Student Paper

Young Lives

An International Study of Childhood Poverty

The Impact of Teacher Gender on Learning Outcomes in Developing Countries: Evidence from Pakistan and Vietnam

Benjamin Lodewijks

Thesis submitted in part fulfilment of the requirements for the degree of MSc in Economics for Development, University of Oxford.

The data used come from Young Lives, a longitudinal study of childhood poverty that is tracking the lives of 12,000 children in Ethiopia, India (Andhra Pradesh), Peru and Vietnam over a 15-year period. **www.younglives.org.uk**

Young Lives is funded by UK aid from the Department for International Development (DFID) and co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014 and by Irish Aid from 2014 to 2015.

The impact of teacher gender on learning outcomes in developing countries: evidence from Pakistan and Vietnam

Thesis submitted in partial fulfilment of the requirements for the
Degree of Master of Science in Economics for Development
at the University of Oxford
01 June 2015

Benjamin Lodewijks

ABSTRACT

Reducing gender gaps in education remains a significant policy concern in many developing countries. One strategy that has been advocated by a range of international organisations to improve learning outcomes for girls is to increase the representation of female teachers. This has been supported by recent empirical evidence from India (Rawal and Kingdon 2010 and Muralidharan and Sheth 2015). This extended essay examines whether the finding that female teachers improve learning outcomes for girls holds in other cultural contexts, drawing on longitudinal data on primary school students from Vietnam and Pakistan.

Using dynamic OLS value-added models, I find that teacher gender has no significant impact on learning outcomes in Vietnam while in Pakistan male teachers significantly improve overall test scores in private schools for both boys and girls by around 0.13 standard deviations per year. This latter finding does not appear to be driven by positive sorting or differences in teacher characteristics or effort levels between male and female teachers. However, male teachers did not have a significant impact on overall test scores in public schools suggesting that the results may be driven by differences in unobservable characteristics between male and female teachers in private schools.

Finally, in both public and private schools in Pakistan male teachers had a stronger impact on test scores in Maths than English and Urdu, although no such effects were found for Vietnam. This finding suggests that so-called 'stereotype effects' may play an important role in teacher-student gender interactions in some developing countries.

ACKNOWLEDGEMENTS

I would like to acknowledge the support of a range of people who have helped me through the process of writing this extended essay. First, I would like to thank Christine Ernst for encouraging me to come to Oxford and for her unending support and willingness to discuss ideas throughout the year. Secondly, I would like to thank my supervisor Dr. Abhijeet Singh for suggesting this topic to me and for providing many valuable insights and suggestions along the way. Finally, I would like to thank Dr. Clement Imbert for his advice on the empirical approach.

The data used in this extended essay comes from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Department for International Development (DFID), with co-funding from 2010 to 2014 by the Netherlands Ministry of Foreign Affairs, and from 2014 to 2015 by Irish Aid. The views expressed here are those of the author. They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders.

The other data source used for this project is the Learning and Educational Achievement in Pakistan Schools (LEAPS) project, a joint project between the World Bank, Harvard University and Pomona College.

Contents

1 Introduction 5
2 Data 8
2.1 LEAPS 8
2.2 Young Lives 9
2.3 Student characteristics 10
2.4 Teacher characteristics 16
3 Empirical strategy 17
3.1 Dynamic OLS value-added models 18
3.2 Estimating the treatment effect of being matched with a female teacher 19
3.3 Potential threats to identification 20
4 Results 22
4.1 Are the results for Pakistan driven by positive sorting? 25
5 Identifying the mechanisms 29
5.1 Can the results be explained by teacher characteristics or effort? 29
5.2 Can the results for Pakistan be partly explained by stereotype effects? 30
6 Conclusion 34

1 Introduction

Significant progress has been made in reducing global gender gaps in education over the last two decades. However, gender gaps in enrolment and learning outcomes still exist in many developing countries. A total of 28 countries remain 'seriously off target' for achieving the Millennium Development Goal of ending gender disparities in primary and secondary enrolment by 2015 and are not expected to eliminate gender disparities in enrolment until at least 2030 (World Bank 2015a).

A range of international organisations including UNESCO (2005) and the United Nations Taskforce for achieving the Millennium Development Goals (2010) have advocated increasing the representation of female teachers as a way of reducing educational gender gaps in developing countries.

Until recently, most empirical evidence on the impact of teacher gender on learning outcomes was based on the US (see Dee 2004, 2005, 2007 and more recently Winters et al 2013 and Antecol et al 2015). However, a number of recent empirical studies in developing countries have supported the view that that female teachers improve learning outcomes for girls. Muralidharan and Sheth (2015) use longitudinal data from India and find that female teachers significantly improve learning outcomes for girls in primary school, without adversely impacting outcomes for boys. This work built upon an earlier study in India by Rawal and Kingdon (2010) which found that both girls and boys benefitted from being matched with a teacher of the same gender. Paredes (2014) also finds that female teachers improve learning outcomes for girls in Chile, although not a developing country.

Finally, Keucken and Valfort (2012) examine data for eleven Sub Saharan African countries and find that female teachers significantly improve reading scores but significantly reduce Maths scores for both boys and girls. However, the positive impact of female teachers on reading scores for girls was larger than their adverse impact on Maths scores, so the results are broadly consistent with other studies.

On the whole, these findings suggest that increasing the representation of female teachers is likely to be a useful policy for reducing gender gaps in education in developing countries. Yet depending on the mechanisms behind this effect it is possible that the impact of female teachers might vary across cultural contexts.

The main contribution of this extended essay is to examine whether the finding that female teachers improve learning for girls holds in different cultural contexts, drawing on longitudinal data from the Learning and Educational Achievement in Pakistan Schools (LEAPS) project and the Young Lives School Survey in Vietnam.

Despite having similar levels of GDP, gender gaps in education and labour force participation are far more pronounced in Pakistan than Vietnam. In 2013, the ratio of girls to boys enrolled in primary and secondary schools in Pakistan was 87% and 73% respectively, while the female labour force participation rate was just 25%. By comparison, enrolments of girls and boys in Vietnam have been broadly similar across all levels of education since 2005 and the female labour force participation rate is currently 73% (World Bank 2015b).

The choice of these two countries extends the existing literature in a number of ways. In the case of Pakistan, I am able to examine a country in which gender gaps in enrolment remain significant. An earlier cross-sectional study of Maths scores in public schools in Pakistan by Warwick and Jatoi (1994) found that male teachers improved Maths scores, although this effect was reversed for teachers with university degrees and was not present in urban areas. The analysis here re-examines this finding using longitudinal data for a range of subjects. I also focus on private schools where identification of the impact of female teachers is arguably clearer since most public schools in Pakistan are gender segregated. In the case of Vietnam, this is to the author's knowledge the first study to examine the impact of teacher gender in a South-East Asian context, where gender gaps in enrolment are generally less pronounced. Finally, the rich set of controls available in both LEAPS and Young Lives allows for the inclusion of a wider range of child characteristics than most previous studies of teacher gender in developing countries.

To examine the impact of female teachers in Vietnam and Pakistan, I adopt a similar methodology to Muralidharan and Sheth (2015) and Antecol et al (2015) and estimate dynamic OLS value-added models. Muralidharan and Sheth (2015) note that this is likely to provide a clearer identification of the impact of teacher gender than the approach of using students fixed effects and variation in teacher gender across subjects which has been used in much of the previous literature (see Dee 2007, Rawal and Kingdon 2010 and Keucken and Valfort 2012) since it is difficult to interpret the impact of higher test scores in a subject without knowing the gender of a student's teacher in the previous year.

In contrast to the previous literature in developing countries, I find no evidence that female teachers significantly improve overall test scores. In Vietnam, teacher gender has no significant impact on test scores for either boys or girls. In Pakistan, I find that male teachers significantly improve overall test scores in private schools by around 0.13 standard deviations a year for both boys and girls. The positive impact of male teachers in private schools in Pakistan does not appear to be driven by positive sorting or differences in teacher characteristics or effort. However, when the analysis was extended to public schools male teachers did not have a significant impact on overall test scores, suggesting that these results may reflect differences in unobservable characteristics between male and female teachers in private schools.

Finally, in both public and private schools in Pakistan the impact of male teachers was stronger in Maths than English and Urdu, which is consistent with the notion of 'stereotype effects' under which students respond to the stereotype that males are better at Maths and females at reading. While there was no evidence of stereotype effects in Vietnam, the results for Pakistan are consistent with those of Keucken and Valfort (2012) for Sub Saharan Africa and suggest that stereotype effects may play a role in teacher-student gender interactions in certain cultural contexts.

The remainder of this essay is structured as follows: Chapter 2 discusses the data and descriptive statistics; Chapter 3 outlines the empirical strategy; Chapter 4 discusses the results; Chapter 5 discusses potential mechanisms; and Chapter 6 concludes.

2 Data

2.1 LEAPS

The LEAPS project collected data on schools in 112 villages in the Punjab province in Pakistan between 2003 and 2006. The purpose of the LEAPS project was to investigate the rise of private schooling so villages were randomly chosen from the list of all villages in the districts of Attock, Faisalabad and Rahim Yar Khan with at least one private school. The sampled villages are generally larger, wealthier and more educated than average rural villages in the Punjab province, although over 40% of the province's rural population live in villages containing a private school (Andrabi et al 2006).

In Pakistan, public schools tend to be gender segregated both in terms of pupils and teachers, especially in larger villages. This makes it difficult to identify the impact of female teachers based on variation in teacher gender over time. By comparison, private schools are usually not gender segregated and many have both male and female teachers. Since the impact of teacher gender in private schools can be identified both within schools and between boys and girls within the same class, the analysis here focuses on private schools. However, the results are extended to public schools in chapter 5 (using the small proportion of students taught by teachers of a different gender) to examine whether the pattern of results found for private schools also holds in public schools.

The first panel of LEAPS consisted of 13,735 third grade students of whom 12,110 were administered exams in English, Maths and Urdu (the official language of Pakistan). These students were tracked and tested along with their peers for two subsequent years, regardless of whether they were promoted. In total, 12,815 students were tested in the second round and 12,123 students in the third round with 8,120 students being tested in all three years. Andrabi et al (2011) note that the high rate of attrition in LEAPS was due to 8.7% of children dropping out of school and absence rates of around 10% on test days. Of those tested, up to 10 students per school (a total of 6,379 in the first round) were administered a child survey each round capturing demographic information.

To examine whether attrition was related to a student's teacher gender match, I created a dummy variable for attrition if no test score was available for a private school student in the following year and regressed it on teacher gender, student gender and their interaction. The coefficients on teacher gender and its interaction were insignificant suggesting that attrition was not related to teacher gender.

In addition to focusing on private school students, I exclude observations where students change schools across consecutive years since changes in test scores for these students may be partly driven by changes in school resources or family circumstances. Teachers in LEAPS were also asked whether or not they shared teaching responsibility for English with another teacher and who this teacher was. To ensure this did not affect the results, I excluded test scores in English for the 155 students taught by both male and female teachers. The resulting sample contained 5,065 student-year observations with test scores for the current and previous year.

2.2 Young Lives

The Young Lives Vietnam School Survey was administered to children studying in grade 5 in the 2011-12 school year. The survey covered 20 communities, chosen to ensure adequate diversity of geographic regions and demographic characteristics. Up to 20 children in each grade 5 class were administered tests in Maths and Vietnamese at the beginning and end of the school year. In total 3,284 children were included in the Young Lives School Survey, with test scores in both rounds being available for 3,187 students in Maths and 3,196 in Vietnamese.

Of these students, I exclude all students who experienced a change in teacher during the school year as no information was available on the characteristics of their subsequent teacher or when the change occurred. Similarly to LEAPS, teachers in Young Lives were asked whether they shared teaching responsibilities in a subject with another teacher. However, no information was provided on who the other teacher was. To address the possibility that teaching might be shared by teachers of a different gender, I exclude test scores for students whose teacher: shared teaching in a subject with another teacher; and who taught in schools with a mix of male and female grade 5 teachers. This led to test scores being excluded for four classes in Maths and two in Vietnamese. The resulting sample contained tests scores in both rounds for 2,913 students in Maths and 2,994 students in Vietnamese.

Test scores were calculated based on student exam responses for both surveys using Item Response Theory (IRT) models. IRT models account for the relative difficulty of each item on a test by assuming a mathematical relationship between the latent ability of an individual and the probability that they will answer a question correctly, which differs depending on the difficulty of each question. Since LEAPS and Young Lives include a subset of questions which are repeated each year, this allows for a valid comparison of student learning growth over time.

As both LEAPS and Young Lives involve multiple choice questions, I use a three parameter logistic (3PL) IRT model and calculate scores based on maximum likelihood with the *openirt* command in Stata written by Tristan Zajonc. A detailed discussion of 3PL IRT models is provided by Das and Zajonc (2010). IRT scores were normalised to have a mean of 0 and a standard deviation of 1 at baseline for Young Lives and in the first round (grade 3) for LEAPs.[1]

2.3 Student characteristics

This section provides descriptive statistics on the children in Young Lives and LEAPS to provide a picture of their family background and to assess whether there are any significant differences between children allocated to male and female teachers.

[1] In LEAPS, IRT scores were normalised for private and public school students separately.

Table 1 summarises the baseline characteristics of private school children in LEAPs. The sample is based on an unbalanced panel of students for whom test scores are available in the current and previous year. As a consequence, baseline test scores and student characteristics are not available for students who only appear in rounds 2 and 3. Demographic characteristics other than age and gender are also not available for children who were not selected for the child survey.

Mother's and father's education in LEAPs are given a value of 1 if the parent did not attend school, 2 if they did not complete primary school, 3 if they completed primary or some secondary schooling and 4 if they studied at the tertiary level. I calculated a household asset index for LEAPS and Young Lives using the first principal component from principal component analysis of a range of household assets. The asset index was standardised to have a mean of 0 and a standard deviation of 1.

There is no explicit question in LEAPS on whether a private school is a single-sex school, so the variable 'Single sex school' in Table 1 was constructed to capture students attending schools with no gender variation in the student body. Only 3% of private school students attend such schools, some of which may not actually be single sex schools given that some private schools have quite small enrolments.

Panel A in Table 1 indicates that there are some significant differences between children allocated to male and female teachers. While girls constitute 47% of students taught by female teachers they constitute only a third of students taught by male teachers. This suggests that parents may choose schools to achieve a teacher-student gender match or that schools may choose teacher gender to match their student population.

There were significant differences in baseline test scores between children allocated to male and female teachers.[2] Children allocated to male teachers had higher baseline test scores in all subjects (although this difference was only significant at the 10% level for Maths) and significantly higher average teacher

[2] Baseline test scores were marginally higher for those who were present across consecutive years so the mean in Table 1 is higher than for the whole sample at baseline.

ratings, implying that better performing students were more likely to be allocated to male teachers. Panels B and C, which provide results for boys and girls separately, indicate that these differences applied to both boys and girls taught by male teachers.

There was not a clear relationship between parental education and teacher gender in private schools. Boys allocated to male teachers had slightly less educated mothers whereas girls allocated to male teachers had more educated fathers. Boys taught by male teachers had higher standardised height and weight scores but lived further from school and had lower average household assets. These differences were not apparent for girls, although girls allocated to male teachers were more likely to live with their parents.

Table 2 replicates Table 1 for Vietnam. In the Young Lives School Survey parental education is recorded as: 0 if they never attended school; 1 if they attended primary school; 2 if they attended lower secondary school (grades 6 to 9); 3 if they attended upper secondary school (grades 10 to 12); and 4 if they studied at the tertiary level.

In Vietnam, there were no significant differences in baseline test scores between students allocated to male or female teachers. However, students allocated to female teachers had more educated parents and higher average household assets. Breaking down the sample into urban and rural areas, students allocated to female teachers had more educated parents and household assets in both regions, although differences in parental education were not significant for urban areas.

To test whether these differences were driven by female teachers being more likely to teach in schools in wealthy areas, I separately regressed parental education, literacy levels and household assets on teacher gender controlling for school fixed effects. There was no significant difference in household assets, father's education or father's literacy for those allocated to female teachers within a school. However, those allocated to female teachers had mothers who were significantly more educated (by 0.23 units on average) and 2.4% more likely to be literate.

While this suggests that some form of sorting may be occurring within schools, since there were no significant differences in baseline test scores by teacher gender there was no evidence of sorting on the basis of student ability in Vietnam.

Table 1: Descriptive statistics for children attending private schools by teacher gender in Pakistan

	PANEL A: ALL STUDENTS				PANEL B: BOYS				PANEL C: GIRLS			
	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male - Female	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male – Female	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male – Female
Girl	5065	0.33	0.47	-0.15***								
Age at baseline	5065	9.23	9.29	-0.06	2836	9.29	9.32	-0.02	2229	9.08	9.25	-0.17**
Single sex school	5065	0.02	0.04	-0.02***	2836	0.02	0.03	-0.01	2229	0.02	0.05	-0.03***
Mother's education	4242	1.85	1.91	-0.06*	2362	1.82	1.90	-0.07*	1880	1.91	1.93	-0.03
Father's education	4242	2.60	2.55	0.05	2362	2.52	2.52	0.00	1880	2.74	2.58	0.16***
Elder brothers	4242	1.04	0.95	0.09*	2362	1.07	0.88	0.18***	1880	0.99	1.03	-0.04
Elder sisters	4242	1.00	1.02	-0.02	2362	1.04	0.99	0.05	1880	0.92	1.05	-0.13
Baseline information												
Maths score	4457	0.12	0.05	0.06*	2483	0.10	0.04	0.07	1974	0.14	0.07	0.07
Urdu score	4456	0.23	0.05	0.18***	2483	0.19	-0.04	0.23***	1973	0.32	0.15	0.18***
English score	4458	0.21	0.04	0.17***	2485	0.17	-0.01	0.18***	1973	0.31	0.10	0.21***
Years of schooling	2739	2.93	3.14	-0.21***	1498	2.83	3.16	-0.33***	1241	3.11	3.12	-0.02
Teacher rating of child (1-10)	2737	6.73	6.39	0.34***	1494	6.52	6.17	0.35**	1243	7.10	6.63	0.47***
Mother at home	2745	0.99	0.98	0.01*	1500	0.98	0.97	0.01	1245	1.00	0.98	0.02***
Father at home	2745	0.89	0.86	0.02	1500	0.86	0.87	0.00	1245	0.93	0.86	0.07***
Standardised height score relative to US	2739	-0.62	-0.94	0.32***	1495	-0.49	-0.84	0.36***	1244	-0.86	-1.05	0.19
Standardised weight score relative to US	2744	-0.76	-1.21	0.44*	1500	-0.41	-0.95	0.54	1244	-1.40	-1.48	0.09
Travels 30 mins or more to school	2745	0.05	0.04	0.01	1500	0.06	0.04	0.02*	1245	0.05	0.05	0.00
Asset index	2743	0.01	0.05	-0.05	1498	-0.04	0.07	-0.11*	1245	0.08	0.03	0.05
Observations	5065	1124	3941		2836	758	2078		2229	366	1863	

Note: (a) The sample consists of all private school students who do not change school across consecutive years and have test score information for at least one subject in consecutive years. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively based on a t test of differences in means by teacher gender.

Table 2: Descriptive statistics for children by teacher gender in Vietnam

	Panel A: All Students				Panel B: Boys				Panel C: Girls			
	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male - Female	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male – Female	Obs.	Male teacher (Mean)	Female teacher (Mean)	Male – Female
Girl	2999	0.47	0.48	-0.01								
Age at baseline	2970	10.46	10.44	0.02	1541	10.46	10.45	0.01	1429	10.45	10.42	0.03
Mother's education	2150	1.70	2.15	-0.45***	1114	1.75	2.15	-0.40***	1036	1.65	2.15	-0.50***
Father's education	2011	1.95	2.33	-0.38***	1063	1.90	2.32	-0.43***	948	2.02	2.34	-0.32***
Mother literate in Vietnamese	2987	0.88	0.93	-0.05***	1557	0.90	0.93	-0.03*	1430	0.87	0.93	-0.07***
Father literate in Vietnamese	2955	0.94	0.95	-0.01	1532	0.93	0.94	-0.02	1423	0.95	0.96	0.00
Number of older siblings	2977	1.06	0.93	0.12**	1551	1.17	0.93	0.24***	1426	0.92	0.93	-0.01
Baseline Maths score	2913	0.00	0.00	0.00	1513	-0.02	-0.03	0.01	1400	0.02	0.03	-0.01
Baseline Vietnamese score	2994	-0.04	0.02	-0.06	1558	-0.15	-0.12	-0.04	1436	0.09	0.16	-0.07
Teacher rating of child (1-5)	2999	3.53	3.50	0.03	1560	3.36	3.35	0.01	1439	3.72	3.66	0.05
Gets 3 meals a day	2993	0.84	0.83	0.01	1556	0.84	0.83	0.01	1437	0.83	0.83	0.00
No health problems	2999	0.71	0.71	0.00	1560	0.72	0.71	0.01	1439	0.70	0.70	0.00
Travels 30 mins or more to school	2979	0.08	0.07	0.01	1547	0.05	0.06	-0.01	1432	0.10	0.07	0.03*
Asset index	2980	-0.27	0.07	-0.34***	1548	-0.28	0.09	-0.36***	1432	-0.26	0.05	-0.32***
Observations	2999	745	2254		1560	393	1167		1439	352	1087	

Note: (a) The sample consists of all students who have information on test scores at the beginning and end of the year for at least one subject and do not change teacher during the year. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively based on a t test of differences in means by teacher gender.

2.4 Teacher characteristics

Table 3 provides some descriptive statistics on male and female teachers in Pakistan and Vietnam. The characteristics of male and female teachers are summarised in Panels A and B respectively, while the final column shows the difference in means between male and female teachers.

The variable teaching college qualification takes a value of 1 if the teacher has completed a post-school course or certification in teaching but not a university degree. In LEAPS, information was available on whether a teacher had a bachelors or masters degree. In Young Lives teachers were only asked whether they had a university degree.

In Pakistan, male teachers in private schools were on average 6.7 years older, 14% more likely to have a bachelors degree and 5% more likely to have a masters degree. They also earned almost twice the salary of female teachers, were more experienced and were less likely to come from the village in which they teach.

Teacher test scores in LEAPS were assessed using the same test given to children and were normalised to have a mean of 0 and standard deviation of 1 across all students in rounds 2 and 3. Average test scores in Table 3 are slightly lower, indicating that teachers with higher test scores had larger average class sizes. Male teachers performed 0.33 standard deviations better in English and 0.42 standard deviations better in Maths than female teachers on average. They also had test scores that were 0.15 standard deviations higher in Urdu, although this difference was not statistically significant.

In Vietnam, male teachers were 4.3 years older on average but generally less educated than female teachers. The proportion of female teachers with a university degree rather than teaching college qualification was 21% higher than for male teachers. Almost all teachers in Vietnam were trained in teaching and held permanent positions.

Table 3: Characteristics of teachers

	Panel A: Male Teachers			Panel B: Female Teachers			*Male - Female*
	Obs.	**Mean**	**St. dev.**	**Obs.**	**Mean**	**St. dev.**	**Mean**
Pakistan							
Age	157	30.23	4.75	505	23.50	11.55	6.73***
Experience^	156	2.38	0.80	505	2.13	0.76	0.25***
Teaching college qualification	157	0.14	0.35	505	0.15	0.35	-0.01
Bachelors degree	157	0.34	0.40	505	0.20	0.47	0.14***
Masters degree	157	0.08	0.15	505	0.02	0.27	0.05**
Trained in teaching	157	0.30	0.44	505	0.25	0.46	0.05
Monthly salary (000 rupees)	157	2.06	0.79	505	1.15	1.55	0.91***
Native to village	157	0.38	0.50	504	0.56	0.49	-0.18***
Permanent contract	157	0.03	0.18	506	0.03	0.18	-0.00
Maths score	129	0.19	1.07	398	-0.22	1.08	0.42***
English score	129	0.05	1.33	399	-0.28	1.24	0.33**
Urdu score	129	0.00	1.04	399	-0.15	1.06	0.15
Vietnam							
Age	45	42.24	7.78	123	37.92	7.11	4.33***
Experience^	45	2.96	0.21	123	2.93	0.29	0.03
Teaching college qualification	45	0.67	0.48	123	0.46	0.50	0.21**
University degree	45	0.29	0.46	123	0.50	0.50	-0.21**
Trained in teaching	45	1.00	0.00	122	0.99	0.01	0.01
Native to province	45	0.76	0.43	123	0.67	0.47	0.08
Permanent contract	45	1.00	0.00	123	0.98	0.13	0.02

Note: (a) The sample consists of all private school teachers teaching in rounds 2 and 3 in LEAPS and all teachers teaching grade 5 in Young Lives. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively based on a t test of differences in means by teacher gender. (c) ^Experience is expressed on a 1-3 scale: 1 indicates less than a year of experience; 2 indicates between 1-3 years of experience; and 3 indicates at least three years of previous experience.

3 Empirical strategy

This section sets out the empirical strategy used to identify the causal impact of teacher gender on student learning. In particular, the approach seeks to identify the treatment effect of being matched with a female teacher and the extent to which this treatment effect might be heterogeneous for boys and girls.

3.1 Dynamic OLS value-added models

To identify the treatment effect of being matched with a female teacher, I use a dynamic OLS value-added model. Andrabi et al (2011) discuss how dynamic OLS value-added models can be derived from an educational production function in which actual student achievement (y^*) is assumed to be a linear function of all past and present inputs (x) and cumulative productivity shocks (μ):

$$y_{it}^* = \alpha_1' x_{it} + \alpha_2' x_{i,t-1} + \cdots \alpha_t' x_{i1} + \sum_{s=1}^{s=t} \theta_{t+1-s} \mu_{is} \qquad (1)$$

Since information on all past and present inputs is not available, Andrabi et al (2011) note that by adding and subtracting $\beta y_{i,t-1}^*$ from the right hand side, assuming all coefficients decline geometrically and normalising θ_1 to 1, we get the dynamic OLS value-added model:

$$y_{it}^* = \alpha_1' x_{it} + \beta y_{i,t-1}^* + \mu_{it} \qquad (2)$$

In this model, $\beta y_{i,t-1}^*$ controls for the impact of all past inputs, student endowments (such as ability) and productivity shocks. Singh (2015) notes that there are a number of sources of potential bias in dynamic OLS value-added models. First, measurement error in test scores can attenuate the coefficient on lagged test scores and potentially also bias the input parameters. Secondly, if student endowments such as ability lead some students to learn more quickly every year, the coefficient on lagged test scores will be biased upwards given the positive correlation between lagged scores and ability.

Nonetheless, a number recent studies comparing results from dynamic OLS value-added models to quasi-experimental evidence or more data intensive methods have found that dynamic OLS models perform relatively well in estimating policy effects. In the context of LEAPS, Andrabi et al (2011) find that because measurement error and unobserved individual heterogeneity tend to offset each other, estimates of the impact of private schools from dynamic OLS value-added models were relatively similar to those from more data intensive dynamic panel models.

While Rothstein (2010) highlights the potential for bias in dynamic OLS value-added models due to student sorting on lagged test scores, Chetty et al (2014) find that dynamic OLS value-added models have minimal (and statistically insignificant) forecast bias relative to both richer models that control for previously unobserved parental characteristics and quasi-experimental results based on teacher transfers. Value-added models have also been found to have minimal bias in experimental studies (Kane and Staiger 2008, Kane et al 2013), studies of contract teachers in India (Muralidharan and Sundararaman 2013) and school choice lotteries (Deeming et al 2014).

Finally, using simulated data Guarino et al (2015) show that dynamic OLS models outperform a range of other estimators (including Arellano-Bond dynamic panel models) in estimating teacher effects under a variety of student and teacher assignment scenarios.

3.2 Estimating the treatment effect of being matched with a female teacher

The main specification used to estimate the treatment effect of being matched with a female teacher in each subject is given below:

$$y_{it} = \gamma + \delta y_{it-1} + \beta_1 FT_{it} + \beta_2 g_i + \beta_3 FT_{it} * g_i + \alpha' x_{it} + \varphi t + \mu_{it} \qquad (3)$$

This is a dynamic OLS value-added model in which y_{it} is the test score for individual i at time t; FT is a dummy variable for being matched with a female teacher; g is a dummy variable for whether the student is female; x is a vector of inputs including child and household characteristics; t is a year dummy and μ is a random error term.

In equation (3), β_1 captures the treatment effect of being matched with a female teacher and β_3 captures any heterogeneous treatment effect on girls relative to boys from being matched with a female rather than a male teacher. This is the same framework used by Muralidharan and Sheth (2015) and Antecol et al (2015). As set out in Muralidharan and Sheth (2015), the treatment effect of female teachers on girls is equal to $\beta_1 + \beta_3$, while the overall treatment effect on all students is equal to $p_g(\beta_1 + B_3) + (1 - p_g)\beta_1 = \beta_1 + p_g\beta_3$ where ρ_g is the proportion of female students in the sample.

3.3 Potential threats to identification

This section discusses three main threats to identifying the treatment effect of female teachers on test scores and how these are addressed in the empirical strategy.

The first main threat to identifying the treatment effect of female teachers is the potential for *non-random of allocation of students and teachers to schools.* For example, female teachers may be more likely to work in wealthier villages where students have stronger learning trajectories. The inclusion of lagged test scores here substantially mitigates potential bias from student sorting. Chetty et al (2014) find that including lagged test scores alone results in minimal forecast bias in the US. The inclusion of number of child and household characteristics also helps reduce potential sorting on observables.

I further control for potential non-random allocation of students and teachers to schools by adding school fixed effects to equation (3). In this case, the treatment effect of being matched with a female teacher is identified based on the relative effectiveness of female and male teachers within the same school, with β_1 being only estimated for students in schools with both male and female teachers. In Young Lives, the identifying variation comes from the third of schools with both male and female grade 5 teachers. In LEAPS, 17% of the sample are in schools with variation in teacher gender. Since almost all private schools in LEAPS have one class per grade this variation occurs largely through changes in teacher gender across rounds.

In addition to controlling for school fixed effects, I also estimate models with classroom fixed effects. This controls for differences in teacher quality, allowing for a direct comparison of differences in treatment effects on boys and girls in the same class. This specification is shown in equation (4) below where φ_c represents the classroom fixed effect and teacher gender drops out because it is collinear with the fixed effect:

$$y_{it} = \gamma + \delta y_{it-1} + \beta_2 g_i + \beta_3 FT_{it} * g_i + \varphi_c + \alpha' x_{it} + \mu_{it} \qquad (4)$$

While in Vietnam there was no evidence of sorting to male or female teachers on baseline test scores, in Pakistan both boys and girls allocated to male teachers had significantly higher baseline scores in English and Urdu. Positive sorting to schools with male teachers should not substantially impact the estimate of β_1 since it is only identified from students in schools with both male and female teachers. However, positive sorting could impact β_3 if high performing girls benefit more from matches with female teachers. To test this, I examine how β_3 differs when estimated on a subsample of students taught by both male and female teachers in rounds 2 and 3.

Since there are two rounds of data in LEAPS after controlling for lagged test scores, it is also possible to use individual fixed effects to control for positive sorting. This is equivalent here to first difference OLS which takes the following form:

$$y_{it} - y_{it-1} = \delta(y_{it-1} - y_{it-2}) + \beta_1(FT_{it} - FT_{it-1}) + \beta_2(FT_{it} - FT_{it-1}) * g_i + \alpha'(x_{it} - x_{i,t-1}) + (\mu_{it} - \mu_{it-1}) \qquad (5)$$

However, this model cannot be estimated with OLS due to the correlation between y_{it-1} and $u_{i,t-1}$. Andrabi et al (2011) address this issue (and potential measurement error in $y_{i,t-2}$) by instrumenting for lagged test score gains using twice lagged test scores in other subjects. They report Arellano-Bond (1991) difference GMM estimates assuming both strictly exogenous and predetermined inputs (although the latter was unable to identify the private school effect in Math) and system GMM estimates. As a robustness check, I also estimate the results here using system GMM.

The second main threat to identification is the potential for *non-random allocation of students to teachers within schools*. This is not a major issue in LEAPS as all private schools had only one class at baseline and the existence of multiple classes in later rounds was generally the result of non-promotion. However, it does potentially affect Vietnam where schools had on average two Grade 5 classes.

To control for non-random allocation within schools, I use teacher responses on how students were allocated to classrooms to exclude classes allocated on the basis of ability (covering 7.5% of all students). I only exclude these students from models which control for school fixed effects since specifications with classroom fixed effects implicitly control for difference in average ability across classes.

The third main threat to identification is the potential for female teachers to be allocated to grades where girls and boys learn at different rates. Muralidharan and Sheth (2015) find that girls in India have stronger learning trajectories in lower grades where they are more likely to be taught by female teachers.

To test for this in LEAPS, I follow Muralidharan and Sheth (2015) and regress scores in each subject on lagged scores, a student gender dummy and an interaction term between student gender and year to compare learning trajectories for boys and girls in round 2 (grade 4) to round 3 (grade 5).[3] While students had a significantly higher trajectory in round 3, a Wald test indicated there were no significant differences in learning trajectories between boys and girls in any subject.

4 Results

The first three columns of Table 4 examine the impact of teacher gender on scores in all subjects, using subject dummy variables to control for differences in learning trajectories across subjects. Column 1 provides results from estimating equation (3) using pooled OLS. Column 2 adds school fixed effects. Column 3 includes classroom fixed effects rather than school fixed effects to examine the relative impact of female teachers on girls and boys in the same classroom.

[3] Results available from the author on request.

All specifications control for child characteristics including age, mother's and father's education, whether a child's mother and father live at home, number of elder brothers and sisters, years of schooling, standardised height and weight scores, household assets and whether the child travels more than 30 minutes to school.

The pooled OLS results in column 1 indicate that female teachers reduce test scores for boys (β_1) by 0.14 standard deviations a year on average, which was significant at the 1% level. The interaction term Female teacher*Girl was not statistically significant, providing no evidence to reject the hypothesis that male and female teachers have a similar impact on girls' test scores relative to boys'.

The overall impact of female teachers on girls test scores is represented by $(\beta_1 + \beta_3)$, with female teachers reducing girls' test scores by 0.13 standard deviations per year, which was significant at the 5% level. The impact of female teachers on all students is indicated by $(\beta_1 + p_g\beta_3)$.

After controlling for school fixed effects in column 2, the positive impact of male teachers on boys fell to 0.12 standard deviations but remained significant at the 5% level. The size of the impact on girls remained similar to that under pooled OLS but was now only significant at the 10% level. In column 3, the coefficient on Female teacher*Girl remained insignificant after controlling for classroom fixed effects.

It is possible that the results of columns 1 to 3 may mask significant differences in the impact of female teachers across subjects. To explore this, the specifications for all subjects are replicated for Maths in columns 4 to 6, for English in columns 7 to 9 and for Urdu in columns 10 to 12.

In Maths, male teachers were found to again have a positive impact on boys, raising boys' Maths scores by 0.18 standard deviations a year after including school fixed effects. This effect was significant at the 5% level. While male teachers also increased test scores for girls by 0.14 standard deviations, this effect was no longer significant after including school fixed effects.

In English, male teachers were found to increase test scores for boys under pooled OLS. The estimated impact of male teachers on boys remained similar after including school fixed effects, although was no longer statistically significant.[4] In Urdu, male teachers had no statistically significant impact on test scores for boys but did appear to improve test scores for girls. After including school fixed effects, male teachers were found to increase test scores for girls by 0.15 standard deviations, although this effect was only significant at the 10% level.

Table 5 shows the estimated impact of teacher gender in Vietnam. As for Pakistan, column 1 shows the results from estimating equation (3) using OLS, while column 2 adds school fixed effects to the model. Column 2 also controls for possible non-random allocation of students to classes within schools by excluding classes allocated on the basis of ability. Column 3 includes classroom fixed effects.

All models include controls for number of child characteristics including age, whether the child's mother and father can read and write Vietnamese, number of older siblings, household assets and whether the child receives at least three meals a day, has any health problems, or has to travel more than 30 minutes to school.[5]

As for Pakistan, the results for all subjects in columns 1 to 3 were replicated for Maths in columns 4 to 6 and for Vietnamese in columns 7 to 9. Across all specifications in Table 5, teacher gender had no statistically significant impact on test scores for either boys or girls. These results were also insignificant when clustered at the classroom level. The interaction term Female teacher* Girl was also insignificant, so I am also unable to reject the hypothesis that male and female teachers have a similar impact on girls' test scores relative to boys' in Vietnam.

---

[4] The F-statistic from comparing the estimate of β_1 in column 7 to column 8 was 0.14 and statistically insignificant. Thus I am unable to reject the hypothesis that the positive impact of male teachers on boys' English scores remained similar after including school fixed effects.

[5] It would have been preferable to include controls for parental education rather than literacy. However, information on both parents' education levels was missing for almost 40% of the sample. When parental education levels were included, the results remained similar and parental education was statistically insignificant. Thus the fact that children with more educated mothers are more likely to be allocated to female teachers within schools is unlikely to substantially affect the results.

4.1 Are the results for Pakistan driven by positive sorting?

In Pakistan, there was evidence of positive sorting towards male teachers based on lagged test scores. This should not substantially affect the estimate of β_1 since this is only identified from students in schools with both male and female teachers. While β_1 could be impacted by non-random attrition within schools, the positive impact of male teachers on test scores was found to be slightly stronger using a balanced panel, suggesting that the results were not driven by non-random attrition.

Positive sorting to male teachers could potentially impact the estimate of β_3 if high performing girls benefit more from being matched with female teachers since β_3 is identified from all schools with girls and female teachers, not just schools with both male and female teachers.

To test this, I replicated the results using the subsample of students who are taught by teachers of different genders in rounds 2 and 3. In this case β_3 is estimated only for students who are taught by both male and female teachers over time. After controlling for child characteristics, this subsample included 425 student-year observations for Maths and Urdu and 358 in English.

The subsample had baseline score that were 0.23 standard deviations higher than the rest of the sample in English, 0.16 standard deviations higher in Urdu and 0.05 standard deviations higher in Maths, which was consistent with positive sorting to schools with male teachers.

Controlling for classroom fixed effects, the estimate of β_3 rose from 0.05 to 0.13 in Maths but fell from -0.04 to -0.10 in Urdu and from -0.02 to -0.22 in English. Combining all subjects β_3 fell from 0.00 to -0.06. In all cases β_3 remained insignificant. Overall, these results suggest that the treatment effect of female teachers on girls was not underestimated due to positive sorting to male teachers.

As a further robustness check, I estimated the results using system GMM. System GMM estimates a system of equations involving the equation in levels (equation 3) and the equation in first differences (equation 5). Based on Andrabi et al's (2011) preferred specification, I assume that time varying inputs are predetermined but have a constant correlation with the individual fixed effect so that differences in time varying inputs can be included as instruments in the levels equation. Estimates of the impact of female teachers were generally similar to those under dynamic OLS.[6]

[6] Results available from the author on request.

Table 4: Impact of teacher gender in Pakistan

	All scores			*Math*			*English*			*Urdu*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lagged score	0.597***	0.482***	0.557***	0.572***	0.449***	0.595***	0.594***	0.422***	0.521***	0.626***	0.518***	0.625***
	(0.022)	(0.028)	(0.016)	(0.033)	(0.042)	(0.021)	(0.026)	(0.028)	(0.024)	(0.028)	(0.032)	(0.025)
Female teacher	-0.141***	-0.122**		-0.252***	-0.178**		-0.128**	-0.093		-0.043	-0.087	
	(0.042)	(0.056)		(0.048)	(0.089)		(0.058)	(0.092)		(0.044)	(0.070)	
Girl	0.076	0.094**	0.070*	-0.036	-0.021	-0.032	0.080	0.134**	0.116**	0.180***	0.166***	0.119***
	(0.047)	(0.041)	(0.036)	(0.060)	(0.057)	(0.050)	(0.059)	(0.055)	(0.055)	(0.050)	(0.044)	(0.038)
Female teacher* Girl	0.008	-0.012	-0.002	0.062	0.035	0.047	0.035	-0.005	-0.016	-0.073	-0.063	-0.039
	(0.055)	(0.047)	(0.042)	(0.069)	(0.066)	(0.057)	(0.068)	(0.063)	(0.062)	(0.058)	(0.052)	(0.046)
$\beta_1 + \beta_3$	-0.133**	-0.134*		-0.190***	-0.142		-0.093	-0.098		-0.115**	-0.150*	
	(0.055)	(0.073)		(0.066)	(0.104)		(0.064)	(0.095)		(0.055)	(0.078)	
$\beta_1 + p_g\beta_3$	-0.138***	-0.128**		-0.225***	-0.162*		-0.112**	-0.095		-0.075*	-0.115*	
	(0.040)	(0.060)		(0.046)	(0.090)		(0.050)	(0.088)		(0.040)	(0.069)	
Child characteristics & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject dummies	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
School FEs	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Classroom FEs	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	11737	11737	11737	3962	3962	3962	3813	3813	3813	3962	3962	3962

Note: (a) Standard errors are in parentheses and are clustered at the school level. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively.

Table 5: Impact of teacher gender in Vietnam

	All scores			***Math***			***Vietnamese***		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lagged score	0.459***	0.409***	0.388***	0.524***	0.455***	0.435***	0.393***	0.367***	0.363***
	(0.032)	(0.030)	(0.023)	(0.038)	(0.038)	(0.030)	(0.036)	(0.031)	(0.026)
Female teacher	-0.027	-0.087		0.046	0.065		-0.095	-0.222	
	(0.093)	(0.158)		(0.094)	(0.166)		(0.112)	(0.176)	
Girl	0.044	0.063	0.053	0.003	0.037	0.026	0.097*	0.093	0.085
	(0.042)	(0.051)	(0.044)	(0.054)	(0.065)	(0.056)	(0.059)	(0.065)	(0.062)
Female teacher* Girl	0.040	0.003	0.001	-0.032	-0.083	-0.076	0.111	0.092	0.076
	(0.062)	(0.069)	(0.066)	(0.071)	(0.080)	(0.074)	(0.078)	(0.085)	(0.084)
$\beta_1 + \beta_3$	0.013	-0.084		0.014	-0.018		0.017	-0.130	
	(0.081)	(0.161)		(0.092)	(0.171)		(0.088)	(0.172)	
$\beta_1 + p_g\beta_3$	-0.008	-0.086		0.030	0.025		-0.041	-0.177	
	(0.082)	(0.156)		(0.086)	(0.164)		(0.094)	(0.168)	
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FEs	No	Yes	No	No	Yes	No	No	Yes	No
Classroom FEs	No	No	Yes	No	No	Yes	No	No	Yes
Observations	5630	5177	5630	2779	2552	2779	2851	2625	2851

Note: (a) Standard errors are in parentheses and are clustered at the school level. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively. (c) Columns with school fixed effects exclude children allocated to classes on the basis of ability.

5 Identifying the mechanisms

The finding that female teachers do not significantly improve learning outcomes for either girls or boys in Vietnam and that male teachers actually improve learning outcomes in private schools in Pakistan, contrasts with a number of recent empirical studies in developing countries (Rawal and Kingdon 2010 and Muralidharan and Sheth 2015). This chapter examines the potential mechanisms behind these results by examining whether they can be explained differences in teacher characteristics, effort levels or 'stereotype effects'.

5.1 Can the results be explained by teacher characteristics or effort?

In theory, differences in the impact of male or female teachers on learning could be driven by differences in *observable teacher characteristics* (such as education, experience or subject specific knowledge) or differences in *effort levels* between male and female teachers. In particular, male teachers in Pakistan were generally more educated and had higher test scores (see Table 3) which might explain their positive impact on test scores.

To test this, Table 6 examines whether the treatment effect of being matched with a female teacher in Pakistan changes as teacher characteristics and measures of effort are included in the model. Column 1 contains the baseline specification, equivalent to column 2 of Table 4. Column 2 then adds observable teacher characteristics while Column 3 adds measures of teacher effort to the specification in column 1. Column 4 includes both teacher characteristics and measures of teacher effort.

The size of the treatment effect of female teachers on boys and girls remains relatively similar across all specifications. F-statistics testing whether β_1 and $(\beta_1 + \beta_3)$ in each specification differed from that in column 1 were statistically insignificant. Adding teacher characteristics also had little impact on the size of the coefficients at a subject level. Including teacher characteristics did increase the standard errors of β_1 and $(\beta_1 + \beta_3)$, with the impact of male teachers now being insignificant for girls and significant at only the 10% level for boys.

In the case of teacher effort, β_1 fell slightly and became insignificant. However, this was actually driven by the fact that information on teacher effort was missing for some students in the sample, with the positive impact of male teachers being weaker for those students with information on teacher effort. An important limitation here is that most of the available measures of teacher effort in LEAPS are self-reported so may not fully capture differences in effort across teachers. They may also be subject to misreporting or measurement error which can lead to attenuation bias.

A similar analysis was also completed for Vietnam.[7] Again the inclusion of teacher characteristics and measures of effort had little effect on the estimated impact of teacher gender, which remained insignificant.

5.2 Can the results for Pakistan be partly explained by stereotype effects?

Since the positive impact of male teachers in Pakistan was stronger in Maths than English and to a lesser extent Urdu, the results for Pakistan (although not Vietnam) are broadly consistent with 'stereotype effects' in which the academic stereotype that males are better at Maths and females at language in turn influences student's assumptions about a teacher's relative proficiency in each subject and their subsequent performance.[8] However, subject-specific stereotype effects do not imply that male teachers should have a positive impact overall.

One way of investigating the channels through which stereotype effects might arise would be to see whether student effort or assessments of teacher quality differ at the subject level. Unfortunately, such information was not available in LEAPS. Household level assessments of school teaching quality in Math and English, based

---

[7] Results available from the author on request.

[8] For a more detailed discussion of stereotype effects see Keucken and Valfort (2012).The literature has also identified *role model effects* and *teacher bias effects* as potential mechanisms to explain gains from a teacher-student gender match. However, I find no evidence of a matching effect here.

on a 1-5 scale, were available from the household survey for a small number of households of public and private school students. The difference in household Math and English teaching ratings (Math–English) was significantly negatively related to the proportion of female teachers in a school, but the effect was small. A 100% increase in female teachers reduced this difference by 0.08 units. Thus the evidence for stereotype effects at the household level was not particularly strong.

Another way of testing the existence of stereotype effects in Pakistan (but not the channels through which they arise) is to see whether they also hold for public schools. Since public schools tend to be gender-segregated in terms of students and teachers, few students experience a change in teacher gender across rounds. However, it is possible to identify the impact of teacher gender based on students who are taught by teachers of a different gender since 320 students in rounds 2 and 3 (excluding those not administered the child survey) are taught by teachers of a different gender, including some students attending non-segregated schools.

One potential concern is that these students may have very different characteristics to those matched with teachers of the same gender. As it turned out both groups had similar baseline test scores, years of schooling, household assets and parental education levels. However, those matched with a teacher of a different gender had teachers with significantly higher test scores and who were 18% more likely to hold a university degree, highlighting the need to control for teacher characteristics here.

Results from applying pooled OLS to equation (3) for public schools are shown in the first and third rows of Table 7.[9] Each cell in the first row shows the estimate of β_1 for all subjects, Math, English and Urdu respectively, while each cell in the third row shows the estimate of $(\beta_1 + \beta_3)$, the treatment effect of female teachers on girls. All estimates include controls for the teacher characteristics listed in Table 6.

[9] Using school fixed effects was considered but variation in teacher gender within public schools was relatively limited.

Interestingly, male teachers did not significantly improve overall test scores in public schools, but the pattern of results was consistent with stereotype effects as male teachers had a stronger impact on test scores in Math than English and Urdu.

Table 6: Evidence on mechanisms

	All scores			
	(1)	(2)	(3)	(4)
Lagged score	0.482***	0.487***	0.488***	0.490***
	(0.028)	(0.029)	(0.030)	(0.030)
Female teacher	-0.122**	-0.139*	-0.106	-0.130
	(0.056)	(0.077)	(0.076)	(0.082)
Girl	0.094**	0.081*	0.083*	0.076*
	(0.041)	(0.042)	(0.044)	(0.043)
Female teacher*Girl	-0.012	0.008	0.003	0.014
	(0.047)	(0.048)	(0.051)	(0.049)
Teacher characteristics				
Age		-0.007		-0.006
		(0.005)		(0.005)
Experience		0.031		0.016
		(0.038)		(0.035)
Teachers college		0.076		0.049
		(0.091)		(0.091)
Bachelor’s degree		0.117*		0.111
		(0.069)		(0.073)
Master’s degree		0.445***		0.407***
		(0.140)		(0.142)
Native to village		-0.154**		-0.150**
		(0.070)		(0.070)
Standardised test score		0.041**		0.036**
		(0.016)		(0.016)
Permanent contract		-0.243		-0.235
		(0.152)		(0.157)
Teacher effort				
Subject class time (mins/day)			0.002	0.001
			(0.011)	(0.010)
Preparation time (mins/day)			0.001	0.000
			(0.001)	(0.001)
Marking time (mins/day)			0.002**	0.002*
			(0.001)	(0.001)
Days absent last month^			-0.015	-0.009
			(0.013)	(0.014)
$\beta_1 + \beta_3$	-0.134*	-0.130	-0.103	-0.117
	(0.073)	(0.086)	(0.090)	(0.091)

$\beta_1 + p_g\beta_3$	-0.128**	-0.135*	-0.105	-0.124
	(0.060)	(0.078)	(0.078)	(0.082)
F-test on β_1		0.10	0.26	0.02
F-test on $(\beta_1 + \beta_3)$		0.01	1.00	0.10
Observations	11737	10683	10829	10497

Note: (a) All columns include child characteristics, subject and year dummies and school fixed effects. (b) Standard errors are in parentheses and are clustered at the school level. (c) ***, **,* denotes significant at 1%, 5% and 10% level respectively. (d)^Reported by head teacher.

The impact of teacher gender in public schools was also estimated using Propensity Score Matching (PSM). In this case, the treatment group was boys (girls) taught by female (male) teachers and the control group was boys (girls) taught by male (female) teachers. An advantage of PSM is that it removes the assumption of a linear relationship between the covariates and test scores, although functional form assumptions are still required to estimate the propensity score.

The sample was stratified by subject and gender and students were matched using nearest neighbour matching on: lagged test scores in all subjects; age; mother's and father's education; whether their father lived at home; teacher's test score in that subject; teacher's age; teacher's experience; teacher's highest general education qualification and whether the teacher was from the village.[10]

One disadvantage of using PSM is that it was not possible to match on all covariates since this created imbalance between treatment and control groups in some blocks of the estimated propensity score. However, matching on lagged scores and teacher scores is likely to capture most of the explained variation in potential outcomes.

The second and fourth rows of Table 7 show the estimated impact of female teachers from PSM. The impact of female teachers is based on the average treatment effect on the treated (ATT) for boys (since boys matched with female teachers are the treatment group) and the negative of the ATT for girls (since girls matched with female teachers are the control group).

[10] I did not include teacher experience in the match for boys in Urdu as this led to imbalance between treatment and controls in some blocks of the propensity score.

The results are broadly similar to those from OLS, although the impact of male teachers was found to be stronger in Math and weaker in English for girls.[11] The results for Maths are consistent with the earlier findings of Warwick and Jatoi (1994), although unlike them I do not find that this effect can be explained by teacher characteristics.

Overall, the results indicate that male teachers do not significantly improve overall test scores in public schools. However, these results should be interpreted cautiously. They are identified from a small proportion of students taught by teachers of a different gender who were more qualified on average. If these teachers also have unobservable characteristics that make them better teachers which are not captured by controlling for (or matching on) their observable characteristics, the impact of female teachers will be biased upward (downward) for boys (girls).

Table 7: The impact of female teachers in public schools in Pakistan

Boys (β_1)	All	Math	English	Urdu
Dynamic OLS with child & teacher characteristics	0.064	-0.066	0.103	0.163**
	(0.110)	(0.124)	(0.168)	(0.063)
Propensity score matching^	0.088	-0.156	0.120	0.126
	(0.251)	(0.199)	(0.237)	(0.139)
Girls ($\beta_1 + \beta_3$)				
Dynamic OLS with child & teacher characteristics	-0.114	-0.202	-0.009	-0.118
	(0.098)	(0.128)	(0.112)	(0.088)
Propensity score matching^	-0.108	-0.360***	0.176	-0.171
	(0.204)	(0.121)	(0.159)	(0.141)

Note: (a) Standard errors are in parentheses and are clustered at the school level. (b) ***, **,* denotes significant at 1%, 5% and 10% level respectively. (c) Unlike OLS, the ATT calculated from PSM for all subjects does not control for differences in average trajectories across subjects.

6 Conclusion

Understanding the impact of teacher gender on both girls and boys is important for policy makers looking to improve learning outcomes in developing countries. While a number of recent empirical studies have found that female teachers

[11] OLS results including lagged alternative subjects were closer to those from PSM. However, including alternate scores in a dynamic OLS framework alters the way in which results can be interpreted. This flexibility in functional form is one of the relative advantages of PSM here.

improve test scores for girls, I find no evidence that female teachers significantly improve test scores for girls or boys in either Vietnam or Pakistan.

These results suggests that unlike other education policies, which have been found to have similar effects in different contexts, the impact of female teachers is likely to be highly context specific. Indeed, the estimated impact of female teachers on girls in Pakistan was very different to that found in studies of neighbouring India.

Where longitudinal studies have found that female teachers improve learning outcomes for girls, the estimated impacts have been relatively small. Muralidharan and Sheth (2015) and Paredes (2014) find that female teachers improve overall test scores by 0.036 and 0.018-0.035 standard deviations a year respectively. From a policy perspective, this suggests that basing hiring decisions on teacher gender is unlikely to be the most effective policy for improving test scores for girls, although identifying more effective educational interventions remains challenging as noted in a recent literature review by Glewwe et al (2013).

The results of this extended essay also suggest that the impact of teacher gender may vary across school sectors. In Pakistan I find that male teachers significantly improve overall test scores for both boys and girls in private schools by around 0.13 standard deviations a year, which did not appear to be driven by differences in observable teacher characteristics or effort levels. However, male teachers were not found to significantly improve overall test scores in public schools which suggests that the results for private schools could be driven by differences in unobservable characteristics between male and female teachers in private schools.

Finally, the impact of male teachers in Pakistan was much stronger in Maths than English and Urdu in both public and private schools. This suggests that 'stereotype effects' may play a role in teacher-student gender interactions in some developing countries. This is consistent with the findings of Keucken and Valfort (2012) for Sub Saharan Africa. While there was no evidence for stereotype effects in Vietnam, gender disparities in school enrolment and labour force participation have been historically much smaller in Vietnam than Pakistan or sub-Saharan Africa.

While the presence of ‘stereotype effects’ suggests that education departments in developing countries may need to actively work to break down these stereotypes, particularly in contexts where girls are less likely to undertake further study in Maths or Science, future research should seek to explore in more detail how stereotype effects arise and how they impact longer term outcomes.

References

Andrabi, T., Das, J. and Khwaja, A. 2006, 'A Dime A Day: The Possibilities and Limits of Private Schooling in Pakistan', World Bank Policy Research Working Paper 4066.

Andrabi, T., Das, J., Khwaja, A., Vishwanath, T. and Zajonc, T. 2007, 'Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate', Report.

Andrabi, T., Das, J., Khwaja, A. and Zajonc, T. 2011, 'Do Value Added Estimates Add-Value? Accounting for Learning Dynamics', *American Economic Journal: Applied Economics,* 3: 29-54.

Arellano, M. and Bond, S. 1991, 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies,* 58(2): 277-97.

Antecol, H., Eren, O. and Ozbeklik, S. 'The Effect of Teacher Gender on Student Achievement in Primary School', *Journal of Labor Economics,* 33(1): 63-89.

Chetty, R., Friedman, J. and Rockoff, J. 2014, 'Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates', *American Economic Review,* 104(9): 2593-2632.

Das, J. and Zajonc, T. 2010, 'India shining and Barat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement', *Journal of Development Economics,* 92: 175-187.

Dee, T. 2004, 'Teachers, Race and Student Achievement in a Randomized Experiment', *The Review of Economics and Statistics*, 86: 195-210.

— 2005, 'A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?', *American Economic Review*, 95: 158-165.

— 2007, ‘Teachers and the Gender Gaps in Student Achievement’, *Journal of Human Resources*, 42: 528-554.

Deming, D., Hastings, J., Kane, T., Staiger, D., 2014, ‘School choice, school quality and academic achievement’, *American Economic Review,* 104 (3): 991–1013.

Glewwe, P., Hanushek, E., Humpage, S. and Ravina, R. 2013, ‘School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010’, in Glewwe, P. (Ed.), *Education Policy in Developing Countries,* University of Chicago Press, 13-64.

Guarino, C., Reckase, M. and Wooldridge, J. 2014, ‘Can value-added measures of teacher performance be trusted?’, *Education Finance and Policy*, 10(1): 117-156.

Kane, T. and Staiger, D. 2008, ‘Estimating teacher impacts on student achievement: an experimental evaluation’, NBER Working Paper 14607.

Kane, T., McCaffrey, D., Miller, T. and Staiger, D. 2013, ‘Have we identified effective teachers? Validating measures of effective teaching using random assignment’, MET Project Research Paper, Bill & Melinda Gates Foundation.

Kuecken, M. and Valfort, M-A 2012, ‘The Impact of Student-Teacher Gender Interactions on Learning Outcomes. Evidence from Sub-Saharan Africa’, mimeo.

Muralidharan, K. and Sheth, K. 2015, ‘Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers, *Journal of Human Resources* (forthcoming).

Muralidharan, K. and Sundararaman, V., 2013, ‘Contract teachers: experimental evidence from India’, NBER Working Paper 19440.

Paredes, V. 2014, ‘A teacher like me or a student like me? Role model versus teacher bias effect’, *Economics of Education Review,* 39: 38-49.

Rawal, S. and Kingdon, G. 2010, ‘Akin to My Teacher: Does Caste, Religious, or Gender Distance between Student and Teacher Matter? Some Evidence from India’, Institute of Education, University of London.

Rothstein, J. 2010, ‘Teacher quality in educational production: tracking, decay, and student achievement’, *Quarterly Journal of Economics,* 125(1): 175-214.

Singh, A. 2015, ‘Private school effects in urban and rural India: Panel estimates at primary and secondary school ages’, *Journal of Development Economics,* 112: 16-32.

United Nations Development Group Taskforce on the Millennium Development Goals 2010, ‘UNDG Millennium Development Goals Thematic Papers: MDG3 Promote Gender Equality and Empower Women’.

United Nations Educational, Scientific and Cultural Organization 2005*, ‘Scaling up’ good practices in girls’ education*.

Warwick, D. and Jatoi, H. 1994, ‘Teacher Gender and Student Achievement in Pakistan’, *Comparative Education Review,* 38(3): 377-399.

Winters M., Haight R., Swaim T., and Pickering K. 2013, ‘The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data’, *Economics of Education Review*, 34: 69–75.

World Bank Group 2015a, ‘Global Monitoring Report 2014/2015: Ending Poverty and Sharing Prosperity’, Washington, DC: World Bank.

World Bank 2015b, ‘World Development Indicators’.