check the robustness in Section V, while Section VI provides concluding remarks.

II. Data description

The International Association for the Evaluation of Educational Achievement (IEA) has conducted the Trends in International Mathematics and Science Study (TIMSS) in the past two decades. The TIMSS data set is enriched by the comprehensive background information related to students and their households, teachers, and schools. In this article, we use the eighth grade math score of Ghanaian students as the measurement of their academic performance. IEA employed a rotated test design in order to balance valid measures of student achievement and reasonable testing time. Students' answers to the tests were scored as correct, partially correct, or incorrect. Based on item response theory (IRT) scaling with marginal estimation, each student's math performance in a set of test questions is then recorded as five plausible values calculated by the expectation and maximization (EM) algorithm.[2] Rubin's (1987) combination rules are used to estimate a variable that is measured by plausible values: the estimates are calculated for each plausible value and then averaged. However, as shown in Table A1 in the appendix, our results remain robust to using any of the five alternative test values. Students' achievements on math are reported with a scale of 0–1000, while their typical scores fall in the range from 300 to 700 and the international centrepoint is 500.

In Ghana, 7323 eighth grade students participated in 2011-TIMSS. All students and their associated schools were randomly chosen. The average score of Ghanaian students, which is barely above 300, is the lowest among 42 countries that assessed eighth grade students at the TIMSS 2011. Even the top five percentile students in Ghana score lower than the international centrepoint. The survey contains a set

[2]A detailed review of the plausible values methodology is given in Mislevy (1991).

of questions regarding whether students suffer from school bullying. These questions were:

> '*During this year, how often were you made fun of or called names at school?*'
> '*During this year, how often were you left out of games or activities by other students at school?*'
> '*During this year, how often did someone spread lies about you at school?*'
> '*During this year, how often was something stolen from you at school?*'
> '*During this year, how often were you hit or hurt by other student(s) at school?*'
> '*During this year, how often were you made to do things you didn't want to do by other students at school?*'

Each respondent was asked to select one out of the following options: 'once a week', 'once or twice a month', 'few times a year', or 'never'. Based on the answers collected from the respondents, the TIMSS data set constructs a measure indicating school bullying. Specifically, a student is graded as being 'bullied weekly' if he/she experienced three of the six bullying behaviours 'once or twice a month' and the other three 'a few times a year', or he/she suffered more than that. Besides school bullying, we also include four clusters of controls that explain students' academic achievements. The first cluster is student's individual characteristics, such as student age and gender. Student's household characteristics, including parents' education level, as well as five indicators on home support for education, comprise the second cluster.[3] The third cluster includes teacher characteristics, consisting of teacher's experience, gender, and education level, while the fourth cluster consists of school characteristics, including school location, percentage of students coming from economically disadvantaged families, school enrolment, and the number of computers as a proxy of school facility.

Table 1 reports the descriptive statistics for the variables discussed above. It shows that the ratio of females to males is close to 1:1. The average age of eighth grade students is approximately 16 years. The statistics show that Ghanaian students score only about 330 points, almost 70 points lower from the Low International Benchmark (400). Furthermore, female students score 24 points behind the male students. Bullying is found to be pervasive in Ghana; with more than one half of surveyed students are bullied weekly. The likelihoods of being bullied are almost equal between the male and female students. Other variables of interest discussed above are also depicted in Table 1.

Table 1. Descriptive statistics.

	TIMSS 2011 eighth grade	
Variables	Mean	Standard deviation
Outcome: math score	333.007	85.620
Male	344.474	86.146
Female	320.489	83.438
Treatment: bullied weekly	0.530	0.499
Male	0.533	0.499
Female	0.528	0.499
Student age	15.744	1.512
Female student	0.478	0.500
Parents' education level		
University or above	0.106	0.307
Post-secondary	0.160	0.366
Upper secondary	0.221	0.415
Lower secondary	0.309	0.462
Primary or no school	0.204	0.403
Computer possession	0.250	0.433
Study desk	0.506	0.500
Own room	0.318	0.466
Internet at home	0.112	0.316
Books at home		
0–10	0.401	0.490
11–25	0.368	0.482
26–100	0.139	0.346
101–200	0.043	0.204
>200	0.048	0.214
School location		
Urban	0.178	0.382
Suburban	0.166	0.372
Large town	0.167	0.373
Small town or village	0.392	0.488
Remote rural	0.098	0.297
Portion of students from disadvantage families		
0–10%	0.066	0.248
11–25%	0.112	0.315
26–50%	0.161	0.367
More than 50%	0.662	0.473
School enrolment	265.153	213.922
School computer		
1 computer for 1–2 students	0.443	0.497
1 computer for 3–5 students	0.118	0.323
1 computer for 6 or more students	0.290	0.454
No computers available	0.149	0.356
Years teacher has been teaching	8.266	6.557
Female teacher	0.121	0.326
Teacher education level		
Upper secondary education	0.079	0.270
Post-secondary non-tertiary level of education	0.450	0.498
Short tertiary education	0.193	0.394
Long tertiary education	0.274	0.446
University or higher	0.004	0.066
Observations	7323	

[3]These home support indicators include computer possession, study desk, having their own room, internet accessibility, and number of books at home.

III. Research methodology

We initiate our study following Ponzo's (2013) strategy of estimating OLS and PSM. The DR estimator is also employed since it is less sensitive to model specification (Imbens and Wooldridge 2009). We commence our analysis by estimating the following model using OLS:

$$Y_i = \beta_0 + \beta_1 bullied_i + \beta_2 X_i + \varepsilon_i, \quad (1)$$

where Y_i denotes the math score of student i, $bullied_i$ is a binary variable indicating whether or not the student has been a victim of school bullying, X_i is a vector of controls (including student, household, teacher, and school characteristics), ε_i is an error term capturing idiosyncratic shocks or unobserved characteristics. β_1 represents the effect of our major interest, that is, the expected mean gap in academic performance between bullied and non-bullied students. However, OLS estimation may be biased due to endogeneity issues.

A matching method can overcome the problems. Intuitively, it matches pairs of individuals by the characteristics from control (non-bullied students) and treatment (bullied students) groups. Accordingly, a pair of matched individuals is essentially similar in all aspects but randomly assigned into control or treatment group. Hence, the matching method makes the comparison between treatment and control group immune to selection bias. Given a large number of covariates in our analysis, the method of PSM proposed by Rosenbaum and Rubin (1983) is mostly suitable in order to avoid the curse of dimensionality. That is, PSM compresses the multidimensional covariates into a propensity score, which refers to the conditional probability of being assigned to treatment group (Abadie and Imbens 2009).

Formally, the propensity score is defined as $P(x) = \Pr(T = 1|X = x)$. T denotes the binary treatment variable (T = 1 if bullied, or 0 otherwise); Y denotes the outcome (math score); and X contains a vector of background variables. Let $Y(0)$ and $Y(1)$ indicate the potential outcomes under control and treatment, respectively. PSM approach is credible if the unconfoundedness and overlap conditions hold: (a) $Y(0), Y(1) \perp T|P(X)$ and (b) $0 < P(T = 1|X) < 1$. The unconfoundedness condition implies that the potential outcomes are independent of treatment conditional on background variables X. The overlap condition requires that for each treated individual there is at least one matched individual in the control group.

We will focus on the average treatment effect on the treated (ATT) instead of the average treatment effect (ATE) as suggested by Heckman (1997). ATT is of greater interest than the ATE because ATE includes the effects on students that were never bullied, whereas ATT explicitly evaluates the effects on those who were actually bullied. Moreover, the ATT estimation requires less restrictive assumptions than ATE. Formally, ATT is given by

$$E[(Y_{1i} - Y_{0i})|T = 1], \quad (2)$$

where Y_{0i} is the value of the outcome variable for individual i if she is not bullied, and Y_{1i} is the value of the outcome variable for individual i if she is bullied.

Several matching algorithms will be implemented, namely nearest neighbour (NN), radius, and Kernel matching (Caliendo and Kopeinig 2008; Imbens 2014). First, NN algorithm matches each bullied student with the non-bullied counterpart with the closest propensity score. NN algorithm is applied with replacement, since a non-bullied student can be an ideal match for more than one bullied student. Second, radius algorithm matches each bullied student with all non-bullied students whose propensity scores fall into predefined neighbour of the propensity score of the bullied student. We set the radius of the neighbourhood as small as 0.005. Finally, we apply Kernel algorithm, which matches each bullied student with a weighted average of non-bullied students. We estimate the weights by the Epanechnikov Kernel function where the bandwidth is 0.06, following Heckman, Ichimura, and Todd (1997).

Moreover, we take the advantages of the DR estimator to reduce the potential bias raised by the propensity score misspecification. DR estimation requires building two models: one to predict outcome (math score) and the other to predict the treatment status (being bullied). The remarkable feature of the DR estimator is that, as long as one of the two models is correctly specified, we could obtain unbiased estimates of the treatment effect. Mathematically, the DR estimator is:

$$\hat{\Delta}_{DR} = n^{-1}\sum_{i=1}^{n}\left[\frac{T_iY_i}{e\left(\boldsymbol{X}_i\hat{\beta}\right)} - \frac{\left\{T_i - e\left(\boldsymbol{X}_i\hat{\beta}\right)\right\}}{e\left(\boldsymbol{X}_i\hat{\beta}\right)}m_1(\boldsymbol{X}_i\hat{\alpha}_1)\right] - n^{-1}\sum_{i=1}^{n}\left[\frac{(1-T_i)Y_i}{1-e\left(\boldsymbol{X}_i\hat{\beta}\right)} - \frac{\left\{T_i - e\left(\boldsymbol{X}_i\hat{\beta}\right)\right\}}{1-e\left(\boldsymbol{X}_i\hat{\beta}\right)}m_0(\boldsymbol{X}_i\hat{\alpha}_0)\right], \quad (3)$$

where $e\left(\boldsymbol{X}_i\hat{\beta}\right)$ is the postulated model for the true propensity score, $m_1(\boldsymbol{X}_i\hat{\alpha}_1)$ and $m_0(\boldsymbol{X}_i\hat{\alpha}_0)$ are postulated models for the true regressions $E[Y|\mathrm{T}=1,\mathbf{X}]$ and $E[Y|\mathrm{T}=0,\mathbf{X}]$. The covariates of the two models are the same as discussed above.

Next, to investigate if bullied students perform better when they have access to female teachers, we employ the heterogeneous treatment effect estimation. First, we separate the full sample into two parts by any possible moderator (e.g. teacher gender), and then apply PSM approach to these subsamples. The difference of the subsample ATT is as follows:

$$ATT_{diff} = E[(Y_{1i} - Y_{0i})|T=1,\mathrm{M}=1] - E[(Y_{1i} - Y_{0i})|T=1,\mathrm{M}=0], \quad (4)$$

where M denotes the moderator variable. An ideal moderator can prominently mitigate the effects of bullying if the differences of ATT estimates are significantly positive.

IV. Empirical results

We start reporting the results by comparing the distributions of scores between bullied and non-bullied students in Ghana. As shown in Figure 1, the score distribution of bullied students' appears to be shifted to the left; thus visually implying a lower performance due to the adverse effect of bullying.

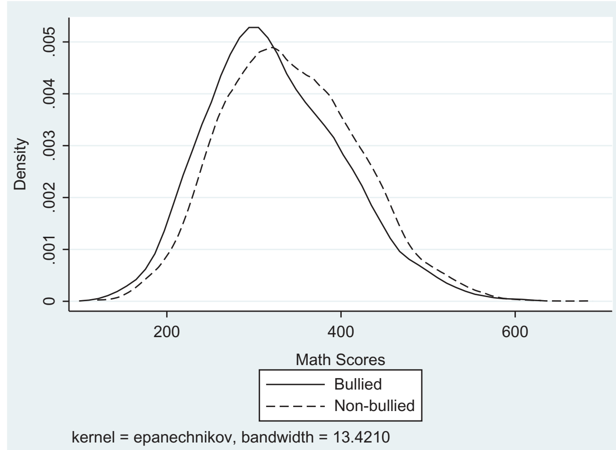

Figure 1. Kernel density estimates, bullied versus non-bullied.

Table 2 contains the estimation results by the OLS approach. Column (1) shows the simplest specification. Only 'bullied', our main variable of interest, is included in the model. In column (2), we add several variables to control for individual and household characteristics: student age, student gender, parents' highest education level, and a set of household facilities. Column (3) includes additional teacher characteristics: teacher's experience, gender, and education level. In column (4), we control for school characteristics: school location, school enrolment, portion of students coming from disadvantaged families, and number of computers. Column (5) adjusts for teacher and school characteristics in an alternative way: school fixed effects are included instead of a set of control variables. In all OLS specifications, standard errors are clustered on school level.

Table 2. Impacts of being bullied weekly on math performance (OLS).

	(1)	(2)	(3)	(4)	(5)
Bullied	−18.144***	−20.440***	−19.630***	−18.634***	−15.447***
	(3.775)	(3.223)	(3.355)	(3.143)	(2.553)
Student age		−13.969***	−13.081***	−10.454***	−6.256***
		(1.636)	(1.573)	(1.452)	(1.069)
Student female		−28.785***	−27.910***	−28.048***	−27.223***
		(3.332)	(3.393)	(3.311)	(2.592)
Household controls	No	Yes	Yes	Yes	Yes
Teacher controls	No	No	Yes	Yes	No
School controls	No	No	No	Yes	No
School fixed effects	No	No	No	No	Yes
Observations	7323	5503	5002	4514	5503
R-squared	0.011	0.140	0.157	0.231	0.456

'Household controls' include: parental education, number of books at home, computer possession, study desk, own room, and Internet accessibility. 'Teacher control' includes teacher's gender, experience, and education level. 'School' controls' include: school location, school enrolment, portion of students coming from disadvantaged families, and instructional computer accessibility. Standard errors are adjusted for school-level clustering. *** indicates 1% level statistical significance.

Although the coefficients of 'bullied' fluctuate from −15.4 to −20.4 across all five specifications, they remain statistically significant at the 1% level. This implies at least a reduction of 20% standard deviation of the sample mean score. The magnitudes of the coefficients create a decreasing trend while the models become more comprehensive. This is reasonable because some control variables may be correlated with school bullying, leading to the overestimate of the impact on bullying. We also find that male and younger students perform better. The rest of the controls appear not to be significantly correlated with students' performance.

Table 3 reports the results from the PSM and DR estimation. The first step of the PSM approach is to predict the propensity score, that is, the probability of student being bullied conditional on pretreatment control variables. For the sake of brevity, we do not report the Logit estimation of propensity score. Three matching algorithms are employed to estimate the ATT: NN matching, radius matching and Kernel matching. NN matching result suggests that on average students being bullied at school achieve 17.1 points lower than their non-bullied fellows in math exam. The similar results generated by radius matching and Kernel matching, about 18.5 and 18.3 points lower test scores respectively, support the above finding. The standard errors in Table 3 are estimated by using bootstrap procedures. We also estimate the Abadie–Imbens standard errors, as a necessary validation to the bootstrap procedure (Abadie and Imbens 2008). The result from the Abadie–Imbens estimator turns out very similar to those shown in Table 3 and provided in Table A2 in the Appendix. The DR estimation result is presented in the bottom of Table 3. It provides slightly larger estimates than the PSM method; a victim scores about 19.4 points lower on average.

To explore whether our estimates are representative of Ghanaian population of students, we also perform weighted least squares (WLS) regression with sampling weights provided by the TIMSS data set. Besides, we also provide PSM estimates accounting for sampling weights. The results, presented in Table A3 in the appendix, are quite similar with the OLS and PSM results in Tables 2 and 3, respectively. Taken together, sampled students across various demographic groups in TIMMS properly and proportionally represent the population estimation.

Do students from developed and developing countries are affected by school bullying in the same degree? To gauge magnitude, we compare our results based on Ghanaian sample to a similar study in Italy (Ponzo 2013). Italian eighth grade students being bullied achieve about 13 points less in mathematics, which is a decrease of 0.18 standard deviation in the outcome measure; the counterparts in Ghana obtain about 18 points less, or a reduction of 0.23 standard deviation in math test scores. It thus appears that bullying leads to more serious impacts to students in the developing countries, which makes sense since they are lack of better access to socio-economic institutions.

Next, we report the heterogeneous effect of teacher gender in the presence of bullying by separately estimating the ATTs according to teacher gender. Table 4 displays a plausible channel to alleviate the impairment of bullying through female teachers through both PSM and DR approaches. The existing adverse effect of bullying sharply declines towards 0 in the presence of a female teacher. On the contrary,

Table 3. Impacts of being bullied weekly on math performance (PSM and DR).

Methods	Eighth grade math scores
Nearest neighbour	−17.137***
	(4.430)
Number of treated	2357
Number of controls	2081
Radius/caliper	−18.547***
	(2.705)
Number of treated	2341
Number of controls	2024
Epanechnikov Kernel	−18.300***
	(2.601)
Number of treated	2357
Number of controls	2081
Doubly robust	−19.373***
	(2.557)

For matching methods: Balancing property and common support are satisfied; Nearest neighbour is applied with replacement; Standard errors, estimated by 100 bootstrap replications, are reported in parentheses. *** indicates 1% level statistical significance.

Table 4. Impacts of being bullied weekly (female versus male teachers).

Methods	Female	Male
Nearest neighbour	−2.491	−20.011***
	(10.759)	(4.725)
Number of treated	276	2077
Number of controls	221	1856
Doubly robust	−10.069	−20.225***
	(6.593)	(2.666)

For matching methods: Balancing property and common support are satisfied; Nearest neighbour is applied with replacement; Standard errors, estimated by 100 bootstrap replications, are reported in parentheses. *** indicates 1% level statistical significance.

in absence of female teachers, bullying leads to a remarkable reduction of 20 points on average. The discrepancy might be explained by their distinct classroom management practices. Specifically, female educators tend to implement a 'feminine' style of management behaviour, while the male peers follow a 'masculine' paradigm (Gray 1987). Under a 'feminine' paradigm, female teachers are more nurturing, affectionate, and empathic than their male counterparts. They tend to be more responsive to school bullying, and more willing to help the victims (Casey and Fuller 1994; Martin and Ross 2005). For example, female teachers determine situations to be more severe than the males. They would communicate with the bullied students, find peer support, and seek parental engagement. In addition, the teaching behaviours of females tend to promote more class engagement and collaboration, which are of great importance in mitigating the adverse effects of bullying on students (Singer 1996). We again validate the heterogeneous effect analysis by the Abadie–Imbens estimator, which is shown in Table A3 in the Appendix.

We have also checked the heterogeneous effects across other categories: student gender, parents' education, teachers' quality, the share of classmates that are also bullied,[4] as well as school characteristics. But we were unable to find any significant differential effects except for teacher gender. Those results can be found in Table A4 in the Appendix.

V. Robustness check

This section evaluates the robustness of our estimates from three aspects. First, it assesses the overlap assumption by a graphical representation. Second, we measure the quality of matching. Third, a placebo regression is employed to test the plausibility of the unconfoundedness assumption. Overlap (or common support) is one of the major assumptions in PSM, which ensures that students with the same propensity score have a positive probability of being both treated and untreated. A straightforward method to test the overlap assumption is to plot the distribution of the propensity scores of the bullied and non-bullied students, and visually inspect whether the two distributions are overlapped. Figure 2 depicts that the two distributions are considerably overlapped. In Figure 3, we validate the overlap condition by examining the distributions of the propensity scores of the bullied and non-bullied students by splitting the sample by teacher gender.

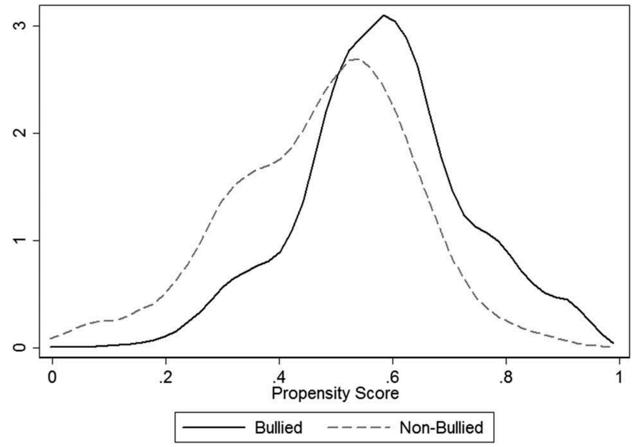

Figure 2. Propensity score distributions of students (full sample).

Another concern may arise over the quality of matching, for example, whether the distributions of the covariates in the control and treatment groups are balanced. If the matching is successfully balanced, then the differences between covariate means of the treatment and control groups should be significantly lower after matching. Following Diamond and Sekhon (2013), we measure the each covariate balance by the mean standardized difference. Smaller mean standardized differences indicate that the individual covariate is well balanced. Table 5 displays the standardized mean differences among covariates pre- and post-matching. With only a few exceptions, the mean differences after matching become significantly small and tend towards zero.

The other major assumption of PSM is unconfoundedness, that is, all covariates that relate with the treatment and potential outcome are included in our analysis. A placebo regression is designed to assess the unconfoundedness assumption. Maintaining all right-hand side variables used in the estimation of the propensity score, we insert a new dependent variable that is assumed to be exogenous with the treatment. If there are omitted variables correlated with the treatment, then the

[4] We acknowledge the suggestion from an anonymous referee. The impact of bullying may also be affected by the share of classmates also being bullied. We define large or small share by the median percentage (55%) of classmates being bullied. But we find insignificant differential impact of the share of classmates also being bullied.

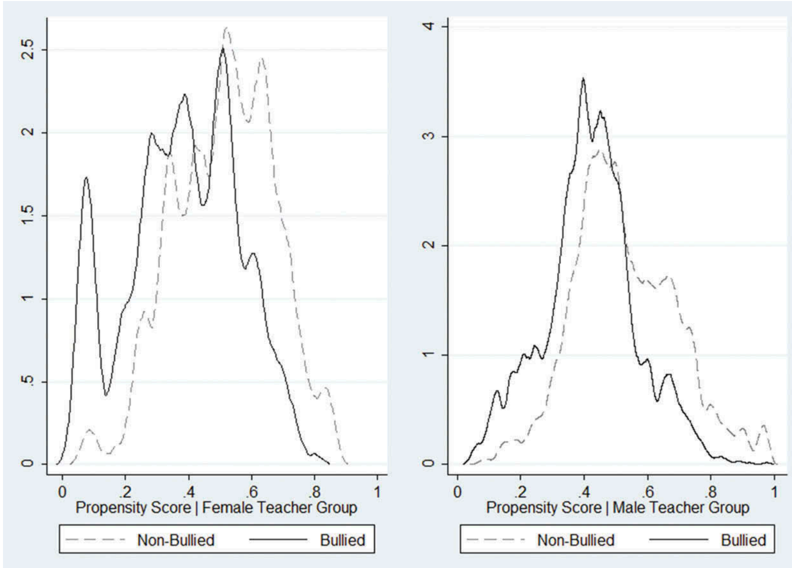

Figure 3. Propensity score distributions of students (sample separated by teacher gender).

Table 5. Matching quality: mean differences in covariates pre- and post-matching.

	Mean difference					
	Whole sample		Female teachers sample		Male teachers sample	
Variable	Before match	After match	Before match	After match	Before match	After match
Home computer	0.018	0.006	0.038	0.012	0.015	0.002
Study desk	0.028	0.001	0.016	0.037	0.030	0.004
Own room	0.028	0.008	0.026	0.024	0.029	0.004
Internet connection	0.017	0.003	0.007	0.003	0.019	0.002
Home books	0.030	0.003	0.063	0.015	0.025	0.001
Student gender	0.008	0.005	0.015	0.019	0.012	0.009
Student age	0.008	0.008	0.037	0.033	0.006	0.007
Parent education	0.017	0.006	0.032	0.008	0.017	0.004
School location	0.015	0.008	0.048	0.009	0.009	0.004
Portion of disadvantaged students	0.007	0.005	0.009	0.005	0.008	0.003
School enrolment	0.010	0.005	0.032	0.030	0.013	0.003
School computer availability	0.021	0.003	0.035	0.019	0.028	0.005
Teacher's experience	0.016	0.003	0.042	0.007	0.017	0.002
Teacher's education	0.014	0.002	0.036	0.014	0.016	0.003
Teacher's gender	0.006	0.002	0.000	0.000	0.000	0.000

coefficient associated with *bullied* should be significantly different from zero. Otherwise, the unconfoundness assumption is more credible. We employ the birth date of each student as the predetermined dependent variable, which is randomly assigned to the students. Table 6 shows the results of the placebo regression. The insignificant coefficient of *bullied* indicates that omitted variables affecting the treatment do not exist.

Table 6. Placebo regression test results.

	Full sample	Female teachers	Male teachers
Bullied	0.385	0.471	0.393
	(0.272)	(0.972)	(0.289)

All regressions include control variables and school fixed effect. Standard errors are adjusted for school-level clustering and heteroscedasticity.

VI. Concluding remarks and policy implications

Our study has made an attempt to fill two imminent gaps in the literature: first, estimating the quantitative effect of bullying on academic performance from a developing African country; and more importantly the possible mitigating mechanism for this persistent impediment through gender based teaching treatment. We refer Ponzo's (2013) framework to estimate our initial results and then improve them by using novel innovations such as the DR estimator, examining quality of matching and placebo regressions. We find that on average an eighth graders performance on a standardized mathematics exam decreases about 18.5 points due to bullying. The decrease in performance is more severe for

Ghana when compared to the revealed decreases in performance of children from developed nations, including Ponzo's estimation in context of Italian school. The performance of children from a developing society is more sensitive to the negative repercussions of bullying due to lack of better access to socio-economic facilities. Hence, it is necessary to build a zero tolerance policy towards school bullying in the educational settings, especially for the developing countries.

Our findings also highlight the importance of tailored anti-bullying programmes involving gender-specific components, which suggests promoting female teacher participation in the long run responding to the prevalence of bullying in school. Additionally, anti-bullying trainings are necessary for the existing male teachers. As it is suggested above, the males' teaching paradigm ('masculine' style) seems not as effective as the females' paradigm ('feminine' style) in addressing school bullying. Hence, male teachers are recommended to adapt themselves to a student-centred pattern and use more collaborative learning techniques rather than the traditional instructional behaviours. This study encourages further investigation of the channels through which teacher's gender affects the impact of school bullying.

Acknowledgements

The authors thank seminar participants at AAEA, USAID. Xu would like to thank for the financial support from the National Natural Science Foundation of China (71203145). All errors are our own responsibility.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China: Grant Number [71203145].

ORCID

Zhicheng Phil Xu http://orcid.org/0000-0003-2198-8612

References

Abadie, A., and G. W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76: 1537–1557. doi:10.3982/ECTA6474.

Abadie, A., and G. W. Imbens 2009. "Matching on the Estimated Propensity Score." National Bureau of Economic Research w15301.

Ammermueller, A. 2012. "Violence in European Schools: A Widespread Phenomenon that Matters for Educational Production." *Labour Economics* 19: 908–922. doi:10.1016/j.labeco.2012.08.010.

Brown, S., and K. Taylor. 2008. "Bullying, Education and Earnings: Evidence from the National Child Development Study." *Economics of Education Review* 27: 387–401. doi:10.1016/j.econedurev.2007.03.003.

Caliendo, M., and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22: 31–72. doi:10.1111/joes.2008.22.issue-1.

Casey, R. J., and L. L. Fuller. 1994. "Maternal Regulation of Children's Emotions." *Journal of Nonverbal Behavior* 18: 57–89. doi:10.1007/BF02169079.

Diamond, A., and J. S. Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 95: 932–945. doi:10.1162/REST_a_00318.

Dunne, M., R. Sabates, C. Bosumtwi-Sam, and A. Owusu. 2013. "Peer Relations, Violence and School Attendance: Analyses of Bullying in Senior High Schools in Ghana." *Journal of Development Studies* 49: 285–300. doi:10.1080/00220388.2012.671472.

Fryer, R. G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics* 31 (2): 373–407. doi:10.1086/667757.

Gray, H. L. 1987. "Gender Considerations in School Management: Masculine and Feminine Leadership Styles." *School Organisation* 7 (3): 297–302. doi:10.1080/0260136870070305.

Heckman, J. J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources* 32: 441–462. doi:10.2307/146178.

Heckman, J. J., H. Ichimura, and P. E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64: 605–654. doi:10.2307/2971733.

Imbens, G. W. 2014. "Matching Methods in Practice: Three Examples." National Bureau of Economic Research w19959.

Imbens, G. W., and J. M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47: 5–86. doi:10.1257/jel.47.1.5.

Le, A. T., P. W. Miller, A. C. Heath, and N. Martin. 2005. "Early Childhood Behaviours, Schooling and Labour Market Outcomes: Estimates from a Sample of Twins." *Economics of Education Review* 24: 1–17. doi:10.1016/j.econedurev.2004.04.004.

Martin, J., and H. Ross. 2005. "Sibling Aggression: Sex Differences and Parents' Reactions." *International Journal of Behavioral Development* 29: 129–138. doi:10.1080/01650250444000469.

Metzler, J., and L. Woessmann. 2012. "The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation." *Journal of Development Economics* 99 (2): 486–496. doi:10.1016/j.jdeveco.2012.06.002.

Mislevy, R. J. 1991. "Randomization-Based Inference about Latent Variables from Complex Samples." *Psychometrika* 56 (2): 177–196. doi:10.1007/BF02294457.

Mullis, I. V., M. O. Martin, P. Foy, and K. T. Drucker. 2012. *PIRLS 2011 International Results in Reading*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Nansel, T. R., M. Overpeck, R. S. Pilla, W. Ruan, B. Simons-Morton, and P. Scheidt. 2001. "Bullying Behaviors among Us Youth: Prevalence and Association with Psychosocial Adjustment." *Journal of the American Medical Association* 285: 2094–2100. doi:10.1001/jama.285.16.2094.

Olweus, D. 1993. *Bullying at School: What We Know and What We Can Do*. Malden, MA: Blackwell.

Ponzo, M. 2013. "Does Bullying Reduce Educational Achievement? an Evaluation Using Matching Estimators." *Journal of Policy Modeling* 35: 1057–1078. doi:10.1016/j.jpolmod.2013.06.002.

Ripski, M. B., and A. Gregory. 2009. "Unfair, Unsafe, and Unwelcome: Do High School Students' Perceptions of Unfairness, Hostility, and Victimization in School Predict Engagement and Achievement?" *Journal of School Violence* 8: 355–375. doi:10.1080/15388220903132755.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55. doi:10.1093/biomet/70.1.41.

Rubin, D. B. 1987. *Multiple Imputations for Nonresponse in Surveys*. New York: John Wiley and Sons.

Saha, L. J. 1983. "Social Structure and Teacher Effects on Academic Achievement: A Comparative Analysis." *Comparative Education Review* 27 (1): 69–88. doi:10.1086/446346.

Singer, E. 1996. "Espoused Teaching Paradigms of College Faculty." *Research in Higher Education* 37 (6): 659–679. doi:10.1007/BF01792951.

Warwick, D. P., and H. Jatoi. 1994. "Teacher Gender and Student Achievement in Pakistan." *Comparative Education Review* 38 (3): 377–399. doi:10.1086/447257.

Appendix

Table A1. Sensitivity check – alternative measures of math performance.

	Impacts of being bullied weekly			
Outcome variable	OLS full sample	PSM full sample	PSM female teachers	PSM male teachers
Proficiency score 1	−19.890***	−19.245***	−6.107	−17.346***
	(2.844)	(3.801)	(10.184)	(4.796)
Proficiency score 2	−17.714***	−15.501***	−1.143	−19.102***
	(2.816)	(3.237)	(10.083)	(4.134)
Proficiency score 3	−18.808***	−14.611***	−3.237	−21.993***
	(2.915)	(3.962)	(11.233)	(4.002)
Proficiency score 4	−17.711***	−16.951***	2.638	−18.263***
	(3.162)	(3.642)	(9.879)	(4.013)
Proficiency score 5	−19.045***	−19.377***	−4.604	−23.351***
	(3.112)	(4.049)	(9.891)	(4.231)

The OLS regression has the same specification as in column 4 of Table 2, which controls for student, household, teacher, and school characteristics. Standard errors in OLS are adjusted for school-level clustering and heteroscedasticity . For PSM method, nearest neighbour algorithm is applied with replacement; Balancing property and common support are satisfied; Standard errors are estimated by 100 bootstrap replications. *** indicates 1% level statistical significance.

Table A2. Abadie–Imbens PSM results.

Panel A: full sample	Eighth grade math scores	
ATT	−21.470***	
	(3.358)	
Number of treated	2380	
Number of controls	2085	
Panel B: Teachers decomposed by gender	Female	Male
	−0.413	−21.710***
	(7.004)	(3.486)
Number of treated	290	2090
Number of controls	228	1857

Balancing property and common support are satisfied. Nearest neighbour is applied with replacement. Abadie–Imbens robust standard errors are reported in parentheses. *** indicates 1% level statistical significance.

Table A3. Sampling weighted linear regression and PSM.

Panel A: Sampling weighted linear regression

	(1)	(2)	(3)	(4)	(5)
Bullied	−19.410***	−21.649***	−20.828***	−20.845***	−15.787***
	(3.960)	(3.515)	(3.589)	(3.257)	(2.629)
Individual and household controls	No	Yes	Yes	Yes	Yes
Teacher level controls	No	No	Yes	Yes	No
School level controls	No	No	No	Yes	No
School fixed effects	No	No	No	No	Yes
Observations	7323	5503	5002	4514	5503
R-squared	0.013	0.129	0.150	0.222	0.463

Panel B: sampling weighted PSM (nearest neighbour)

	Full sample	Female teacher	Male teacher
Bullied	−16.405***	−2.491	−20.011***
	(4.535)	(10.435)	(4.914)
Number of treated	2357	276	2077
Number of controls	2081	221	1856

For WLS method: 'Individual and household controls' include: student age, gender, parents' highest education level, number of books at home, computer possession, study desk, own room, and Internet accessibility. 'Teacher level control' includes teacher's gender, experience, and education level. 'School level controls' include: school enrolment, location, portion of students coming from disadvantaged families, and instructional computer accessibility. Weighted standard errors are adjusted for school-level clustering. For PSM method: Nearest neighbour algorithm is applied with replacement; Balancing property and common support are satisfied; Standard errors are estimated by 100 bootstrap replications. *** indicates 1% level statistical significance.

Table A4. Heterogeneous impacts of being bullied across other categories.

A. Students decomposed by gender	Female	Male
	−22.314***	−12.973***
	(7.757)	(5.529)
Number of treated	1134	1215
Number of controls	974	1063
B. Students decomposed by school location	**Urban**	**Rural**
	−12.019**	−12.465**
	(6.138)	(5.943)
Number of treated	1238	1128
Number of controls	1051	1026
C. Students decomposed by parents' education	**Post-secondary**	**Secondary or lower**
	−18.71**	−21.668***
	(8.239)	(4.444)
Number of treated	633	2113
Number of controls	557	1811
D. Students decomposed by teachers' quality	**Low teaching quality**	**High teaching quality**
	−21.499***	−16.3753**
	(5.313)	(7.618)
Number of treated	1613	732
Number of controls	1479	601
E. Students decomposed by school facility	**With computer**	**Without computer**
	−16.992***	−17.254*
	(4.196)	(9.740)
Number of treated	2037	558
Number of controls	1725	553
F. Students decomposed by students' economic backgrounds	**Low portion of poor students in school**	**High portion of poor students in school**
	−19.201***	−16.261***
	(7.112)	(5.050)
Number of treated	857	1522
Number of controls	776	1367
G. Students decomposed by share of classmates also bullied	**Small share of classmates also bullied**	**Large share of classmates also bullied**
	−19.288***	−23.248***
	(5.797)	(5.704)
Number of treated	935	1428
Number of controls	1318	761

Balancing property and common support are satisfied. Nearest neighbour is applied with replacement. Standard errors, estimated by 100 bootstrap replications, are reported in parentheses. *, **, *** indicate 10, 5, and 1% level statistical significance, respectively.