

Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers

Karthik Muralidharan[†]

Ketki Sheth

26 August 2013

Abstract: Recruiting female teachers is frequently suggested as a policy option for improving girls' education outcomes in developing countries, but there is surprisingly little evidence on the effectiveness of such a policy. We study gender gaps in learning outcomes, and the effectiveness of female teachers in reducing these gaps using a large, representative, annual panel data set on learning outcomes in rural public schools in the Indian state of Andhra Pradesh. We report six main results in this paper. (1) We find a small but significant negative trend in girls' test scores in both math ($0.02\sigma/\text{year}$) and language ($0.01\sigma/\text{year}$) as they progress through the public primary school system; (2) Using five years of panel data, school-grade and student gender by grade fixed effects, we find that both male and female teachers are more effective at teaching students of their own gender; (3) However, female teachers are more effective overall, resulting in girls' test scores improving by an additional 0.036σ in years when they are taught by a female teacher, with no adverse effects on boys when they are taught by female teachers; (4) The overall gains from having a female teacher are mainly attributable to their greater effectiveness at improving math test scores than male teachers (especially for girls); (5) We find no effect of having a same-gender teacher on student attendance, suggesting that the mechanism for the impact on learning outcomes is not on the extensive margin of increased school participation, but on the intensive margin of more effective classroom interactions; (6) Finally, the increasing probability of having a male teacher in higher grades can account for around 10-20% of the negative trend we find in girls' test scores as they move to higher grades.

JEL Classification: I21, J16, O15

Keywords: education gender gaps, girls' education, female teachers, India, longitudinal data, MDG's, social distance of teachers

[†] Karthik Muralidharan: UC San Diego, NBER, BREAD, J-PAL; E-mail: kamurali@ucsd.edu
Ketki Sheth: UC San Diego; E-mail: ksheth@ucsd.edu

We thank Prashant Bharadwaj, Julie Cullen, Gordon Dahl, Craig McIntosh, and several seminar participants for comments. We also thank the AP RESt team and the Azim Premji Foundation for collecting the data used in this paper, and thank Venkatesh Sundararaman for the overall support provided to the AP RESt project. Financial assistance for the data collection was provided by the Government of Andhra Pradesh, the UK Department for International Development (DFID), the Azim Premji Foundation, and the World Bank. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of any of the organizations that supported the data collection.

1. Introduction

Reducing gender gaps in education attainment has been an important priority for international education policy, and is explicitly listed as one of the United Nations Millennium Development Goals (MDGs). This commitment has been reflected in the policies of many developing countries, and substantial progress has been made in the past decade in reducing gender barriers in primary school enrollment. One key policy that is credited with increasing girls' education is the increased recruitment of female teachers (UNESCO 2012, Herz and Sperling 2004, UN 2012). UNICEF has documented the practice in a variety of countries, including Bangladesh, India, Liberia, Nepal, and Yemen, and the United Nations Task Force for achieving the MDGs has advocated hiring more female teachers as an effective policy mechanism for reaching the goal of universal primary education of girls (UNDG 2010, Rehman 2008, Slavin 2006).

While the idea that hiring more female teachers can bridge gender gaps is widely prevalent among policy makers, there is very little empirical evidence on testing this hypothesis in developing countries. In this paper, we study the causal impact of having a female teacher on the learning gains of female students, using one of the richest datasets on primary education in a developing country. The dataset features annual longitudinal data on student learning measured through independent assessments of learning conducted over five years across a representative sample of 500 rural schools and over 90,000 students in the Indian state of Andhra Pradesh. The data also includes detailed information on teacher characteristics and on their assignments to specific classrooms in each year.

The combination of panel data and variation in the gender of teachers and students allows us to estimate the causal impact of matching teacher and student gender in a *value-added* framework. Identification concerns are addressed by showing that our causal estimates of gender matching do not change under an increasingly restrictive set of specifications including school, school-grade, and student gender by grade fixed effects. We also show that there is no correlation between the probability of being assigned a female teacher and either the fraction of female students in the class or their mean test scores at the start of the year. Further, our estimation sample is restricted to schools that only have one section per grade, which precludes the possibility that students may be tracked across sections and that female teachers may be assigned to different sections based on unobservables.

We report six main findings in this paper. First, we find a small but significant negative trend in girls' test scores in both math ($0.02\sigma/\text{year}$) and language ($0.01\sigma/\text{year}$) as they advance through the five grades of primary school.¹ Girls have significantly higher test scores in language and equal test scores in math relative to boys at the end of grade one, but score almost on par with boys in language and significantly worse in math by the end of grade five. These results are consistent with evidence of gender gaps in test scores (particularly in math) documented in both high and low income countries (Fryer and Levitt 2010, Bharadwaj et al. 2012), and suggest that the growing gender gaps documented at later ages in both these papers probably reflect a cumulative effect of a trend that starts as early as primary school.

Second, using five years of panel data and school-grade and student gender by grade fixed effects, we find that teachers are $.034\sigma/\text{year}$ more effective in teaching students of their own gender *relative* to teachers of the opposite gender. In other words, female teachers are $.034\sigma/\text{year}$ more effective at reducing the gender gap in achievement than male teachers. Since female teachers differ from male teachers on several characteristics that may be correlated with teacher quality, we test the robustness of the 'gender-match' result by including interactions between student gender and each of the teacher characteristics on which female and male teachers differ, and find that our estimates above are essentially unchanged.

The result above is a difference-in-difference estimate that compares the relative advantage of female teachers in teaching girls rather than boys with the relative disadvantage of male teachers in teaching girls rather than boys, and is *symmetric by construction*. However, the overall effectiveness of a teacher is also determined by his or her effectiveness at teaching students of the opposite gender. Our third result speaks to this issue and we find that female teachers in our setting are more effective overall than male teachers. We find that girls who have a female teacher in a given year have $.036\sigma$ higher annual test score gains than if she had a male teacher. However, boys perform similarly regardless of the gender of their teacher. Thus, girls are likely to benefit from a policy of hiring more female teachers, and overall educational performance is likely to increase due to the lack of any offsetting effect on boys.

Fourth, these effects differ by subject. In particular, female teachers are more effective at teaching math relative to language when compared to male teachers. While girls continue to

¹ As we discuss later, this estimate is based on the sample of test takers in public schools, and cannot account for the biases that may occur due to differential migration to private schools and differential absence on the day of the test by gender.

fare better with female teachers relative to male teachers in both language and math, the effect is greater in math relative to language. Boys though fare a little worse with female teachers (relative to male teachers) in language, and experience no differential effect of teacher gender in math. Together, these results suggest that the overall gains from hiring female teachers come mainly from improving mean math test scores relative to male teachers (positive for girls, no effect for boys) than from language (positive for girls, negative for boys, and no overall effect).

Fifth, we also study the impacts of a teacher-student gender match on student attendance, and find no evidence that teachers are more effective at raising the attendance for students of the same gender. This suggests that the likely mechanism for the 'matching' effect on test scores is not on the extensive margin of increased student-teacher contact time, but rather on the intensive margin of more effective classroom interactions.

Finally, we document that female teachers are more likely to teach in earlier grades. Combined with the results above, we estimate that around 10-20% of the trend of increasing gender gaps in test scores over time can be attributed to the reduction in the probability of girls being taught by female teachers as they advance to higher grades. Since teachers in higher grades are more likely to be male across several countries (NCES 2011, UNESCO 2010), our results suggest that an important channel for growing gender gaps in achievement (especially in math) could be the greater likelihood of having male teachers in higher grades.

While there have been several studies on the impact of shared gender between teachers and students on learning outcomes in developed country contexts, there is surprisingly little well-identified evidence on this question from developing countries. In the US and UK, studies have shown improved test scores, teacher perception, student performance, and engagement of girls when taught by a female teacher in schools, with magnitudes of test score impacts similar to those found in our paper (Dee 2007, Dee 2005, Nixon and Robinson 1999, Ehrenberg et al. 1995, Ouazad and Page 2012). However, other studies conducted in both the US and in European countries have failed to find such an effect (Holmund and Sund 2008, Carrington, Tymms and Merrell 2008, Lahelma 2006, Winters et al. 2013, Marsh et al. 2008, Driessen 2007, Neugebauer et al. 2011). In higher education institutions in the US, female professors have been found to

have small effects on female students' course selection, achievement, and major choice (Bettinger et al. 2004, Carrell et al. 2010, Hoffmann and Oreopoulos 2009).²

However, the question of the role of female teachers in reducing gender gaps is much more salient in developing country contexts, where gender gaps in school enrollment and attainment are much larger and increased recruitment of female teachers is actively advocated (OECD 2010, Hausmann et al. 2012, Muralidharan and Prakash 2013, Bharadwaj et al. 2012). The only related paper in a developing country setting is Rawal and Kingdon (2010), who use test score data on 2nd and 4th grade students in the Indian states of Bihar and Uttar Pradesh, and find a positive impact on educational achievement for girls taught by female teachers, but find no similar effect for boys.

In addition to providing well-identified estimates of the impact of matching teacher and student gender on learning outcomes in a developing country (where the literature is very sparse), our dataset allows us to make advances relative to both the developed and developing country literatures on this subject. First, while several existing papers in this literature (especially those looking at college-level outcomes) use grades or test scores assigned by the students' own teachers, the test scores used in this paper are based on independent assessments and grading. This limits the concern that the measured effects of gender matching may reflect more generous grading by teachers towards students who share their own gender and allows us to be confident that the effects we measure reflect genuine impacts on learning.

Second and more important, the majority of papers in the global literature on this question (including Dee 2007 and Rawal and Kingdon 2010) use student fixed effects and variation in the gender of teachers across different subjects to identify the impact of the gender match on learning, but they are based on comparing *levels* of test scores as opposed to value added. Thus, it can be difficult to interpret the magnitudes of the estimated effects without knowing the gender composition of the teachers in that subject in previous grades.³ Our use of five years of annual

²Analogous to gender, studies in the United States have also looked at the effect of sharing the ethnicity of a teacher and have generally found positive effects on such educational outcomes as drop outs, pass rates, and grades at the community college level, and teacher perceptions and student achievement in school going children (Dee 2004, Dee 2005, Farlie et al. 2011). We find no similar effect on other important dimensions in the Indian context, particularly disadvantaged castes and minority religions. We do not focus on caste and religion because the fraction of teachers and students in the relevant categories are small (typically less than 20%) and as a result the fraction of 'matches' are usually less than 5% (and often much smaller), which makes the estimates less stable to the series of robustness checks that we use in this paper to ensure that the estimates of the 'match' are well identified.

³ Thus, if this approach finds that a girl in eighth grade who has a female language teacher and a male math teacher does better in language, the interpretation of the point estimate is confounded by the possibility that the girl is also

panel data on test scores allows us to estimate the impact of a gender match on the *value-added in the year that the match occurred*, which has a much clearer interpretation relative to the standard in the literature. Finally, we observe students at a younger and more formative age than most of the literature, when the role of sharing gender may be especially important. This is also the age that is most relevant to policy for reducing education gender gaps in developing countries since the majority of students do not complete more than eight years of school education.

The remainder of this paper is organized as follows: Section 2 describes the dataset and presents summary statistics on students and teachers; Section 3 lays out the estimation and identification strategies; Section 4 presents the main results, and section 5 concludes.

2. Context and Dataset

India has the largest primary schooling system in the world, catering to over 200 million children. As in other developing countries, education policy in India has placed a priority on reducing gender disparities in education, and both the Five Year Plans and Sarva Shiksha Abhiyan (SSA), the flagship national program for universal primary education, have called for an increase in recruiting female teachers as a policy for increasing girls' education. SSA requires that 50% of new teachers recruited be women, and the 11th Five Year Plan suggested that it be increased to 75% (Government of India 2008). These calls for increased female teachers reflect a belief that through such mechanisms as role model effects, increased safety, reduced prejudices, and greater identification and empathy, female teachers are arguably more effective in increasing girls' achievement in primary school relative to their male counterparts (Ehrenberg et al. 1995, Stacki 2002, Dee 2005).

This paper uses data from the Indian state of Andhra Pradesh (AP), which is the 5th most populous state in India, with a population of over 80 million (70% rural). The data was collected as part of the Andhra Pradesh Randomized Evaluation Studies (AP RESt), a series of experimental studies designed to evaluate the impact of various input and incentive-based interventions on improving education outcomes in AP.⁴ The project collected detailed panel data over five years (covering the school years 2005-06 to 2009-10) on students, teachers, and households in a representative sample of 500 government-run primary schools (grades 1 through

more likely to have had female language teachers in earlier grades (especially if teacher gender is correlated with subjects taught across grades, which is likely to be true).

⁴ These interventions are described in Muralidharan and Sundararaman (2011).

5) across 5 districts in Andhra Pradesh. The dataset includes annual student learning outcomes as measured by independently conducted and graded tests in language (Telugu) and math (conducted initially at the start of the 2005-06 school year as a baseline, and subsequently at the end of each school year), basic data on student and teacher demographics, and household socioeconomic data for a subset of households. The test scores are normalized within each year-grade-subject combination and all analysis is conducted in terms of normalized test scores, with magnitudes being reported in standard deviations.

The Appendix provides further details on the dataset, including sample size and attrition between years. There is some differential attrition in the sample over time by gender (where attrition is defined as the fraction of students who had taken a test at the end of year 'n-1', but did not take a test at the end of year 'n'), with female students more likely to be in the test-taking sample (around 3% each year). However, this attrition over time is not a first-order concern for this paper because it is highly unlikely that the additional 3% of female students who appear for the test each year (relative to boys) would have test scores that are *differentially affected* by having a female teacher. This is further supported by Table 8 and by Appendix Table 3, where we show that having the same gender as the teacher does not have any impact on either student attendance on a typical school day or on student presence in the end of year test.

Table 1 - Panel A, presents descriptive statistics on students who have at least one recorded test score and data on gender in the dataset.⁵ Girls comprise 51% of the sample of public-school students in our sample. This does not imply that more girls are going to school than boys since it is likely that more boys are attending private schools (Pratham 2012). However, it does illustrate that on average, girls are well represented in public primary schools and in our sample. The girls in the sample come from modestly better off socioeconomic backgrounds than the boys, and have parents who are slightly more educated and affluent. These differences probably reflect two dimensions of selection into the sample – better off households are more likely to send girls to school, and better off households are more likely to send boys to private schools. However, the magnitudes of these differences are quite small (often in the range of 0-2 percentage points), and the statistical significance reflects the very large sample size. Since the household surveys

⁵ Less than 3 percent of students with test scores have no recorded gender.

were completed for only 70% of the sample of students for whom we have test score data, our main specifications do not include household controls.⁶

Table 1 - Panel B, presents summary statistics for the teachers in our analysis. Female teachers comprise 46% of the total teacher body, but are less experienced, less likely to have completed high school or a masters degree, and less likely to hold a head-teacher position. Not surprisingly, their mean salaries are also lower. They also comprise a much greater share of the contract teacher work-force than that of regular civil-service teachers. Since teacher characteristics vary systematically by gender, we will report our key results on the impact of matching teacher and student gender, both with and without controls for these additional teacher characteristics. We will also conduct robustness checks of our main results on the effects of a teacher-student 'gender match' on learning outcomes, by including interactions of student gender with each of the teacher characteristics that are different across male and female teachers.

Table 2 - Panel A presents summary statistics on gender differences in test scores by grade. We see that girls score as well as boys in math and score 0.05σ *higher* on language in grade 1. However, there is a steady decline in girls' test scores in both math and language as they advance through higher grades, and by the last two years of primary school (grades 4 and 5) we see that girls' initial advantage in language scores has declined and they do significantly worse than boys on math (by around $.1\sigma$). Table 2 - Panel B quantifies the annual decline in girls' relative scores by including an interaction term between student gender and grade in a standard value-added specification. We find evidence of a growing education gender gap among test takers in public primary school, with a mean decline of 0.02σ /year in math scores and 0.01σ /year decline in language scores for girls relative to boys. Since the data includes nine different cohorts of students (see Appendix), we also include cohort fixed effects, and see that the estimates of the gender gaps and of the trend in the gender gap across grades are unchanged. Similarly, the results are also robust to including school fixed effects.

One caveat to the interpretation of the above numbers is that they are based on a representative sample of test-taking students in public schools. Relative to the gender gap in the

⁶ While there are a few observable differences between the boys and girls in the sample, including these in the estimation will only matter if there are differential interactions between these household characteristics and teacher gender across boys and girls. We verify that our results are robust to the inclusion of household characteristics, but prefer to not include household characteristics in our main estimating equations because doing so reduces the sample size by 30% and it is possible that the remaining sample may have some non-random attrition. Results with household controls are available on request.

universe of primary-age school children, our estimate may be biased downwards if higher-scoring boys are differentially more likely to attend private schools. Conversely, they may be biased upwards if lower-scoring boys are more likely to be absent on the day of testing. While we cannot estimate these, it is more likely that we under-estimate the gender gap, because boys aged 7-10 in rural AP are around 10 percentage points more likely to be enrolled in a private school during this period (45% versus 35% in 2010 - Pratham 2010), whereas girls in public schools are only 3% more likely to be present on the day of testing (Appendix Table 3).

In spite of these caveats, this documentation of gender gaps in a representative sample of public schools in rural AP is a useful contribution to the literature on gender gaps in test scores in developing countries, because there are very few longitudinal data sets on student test scores in low-income settings, and no other paper that we are aware of is able to document these gaps with *cohort fixed effects*. Further, the literature on gender gaps in test scores mostly relies on samples of students who take tests in schools, and therefore has the same limitations we discuss above.

3. Estimation and Identification

Our main estimating equation takes the form:

$$(1) \quad E_{itjk} = \alpha + \gamma E_{it-1j-1k} + \beta_1(F * g)_{itjk} + \beta_2 g_{itjk} + \beta_3 F_{itjk} + \delta T_{itjk} + \mu_{itjk}$$

where E_{itjk} are student educational outcomes (test scores and attendance) for student i , in year t , grade j , and school k respectively. g_{itjk} is an indicator for whether the student is a girl, F_{itjk} is an indicator for whether the student's current teacher is female, and $F * g_{itjk}$ is an indicator for whether a girl student shares her teacher's gender in the current year. T_{itjk} is a vector of additional teacher characteristics, and μ_{itjk} is a stochastic error term. The inclusion of the lagged test score on the right-hand side of (1) allows us to estimate the impact of contemporaneous inputs in a standard *value-added* framework. Since all test scores are normalized by grade and subject, the estimated coefficients can be directly interpreted as the correlation between the covariate and annual gains in normalized test scores.⁷ When studying attendance we do not include the lagged attendance of the previous year.

⁷ In the case of grade 1 where there is no lagged score (since there was no testing prior to enrolling in school), we set the normalized lagged score to zero. Our results on the impact of 'gender matching' on test score gains are unchanged if we drop grade 1 from the analysis.

The above estimating equation allows us to calculate the marginal impact of changing each component of the feasible student-teacher gender combinations relative to boys taught by male teachers (the omitted category).

The first coefficient of interest in this paper is β_1 , which indicates the extent to which teachers are relatively more effective at teaching to their own gender compared to teachers of the opposite gender. Since the indicator variable is based on the interaction of dummies for teacher and student gender, the coefficient is a 'difference in difference' estimate of the impact of female teachers when teaching girls rather than boys *relative* to their male counterparts teaching girls rather than boys. The coefficient on the interaction term therefore reflects the sum of the relative advantage of female teachers when teaching girls (rather than boys) and the relative disadvantage of male teachers when teaching girls (rather than boys). (i.e., $\beta_1 = (\text{female teachers teaching girls} - \text{female teachers teaching boys}) - (\text{male teacher teaching girls} - \text{male teachers teaching boys})$).

A more intuitive way of understanding this is to note that β_1 represents the relative effectiveness of female teachers (compared to male teachers) in reducing the test score gap between girls and boys. By construction, this is symmetric and equivalent to the relative effectiveness of male teachers teaching boys compared to girls *relative* to female teachers teaching boys compared to girls. It is important to highlight that a positive β_1 does not necessarily imply that both boys and girls have better outcomes when sharing their teacher's gender. For example, a positive β_1 could co-exist with a situation where all students are better off with female (or male) teachers because the general effectiveness of female (or male) is considerably higher (even for students of the opposite gender).

β_2 is the difference in test score gains of girls taught by male teachers relative to boys taught by male teachers (i.e., $\text{male teachers teaching girls} - \text{male teachers teaching boys}$). β_3 is the difference in test score gains of boys taught by female teachers relative to when taught by male teachers (i.e., $\text{female teachers teaching boys} - \text{male teachers teaching boys}$). Thus, β_3 estimates the extent to which boys perform differently when they are taught by a female teacher relative to a male teacher.

Starting with the omitted category (of male teachers teaching boys), adding combinations of β_1, β_2 , and β_3 allow us to measure other marginal effects of interest. Analogous to β_3 for boys,

testing if $\beta_1 + \beta_3 > 0$ provides a formal test of whether girls gain by being paired with female teachers relative to male teachers. The derivation is below:

$$(2) \text{ Female teachers teaching girls} - \text{Male teachers teaching girls} > 0 \\ \Rightarrow (\alpha + \beta_1 + \beta_2 + \beta_3) - (\alpha + \beta_2) > 0 \Rightarrow \beta_1 + \beta_3 > 0$$

As highlighted earlier, it is possible that female teachers are relatively more effective at teaching girls than boys compared to male teachers (a positive β_1), but that female teachers are overall less effective (a negative β_3), resulting in girls being better off with male teachers despite the loss in gains from not sharing their teacher's gender ($\beta_1 + \beta_3 < 0$).

Additionally, if we value both boys' and girls' educational achievement equally, then we would be interested in knowing whether the positive gain for girls taught by female teachers outweighs any adverse effects from mismatching boys to being taught by female teachers (i.e., *potential gain to girls + potential loss to boys*). The formal test for this is $\lambda_g \beta_1 + \beta_3 > 0$ where λ_g is the proportion of girls in schools. The derivation is below:

$$(3) \lambda_g * \text{potential gain to girls} + (1 - \lambda_g) * \text{potential loss to boys} > 0 \\ \Rightarrow \lambda_g(\beta_1 + \beta_3) + (1 - \lambda_g)(\beta_3) > 0 \Rightarrow \lambda_g \beta_1 + \beta_3 > 0$$

Thus, if the effect of female teachers on boys was negative, but their effect on girls was positive, we would find that: $\beta_3 < 0$ and $\beta_1 + \beta_3 > 0$. The test outlined in Eq (3) can also be interpreted as the overall effectiveness of female teachers relative to male teachers. Intuitively the impact of replacing a male teacher in a classroom with a female teacher is equal to the sum of the impact of the female teacher on all students (β_3), and the additional gains to female students from matching with a female teacher (β_1), weighted by the fraction of female students in the classroom (λ_g).

The main identification challenge in interpreting these coefficients causally is that teachers are not randomly assigned to schools, and it is possible that schools with more female teachers are in areas with greater overall girls' education levels and steeper learning trajectories. Thus, it is possible that girls would perform well in these schools regardless of their teacher's gender. In such a case, the estimate of β_1 could be confounded by omitted variables correlated with both the probability of having a female teacher and steeper learning trajectories for girls. We address this concern by augmenting (1) with school fixed effects, and thereby estimating the impact of a

gender-match on value-added *relative* to the schools' average effectiveness at improving value-added.

A further concern could be that teachers are not assigned randomly to grades within schools, and a similar omitted variable concern would apply if female teachers are differentially assigned to grades in which girls are more likely to show greater test score gains (for instance, if female teachers are more likely to be assigned to younger grades and if girls outperform boys in earlier grades). To address this concern, we include school grade fixed effects, which controls for the average performance in a given *grade* in the school (instead of the overall performance of the school). Finally, to account for potentially differential trajectories of learning in different grades *by gender*, we also include grade fixed effects by student gender to estimate the parameters of interest by comparing educational outcomes relative to girls' and boys' average learning trajectories in each grade. Our preferred specification therefore includes both school-grade fixed effects and grade fixed effects by gender to address this concern.⁸

A final concern could be that if grades in a school have multiple sections, then the assignment of teachers to sections within grades could be based on omitted variables such as a greater probability of assigning female teachers to sections that have girls with a greater likelihood of improving test scores. However, this is not an important factor in our setting because schools typically have fewer teachers than grades, and the typical teaching arrangement is one of multi-grade teaching (where the same teacher simultaneously teaches multiple grades) and so there are only few cases where there are multiple sections per grade with different teachers assigned to different sections. We drop all such cases (6% of observations) where there are multiple teachers per grade.

Note that our identification strategy does not require teacher gender to switch in a given school grade over time, and neither does it require teacher gender to switch within a cohort over time (across different grades).⁹ Rather, the inclusion of school-grade and gender grade fixed effects implies that the identifying variation is coming from the differential effectiveness of

⁸ Since the data are drawn from schools that were exposed to various experimentally-assigned programs, all estimates include dummy variables indicating the treatments assigned to the school. This turns out to not matter in practice because our main specifications of interest use school-fixed effects, which makes the treatment status of the school irrelevant for identification purposes.

⁹ We avoid using a student fixed effects estimate because the identifying variation in a specification with student fixed effects would come from changes in teacher gender in different grades. However, as we see in 3.1, girls have higher value-added in lower grades, and female teachers are more likely to be assigned to lower grades. This would therefore create an upward bias in the 'matching estimate'.

teachers (by gender) at teaching girls versus boys relative to (a) the mean value added experienced by students in that school and grade over the five years of data, and (b) the mean value added for girls relative to boys in that grade across all schools in the sample.

3.1. Testing the Identifying Assumptions

Table 3 shows the correlation between various classroom characteristics and the probability of the classroom having a female teacher. We see that there is no significant correlation between having a female teacher and the fraction of girls in the classroom or with the average test scores of incoming cohorts for either gender. Female teachers *are* more likely to be assigned to younger grades. But once school-grade fixed effects are included, this is no longer an issue for average female teacher effects, and it continues to be case that there is no significant correlation between having a female teacher in the class and either the fraction of female students or the test scores of the incoming cohort (columns 5 and 6).

However, we see in Table 4 that girls do have a slightly more concave learning trajectory than boys. We estimate a standard value-added model that controls for lagged test scores (as in Eqn. 1), but allow for an interaction between student gender and grade, and we see that female students have lower value-added in higher grades. Since female teachers are more likely to be assigned to lower grades, the inclusion of school-grade fixed effects (i.e., the average test score gain in a grade within a school over the five years across both student genders) does not address the possible spurious correlation from female teachers being more likely to be assigned to grades where female students fall behind boys at a lower rate. Therefore, we also include grade fixed effects by student gender in our main specifications to control for average value-added test scores in each grade *by student gender*. Thus, the parameters of interest in Eq. (1) are identified relative to the average learning trajectory for girls in the same grade (student gender grade fixed effects) and relative to the average learning trajectory in the same school for that grade (school-grade fixed effects).

We also verify that there is no significant difference between classrooms taught by male and female teachers on any of the household socio-economic variables listed in Table 1 (tables available on request), but we focus our attention on the test-scores of incoming cohorts as the most useful summary statistic of previous inputs into education to test balance on, because the sample size with the household survey is 30% smaller than that of just the test scores.

4. Results

The main results of the paper (from the estimation of Equation 1) are presented in Table 5, which pools the results across subjects (results separated by subject are in Table 7). The columns show increasingly restrictive identification assumptions with school fixed effects (Column 2), school-grade fixed effects (Column 3), and both of these with grade fixed effects by student gender (Column 4 and 5). Column 6 expands the preferred specification in Column 5 with the inclusion of teacher covariates to differentiate between a pure "gender effect" versus effects driven by teacher characteristics correlated with teacher gender. Thus, the estimates in column 5 are relevant to the policy question: "What will happen if we replace a male teacher with a female teacher whose characteristics are the same as those of the average female teacher?" On the other hand, the estimates in column 6 answer the question: "What will happen if we just switch a teacher's gender from male to female holding other observable characteristics constant?" While our main results are remarkably stable and robust under the various specifications, our discussion below will use the estimates in columns 5 and 6, unless mentioned otherwise.

Averaged across subjects, we see that teachers are $.034\sigma$ /year more effective in teaching to their own gender relative to a student of the opposite gender compared to teachers of the other gender. In other words, female teachers are $.034\sigma$ /year more effective in reducing the gender gap between girls and boys relative to male teachers. We find no negative effect on boys from being taught by female teachers relative to male teachers (β_3 is close to zero). We estimate that girls gain an extra $.036\sigma$ /year when taught by female teachers instead of male teachers ($\beta_1 + \beta_3$), and that there is a statistically significant net increase in annual test score gains of $.019\sigma$ /year from replacing a male teacher with a female one ($\lambda_g * \beta_1 + \beta_3$). However, once we control for teacher characteristics, this net welfare effect drops to $.013\sigma$ /year, suggesting that characteristics correlated with female teachers may partly contribute to female teachers being more effective overall.

This discussion points to an important caveat to the interpretation of these results. Since female teachers are systematically different from their male counterparts (Table 1 - Panel B), it is possible that the β_1 estimated in (1) reflects not just the effect of female students matching with female teachers, but the effect of female students matching with teacher *characteristics* that are systematically more commonly found in female teachers. We address this concern in Table 6, where we show a series of regressions where we follow the specification in (1), but include

teacher characteristics and the interaction of this characteristic with student gender. These include teacher demographic characteristics that may be correlated with teaching effectiveness (such as education, training, contractual status, seniority, and salary) as well as teaching conditions (multi-grade teaching) and measures of teacher effort (absence). Doing so allows us to test the extent to which the positive β_1 found in Table 5 reflects a 'gender' match as opposed to other characteristics of female teachers that differentially effect girl students.

Panel A of Table 6 reports the key results without controlling for other teacher characteristics, while Panel B includes them. The main result is that the estimates of β_1 are remarkably robust to including the student interactions with teacher characteristics that vary by teacher gender.¹⁰ In all cases, the estimate of the gain to a female student from switching to a female teacher ($\beta_1 + \beta_3$) is positive and significant (ranging from 0.03 to 0.04 σ /year), and so is the estimate of the overall gain to a classroom ($\lambda_g * \beta_1 + \beta_3$) from having a female instead of a male teacher (ranging from 0.015 to 0.025 σ /year). The results in Panel B show that the figures are even more consistent (and always significant) when controls for other teacher characteristics are included. The range of the magnitudes is much tighter with ($\beta_1 + \beta_3$) mostly being 0.031 σ /year and ($\lambda_g * \beta_1 + \beta_3$) *always* being 0.015 σ /year.

Table 7 breaks down the results by subject (Panels A and B) and also conducts formal tests of equality across subject for the key quantities of interest (Panel C). Comparing ($\beta_1 + \beta_3$) across subjects suggests that the gains to girls from having a female teacher are higher in math. Finally, comparing the total social gains of shifting from a male to a female teacher ($\lambda_g * \beta_1 + \beta_3$) across subjects, we see that the gains in math are significantly larger than those in language (Panel C - columns 5 and 6). Further, once we control for teacher characteristics, all the gains in Column 6 of Table 5 can be attributed to the better performance of female teachers in math (where female teachers do much better with girls and no worse with boys) with the net effects in language being close to zero (positive for girls and negative for boys).

We also study the impact of a teacher-student gender match on student attendance. We find no significant effect of a gender-match on student attendance (Table 8). We do find that female teachers are slightly more effective at increasing attendance overall (by around 0.6 percent), but

¹⁰ In the interest of space, we only show these results for characteristics that are significantly different across teacher gender (see Table 1 - Panel B). The estimate of β_1 is unchanged and significant for interactions with other teacher characteristics (such as religion and caste) as well.

there is no differential impact by student gender. This result is interesting because the rhetoric of hiring female teachers is often based on the belief that having female teachers increases the safety and comfort of girls in school, and that their presence therefore encourages girls to attend school. Our results suggest however, that the mechanism for the positive impact of a gender match on test scores is less likely to be due to effects on the extensive margin of school participation, but more due to the increased effectiveness of classroom transactions between teachers and students.

Of course, this result could be reflecting a scenario where total primary school enrollment for both boys and girls is over 98% (Pratham 2012) and the role of female teachers in increasing attendance of female students may be more limited in such a setting. Nevertheless, our results suggest that even after achieving gender parity in school enrollment, there may be continued benefits to a policy of preferred hiring of female teachers due to their greater overall effectiveness in improving learning outcomes, and specifically due to their effectiveness in reducing gender gaps in test scores.

Finally, we calculate what proportion of the growing gender gap calculated in Table 2 can be attributed to girls being less likely to have a female teacher as they advance through primary school. Regressing the probability of a female teacher on the grade taught (with school fixed effects), we find that there is a 4 percentage point reduction in the probability of a student having a female teacher at each higher grade. Multiplying the reduced probability of a female teacher by the cost to girls of not having a female teacher in a given year ($\beta_1 + \beta_3$), and dividing this by the total annual increase in the test score gender gap (estimated in Table 2), we estimate that the reduced likelihood of female teachers in higher grades accounts for 9% of the annual growth in the gender gap in math and 21% in language (the fraction of the growing gender gap in language that is accounted for by this channel is higher than in math because the absolute magnitude of the annual growth in the gender gap is lower in language). Using estimates without school fixed effects, these figures would be 8% and 15% respectively (because the overall trend in the gender gap is slightly larger without school fixed effects - see Table 2).

5. Conclusion

We study gender gaps in primary school learning outcomes in a low-income setting using one of the richest datasets on primary education in a developing country. We find that at the

start of primary school, girls have a slight advantage in the local language (approximately $.05\sigma$) and are at par in math. However, girls lose this advantage in both language (by $0.01\sigma/\text{year}$) and in math (by $0.02\sigma/\text{year}$) as they progress through the schooling system.

While these trends likely reflect a broad set of household, school, and social factors, one specific school-level policy that has been posited as a promising channel for mitigating these trends is the greater use of female teachers in low-income settings. This is a policy that has been widely recommended and adopted, but there has been very little well-identified evidence to support this claim. In this paper, we present some of the first well-identified empirical tests of this hypothesis in a low-income setting, using an extremely rich data set collected annually over five years in the Indian state of Andhra Pradesh.

Our results suggest that female (and male) teachers are relatively more effective when teaching to their own gender, that learning for girls increases when they are taught by female teachers relative to male teachers, and that boys do not suffer adverse effects when taught by female teachers relative to male teachers, even when controlling for teacher observables. These results differ across subjects, and the value to girls of having a teacher of the same gender is greater in math than in language. One possible explanation for this could be that boys and girls face different stereotypes in math and language and that shared teacher gender matters more in areas with negative stereotypes, such as a stereotype that girls are less good at math).

From a policy perspective, our estimates suggest that expanding the hiring of female teachers - both at the margin of the current patterns of hiring (assuming that the marginal female teacher hired has the same characteristics as the average female teacher), and also when holding other characteristics constant, would improve overall learning outcomes and be especially useful as a tool for bridging gender gaps in learning trajectories over time. While we find evidence to suggest that the mechanism of impact is through more effective classroom interactions (as opposed to increased teacher-student contact time), our data does not allow us to explore the further granularity of the specific mechanisms through which shared gender may influence learning (such as role model effects, greater empathy, and closer identification between teachers and students of the same gender). Decomposing the reduced form effects further could help in crafting more nuanced policies to capture these positive gains without having adverse effects on either gender.

References

- BETTINGER, E. P., and B. T. LONG (2005): "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *American Economic Review*, 95, 152-157.
- BHARADWAJ, P., G. D. GIORGI, D. HANSEN, and C. NEILSON (2012): "The Gender Gap in Mathematics: Evidence from Low and Middle Income Countries," NBER Working Paper 18464.
- CARRELL, S., M. PAGE, and J. WEST (2010): "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *Quarterly Journal of Economics*, 125, 1101-1144.
- CARRINGTON, B., P. TYMMS, and C. MERRELL (2008): "Role Model, School Improvement and the Gender Gap - Do Men Bring out the Best in Boys and Women the Best in Girls?," *British Educational Research Journal*, 34.
- 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.
- DRIESSEN, G. (2007): "The Feminization of Primary Education: Effects of Teachers' Sex on Pupil Achievement, Attitudes and Behavior," *Review of Education*, 53, 183-203.
- EHRENBERG, R. G., D. D. GOLDBERGER, and D. J. BREWER (1995): "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988," *Industrial and Labor Relations Review*, 48, 547-561.
- FAIRLIE, R., F. HOFFMAN, and P. OREOPOULOS (2011): "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom ": NBER Working Paper 17381.
- FRYER, R. G., and S. D. LEVITT (2010): "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2, 210-40.
- GOI (2008): "Eleventh Five Year Plan 2007 - 2012," New Delhi: Oxford University Press.
- HAUSMANN, R., L. D. TYSON, and S. ZAHIDI (2012): "The Global Gender Gap Report 2012," *World Economic Forum*.
- HERZ, B., and G. B. SPERLING (2004): "What Works in Girls' Education Evidence and Policies in the Developing World. ," USA: Council on Foreign Relations.
- HOFFMAN, F., and P. OREOPOULOS (2009): "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources*, 44.
- HOLMLUND, H., and K. SUND (2008): "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?," *Labor Economics*, 15.
- LAHELMA, E. (2000): "Lack of Male Teachers: A Problem for Students or Teachers?," *Pedagogy, Culture and Society*, 8, 173-86.
- MARSH, H. W., A. J. MARTIN, and J. H. S. CHEND (2008): "A Multilevel Perspective on Gender in Classroom Motivation and Climate: Potential Benefits of Male Teachers for Boys?," *Journal of Educational Psychology*, 100, 78-95.
- MURALIDHARAN, K., and N. PRAKASH (2013): "Cycling to School: Increasing Secondary School Enrollment for Girls in India," NBER Working Paper 19305.
- MURALIDHARAN, K., and V. SUNDARARAMAN (2011): "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119.
- NCES (2011): "National Assessment of Educational Progress (Naep) Data Explorer. ," NCES.

- NEUGEBAUER, M., M. HELBIG, and A. LANDMANN (2011): "Unmasking the Myth of the Same-Sex Teacher Advantage," *European Sociological Review*, 27, 669-689.
- NIXON, L., and M. ROBINSON (1999): "The Educational Attainment of Young Women: Role Model Effects of Female High School Faculty," *Demography*, 36.
- OECD (2010): "Pisa 2009 Results: What Students Know and Can Do -- Student Performance in Reading, Mathematics and Science (Volume 1)".
- OUAZAD, A., and L. PAGE (2012): *Students' Perceptions of Teacher Biases: Experimental Economics in Schools*. London, UK: Center for the Economics of Education. London School of Economics.
- PRATHAM (2010): *Annual Status of Education Report*.
- (2012): *Annual Status of Education Report*.
- RAWAL, S., and G. KINGDON (2010): "Akin to My Teacher: Does Caste, Religious, or Gender Distance between Student and Teacher Matter? Some Evidence from India," London: Institute of Education, University of London.
- REHMAN, N.-U. (2008): "Yemen Makes Progress in Girls' Education with Unicef - Supported Literacy Programmes", http://www.unicef.org/education/yemen_43424.html.
- SLAVIN, P. (2006): "Liberia Launches Girls' Education National Policy with Support from Unicef" At a glance: Liberia, http://www.unicef.org/education/liberia_33458.html.
- STACKI, S. (2002): "Women Teachers Empowered in India: Teacher Training through a Gender Lens," UNICEF.
- UN (2012): "We Can End Poverty 2015 Millenium Development Goals."
- UNDG (2010): "Thematic Paper on Mdg3: Promote Gender Equality and Empower Women," United Nations Development Group.
- UNESCO (2010): "Institute for Statistics."
- (2012): "Enrolment and Gender Trends: Primary Education," UNESCO.
- WINTERS, M. A., R. C. HAIGHT, T. T. SWAIM, and K. A. PICKERING (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.

Table 1: Summary Statistics by Gender

Panel A: Students					
	No. Obs.	Mean	Male	Female	Female - Male
Female	94599	0.509			
Literate Father	66511	0.592	0.582	0.600	0.0185***
Literate Mother	66827	0.439	0.429	0.449	0.0199***
Proper House	66851	0.311	0.306	0.315	0.00981***
Has Toilet	66974	0.289	0.284	0.294	0.0106***
Panel B: Teachers					
	No. Obs.	Mean	Male	Female	Female - Male
Female	2680	0.457			
Head Teacher	2680	0.288	0.377	0.182	-0.195***
Regular Teacher	2680	0.503	0.497	0.511	0.0141
Contract Teacher	2680	0.188	0.116	0.273	0.157***
Completed Education: 12th Pass	2680	0.931	0.962	0.893	-0.0696***
Completed Education: Masters	2680	0.226	0.270	0.174	-0.0964***
Has Teacher Training	2661	0.833	0.909	0.743	-0.166***
Native to Village	2679	0.234	0.175	0.304	0.128***
Married	2676	0.810	0.845	0.769	-0.0762***
Active in Union	2674	0.183	0.276	0.074	-0.202***
Salary (monthly)	2674	9560	10697	8209	-2487.5***
Age	2660	36.905	39.542	33.750	-5.791***
Years Experience	2285	12.953	14.465	11.076	-3.389***
Teacher Absence	2666	0.191	0.197	0.184	-0.0135**
Multigrade Classroom	2680	0.458	0.475	0.437	-0.0386**
Classroom Enrollment	2680	23.225	22.869	23.647	0.778

Notes: (1) All variables are binary indicators, except for salary which ranges from 300 to 38400 (with a standard deviation of 5776), age which ranges from 12 to 58 (with a standard deviation of 9.76), and years of experience which ranges from 1 to 42 (with a standard deviation of 7.94). (2) Significance levels are as follows: *10%, **5%, and ***1%.

Table 2 - Learning Gaps by Gender and Grade**Panel A: Gender Differentials in Test Scores by Grade**

Dependent Variables: Normalized Test Score (Within Grade)

	Pooled Across Subjects			Math			Telugu		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female (Grade 1)	0.0279** (0.0122)	0.0279** (0.0122)	0.0207** (0.00925)	0.00238 (0.0135)	0.00237 (0.0135)	-0.00377 (0.0101)	0.0531*** (0.0127)	0.0531*** (0.0127)	0.0448*** (0.0100)
No. of observations	66660	66660	66660	33187	33187	33187	33473	33473	33473
Female (Grade 2)	0.00526 (0.0114)	0.00507 (0.0114)	0.00580 (0.00828)	-0.0271** (0.0117)	-0.0273** (0.0117)	-0.0241*** (0.00881)	0.0376*** (0.0122)	0.0374*** (0.0121)	0.0356*** (0.00889)
No. of observations	70953	70953	70953	35453	35453	35453	35500	35500	35500
Female (Grade 3)	-0.0217* (0.0118)	-0.0217* (0.0117)	-0.0225*** (0.00813)	-0.0569*** (0.0120)	-0.0570*** (0.0119)	-0.0572*** (0.00863)	0.0136 (0.0128)	0.0135 (0.0127)	0.0122 (0.00894)
No. of observations	74715	74715	74715	37349	37349	37349	37366	37366	37366
Female (Grade 4)	-0.0442*** (0.0120)	-0.0444*** (0.0120)	-0.0375*** (0.00770)	-0.0956*** (0.0122)	-0.0957*** (0.0121)	-0.0876*** (0.00815)	0.00709 (0.0130)	0.00698 (0.0130)	0.0126 (0.00864)
No. of observations	79972	79972	79972	39973	39973	39973	39999	39999	39999
Female (Grade 5)	-0.0262** (0.0115)	-0.0263** (0.0115)	-0.0206*** (0.00738)	-0.0749*** (0.0123)	-0.0750*** (0.0123)	-0.0669*** (0.00771)	0.0225* (0.0123)	0.0224* (0.0123)	0.0256*** (0.00846)
No. of observations	85572	85572	85572	42777	42777	42777	42795	42795	42795

Panel B: Trends in Gender Differentials in Test Scores from Lower to Higher Grades

Female	0.0311** (0.0132)	0.0311** (0.0132)	0.0271*** (0.00993)	0.0115 (0.0142)	0.0115 (0.0141)	0.00814 (0.0106)	0.0506*** (0.0139)	0.0505*** (0.0139)	0.0458*** (0.0107)
Female*Grade	-0.0144*** (0.00383)	-0.0144*** (0.00382)	-0.0126*** (0.00281)	-0.0207*** (0.00410)	-0.0207*** (0.00409)	-0.0189*** (0.00298)	-0.00805* (0.00410)	-0.00803* (0.00410)	-0.00631** (0.00308)
No. of Observations	377872	377872	377872	188739	188739	188739	189133	189133	189133
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes

Notes: (1) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including school fixed effects, and are clustered at the student level for OLS regressions including school fixed effects. (2) Significance levels are as follows: *10%, **5%, and ***1%.

Table 3: Characteristics of Classrooms Assigned to Female Teachers

	Dependent Variable: Classroom Has a Female Teacher					
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of Female Students	-0.000673 (0.0338)	-0.0121 (0.0323)	0.00109 (0.0209)	-0.0103 (0.0204)	0.00523 (0.0205)	-0.00914 (0.0204)
Grade 1	0.0156 (0.0163)	0.0641*** (0.0153)	0.0243** (0.0121)	0.0658*** (0.0116)		
Grade 2	0.0228 (0.0150)	0.0491*** (0.0141)	0.0278** (0.0121)	0.0460*** (0.0116)		
Grade 4	-0.0671*** (0.0150)	-0.0398*** (0.0146)	-0.0676*** (0.0123)	-0.0358*** (0.0118)		
Grade 5	-0.140*** (0.0170)	-0.0629*** (0.0163)	-0.134*** (0.0121)	-0.0539*** (0.0119)		
Test Score of Incoming Cohort of Male Students	-0.0142 (0.0228)	-0.0111 (0.0214)	-0.00149 (0.0134)	-0.000472 (0.0129)	-0.00906 (0.0136)	-0.00380 (0.0134)
Test Score of Incoming Cohort of Female Students	0.0189 (0.0191)	0.00683 (0.0188)	-0.00698 (0.0119)	-0.00476 (0.0113)	0.00172 (0.0124)	0.00571 (0.0119)
Number of Observations	10974	9641	10974	9641	10974	9641
Teacher Characteristics	No	Yes	No	Yes	No	Yes
School Fixed Effects	No	No	Yes	Yes	No	No
School*Grade Fixed Effects	No	No	No	No	Yes	Yes
Boys' Test Score = Girls' Test Score (p-value)	0.3168	0.5708	0.7932	0.8298	0.6117	0.6483

Notes: (1) "Teacher Characteristics" are salary, age, experience, teacher absence, class enrollment size and indicators for caste, teacher status, education, training, native to school location, marital status, union status, and a multigrade class. (2) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (3) Significance levels are as follows: *10%, **5%, and ***1%.

Table 4: Gender Differentials in Learning Trajectories from Lower to Higher Grades

	Dependent Variable: Normalized Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0251** (0.0120)	0.0255*** (0.00923)	0.00464 (0.0132)	0.00667 (0.0102)	0.0464*** (0.0125)	0.0453*** (0.00980)
Female*Grade	-0.00624* (0.00322)	-0.00725*** (0.00252)	-0.00830** (0.00368)	-0.0106*** (0.00281)	-0.00563* (0.00339)	-0.00562** (0.00271)
No. of Observations	304410	304410	151785	151785	152625	152625
School Fixed Effects	No	Yes	No	Yes	No	Yes

Notes: (1) Regressions include student's previous year's test score as an independent variable. (2) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including school fixed effects, and are clustered at the student level for OLS regressions including school fixed effects. (3) Significance levels are as follows: *10%, **5%, and ***1%.

Table 5: Impact of Female Teachers on the Learning Gains of Female Students (Pooled Across Math and Language)

	Dependent Variable: Normalized Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
(β_1) Female Student * Female Teacher	0.0383*** (0.00997)	0.0362*** (0.00788)	0.0354*** (0.00753)	0.0350*** (0.00792)	0.0343*** (0.00757)	0.0347*** (0.00804)
(β_2) Female Student	-0.0120* (0.00676)	-0.0140*** (0.00522)	-0.0126** (0.00498)			
(β_3) Female Teacher	-0.0154 (0.0188)	-0.00344 (0.00629)	0.000700 (0.00697)	0.00212 (0.00634)	0.00132 (0.00699)	-0.00305 (0.00805)
$\beta_1 + \beta_3$	0.023	0.033	0.036	0.037	0.036	0.032
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	1.575	30.113***	29.585***	37.954***	28.615***	16.722***
$\lambda_g * \beta_1 + \beta_3$	0.004	0.015	0.019	0.020	0.019	0.015
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	0.054	10.194***	10.913***	17.643***	10.944***	4.625**
Number of Observations	268548	268548	268548	268548	268548	235022
Teacher Characteristics	No	No	No	No	No	Yes
School Fixed Effects	No	Yes	No	Yes	No	No
School*Grade Fixed Effects	No	No	Yes	No	Yes	Yes
Grade Fixed Effects by Student Gender	No	No	No	Yes	Yes	Yes

Notes: (1) Regressions include student's previous year's test score as an independent variable. (2) "Teacher Characteristics" are salary, age, experience, teacher absence, class enrollment size and indicators for caste, teacher status, education, training, native to school location, marital status, union status, and a multigrade class. (3) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (4) Significance levels are as follows: *10%, **5%, and ***1%.

Table 6: Heterogeneous Effects on Test Score Gains of Girls by Teacher Characteristics and Teacher Gender

Dependent Variable: Normalized Test Scores												
Panel A: Excludes Additional Teacher Correlates												
Teacher Characteristic:	Head Teacher	Contract Teacher	Completed 12th	Completed Masters	Teacher training	Native to Village	Married	Active in Union	Salary	Experience	Absence	MG
(β_1) Female Student * Female Teacher	0.0312*** (0.00772)	0.0342*** (0.00765)	0.0351*** (0.00758)	0.0343*** (0.00759)	0.0346*** (0.00764)	0.0338*** (0.00759)	0.0333*** (0.00758)	0.0319*** (0.00779)	0.0360*** (0.00769)	0.0303*** (0.00816)	0.0352*** (0.00761)	0.0335*** (0.00758)
(β_3) Female Teacher	-0.00228 (0.00708)	0.00243 (0.00708)	-0.000149 (0.00700)	0.00247 (0.00701)	0.000203 (0.00709)	-0.00156 (0.00702)	-0.000250 (0.00699)	0.00834 (0.00717)	-0.00201 (0.00711)	-0.000133 (0.00776)	-0.00350 (0.00704)	0.000449 (0.00698)
(δ_1) Female Student *Characteristic	-0.0179** (0.00846)	0.000603 (0.0122)	0.0260 (0.0188)	-0.000566 (0.00862)	-0.00151 (0.0127)	0.00786 (0.0100)	-0.0121 (0.0111)	-0.0132 (0.00977)	0.000000770 (0.000000727)	-0.000939* (0.000536)	-0.00428 (0.0178)	-0.0295*** (0.00770)
(δ_3) Teacher Characteristic	-0.0185** (0.00743)	-0.00902 (0.0105)	-0.0423*** (0.0158)	0.0146* (0.00759)	0.0102 (0.0109)	0.00785 (0.00837)	-0.0308*** (0.00905)	0.0413*** (0.00806)	-0.00000177*** (0.000000632)	-0.00165*** (0.000513)	-0.0666*** (0.0142)	-0.0162** (0.00767)
$\beta_1 + \beta_3$	0.029	0.037	0.035	0.037	0.035	0.032	0.033	0.040	0.034	0.030	0.032	0.034
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	18.248***	29.589***	27.669***	30.250***	26.730***	23.338***	24.694***	34.710***	25.099***	16.541***	22.437***	26.115***
$\lambda_g * \beta_1 + \beta_3$	0.014	0.020	0.018	0.020	0.018	0.016	0.017	0.025	0.016	0.015	0.014	0.017
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	5.586**	11.894***	9.762***	12.220***	9.567***	7.547***	8.678***	17.867***	7.948***	5.717**	6.361**	9.532***
Number of Observations	268548	268548	268548	268548	267475	268482	268442	268318	268215	239932	264581	268264
Panel B: Includes Additional Teacher Correlates												
(β_1) Female Student * Female Teacher	0.0317*** (0.00818)	0.0335*** (0.00811)	0.0356*** (0.00805)	0.0344*** (0.00806)	0.0338*** (0.00809)	0.0343*** (0.00805)	0.0341*** (0.00805)	0.0319*** (0.00828)	0.0354*** (0.00814)	hort fixed effe hool fixed effe	0.0347*** (0.00804)	0.0334*** (0.00805)
(β_3) Female Teacher	-0.00148 (0.00809)	-0.00241 (0.00807)	-0.00361 (0.00806)	-0.00293 (0.00805)	-0.00254 (0.00807)	-0.00282 (0.00805)	-0.00274 (0.00805)	-0.00156 (0.00812)	-0.00342 (0.00807)	-0.00174 nake landscap	-0.00305 (0.00805)	-0.00249 (0.00805)
(δ_1) Female Student *Characteristic	-0.0197** (0.00885)	0.0169 (0.0143)	0.0372* (0.0218)	-0.00361 (0.00903)	-0.0146 (0.0150)	0.0118 (0.0111)	-0.0173 (0.0122)	-0.0141 (0.0102)	0.000000453 (0.000000783)	-0.000854 (0.000540)	-0.000218 (0.0191)	-0.0315*** (0.00815)
(δ_3) Teacher Characteristic	0.0972*** (0.0351)	0.0179 (0.0292)	-0.0449** (0.0219)	0.00870 (0.00828)	0.0455** (0.0213)	0.0259** (0.0112)	-0.0201* (0.0109)	0.0430*** (0.00862)	-0.000000291 (0.000000920)	0.000493 (0.000731)	-0.0711*** (0.0156)	-0.0172** (0.00843)
$\beta_1 + \beta_3$	0.030	0.031	0.032	0.031	0.031	0.031	0.031	0.030	0.032	0.030	0.032	0.031
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	14.997***	16.044***	17.045***	16.476***	16.191***	16.436***	16.346***	15.135***	16.866***	15.324***	16.639***	15.878***
$\lambda_g * \beta_1 + \beta_3$	0.015	0.015	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.014
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	4.592**	4.609**	4.527**	4.573**	4.605**	4.590**	4.587**	4.636**	4.561**	4.622**	4.577**	4.514**
Number of Observations	235022	235022	235022	235022	235022	235022	235022	235022	235022	235022	235022	235022

Notes: (1) Regressors include student's lagged normalized test score, school*grade fixed effects and grade fixed effects by student gender (Specification from Column 5 and 6 of Table 5). (2) "Teacher Characteristics" are salary, age, experience, teacher absence, class enrollment size and indicators for caste, teacher status, education, training, native to school location, marital status, union status, and a multigrade class. (3) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (4) Significance levels are as follows: *10%, **5%, and ***1%.

Table 7: Impact of Female Teachers on the Learning Gains of Female Students by Subject

	Dependent Variable: Normalized Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Math						
(β_1) Female Student * Female Teacher	0.0338*** (0.0111)	0.0335*** (0.00876)	0.0336*** (0.00843)	0.0312*** (0.00880)	0.0315*** (0.00847)	0.0302*** (0.00901)
(β_2) Female Student	-0.0374*** (0.00729)	-0.0428*** (0.00575)	-0.0408*** (0.00553)			
(β_3) Female Teacher	-0.0139 (0.0209)	0.00240 (0.00701)	0.00806 (0.00785)	0.00916 (0.00706)	0.00917 (0.00786)	0.00924 (0.00909)
$\beta_1 + \beta_3$	0.020	0.036	0.042	0.040	0.041	0.039
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	0.928	28.781***	30.823***	35.745***	29.364***	20.412***
$\lambda_g * \beta_1 + \beta_3$	0.003	0.019	0.025	0.025	0.025	0.025
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	0.027	13.635***	15.392***	22.120***	15.464***	10.257***
Number of Observations	133907	133907	133907	133907	133907	117205
Panel B: Language (Telugu)						
(β_1) Female Student * Female Teacher	0.0429*** (0.0104)	0.0393*** (0.00851)	0.0373*** (0.00819)	0.0385*** (0.00856)	0.0364*** (0.00824)	0.0392*** (0.00875)
(β_2) Female Student	0.00971 (0.00724)	0.0104* (0.00563)	0.0113** (0.00539)			
(β_3) Female Teacher	-0.0174 (0.0182)	-0.00858 (0.00687)	-0.00531 (0.00762)	-0.00389 (0.00692)	-0.00485 (0.00764)	-0.0140 (0.00880)
$\beta_1 + \beta_3$	0.026	0.031	0.032	0.035	0.032	0.025
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	2.130	22.243***	19.361***	27.818***	18.819***	8.821***
$\lambda_g * \beta_1 + \beta_3$	0.005	0.011	0.014	0.016	0.014	0.006
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	0.069	4.894**	4.830**	9.099***	4.848**	0.642
Number of Observations	134641	134641	134641	134641	134641	117817
Panel C: Subject Differences (Math - Language)						
$\beta_{1M} - \beta_{1L}$	-0.009	-0.006	-0.004	-0.007	-0.005	-0.009
χ^2 [$H_0: \beta_{1M} - \beta_{1L} = 0$]	1.30	0.66	0.29	1.06	0.50	1.50
$\beta_{3M} - \beta_{3L}$	0.004	0.011	0.013	0.013	0.014	0.023
χ^2 [$H_0: \beta_{3M} - \beta_{3L} = 0$]	0.11	3.55*	4.17**	4.92**	4.56**	9.26***
$(\beta_{1M} + \beta_{3M}) - (\beta_{1L} + \beta_{3L})$	-0.006	0.005	0.010	0.006	0.009	0.014
χ^2 [$H_0: (\beta_{1M} + \beta_{3M}) - (\beta_{1L} + \beta_{3L}) = 0$]	0.27	0.88	2.34	1.03	2.07	3.69*
$(\lambda_g * \beta_{1M} + \beta_{3M}) - (\lambda_g * \beta_{1L} + \beta_{3L})$	-0.001	0.008	0.011	0.009	0.012	0.019
χ^2 [$H_0: (\lambda_g * \beta_{1M} + \beta_{3M}) - (\lambda_g * \beta_{1L} + \beta_{3L}) = 0$]	0.02	3.22*	4.46**	4.21**	4.49**	8.10***
Teacher Characteristics	No	No	No	No	No	Yes
School Fixed Effects	No	Yes	No	Yes	No	No
School*Grade Fixed Effects	No	No	Yes	No	Yes	Yes
Grade Fixed Effects by Student Gender	No	No	No	Yes	Yes	Yes

Notes: (1) Regressions include student's previous year's test score as an independent variable. (2) "Teacher Characteristics" are salary, age, experience, teacher absence, class enrollment size and indicators for caste, teacher status, education, training, native to school location, marital status, union status, and a multigrade class. (3) Standard errors (in parantheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (4) Significance levels are as follows: *10%, **5%, and ***1%.

Table 8: Impact of Female Teachers on the Attendance of Female Students

	Dependent Variable: Student Attendance					
	(1)	(2)	(3)	(4)	(5)	(6)
(β_1) Female Student * Female Teacher	-0.00266 (0.00349)	-0.00380 (0.00306)	-0.00346 (0.00307)	-0.00326 (0.00308)	-0.00288 (0.00310)	-0.00178 (0.00333)
(β_2) Female Student	0.00759*** (0.00264)	0.00740*** (0.00207)	0.00679*** (0.00207)			
(β_3) Female Teacher	0.000113 (0.00461)	-0.00372 (0.00247)	0.00752*** (0.00287)	0.00375 (0.00250)	0.00720** (0.00288)	0.00587* (0.00336)
Number of Observations	148791	148791	148791	148791	148791	129890
Male Student with Male Teacher Mean	0.777	0.777	0.777	0.777	0.777	0.777
$\beta_1 + \beta_3$	-0.003	-0.008	0.004	0.000	0.004	0.004
F-statistic ($H_0: \beta_1 + \beta_3 = 0$)	0.346	10.317***	2.172	0.043	2.438	1.600
$\lambda_g * \beta_1 + \beta_3$	-0.002	-0.011	0.012	0.004	0.012	0.010
F-statistic ($H_0: \lambda_g * \beta_1 + \beta_3 = 0$)	0.087	9.142***	6.020**	1.268	5.949**	3.065*
Teacher Characteristics	No	No	No	No	No	Yes
School Fixed Effects	No	Yes	No	Yes	No	No
School*Grade Fixed Effects	No	No	Yes	No	Yes	Yes
Grade Fixed Effects by Student Gender	No	No	No	Yes	Yes	Yes

Notes: (1) Attendance is calculated as the average of the indicator of whether the student was present or not on the day of 2 to 6 visits per year. (2) "Teacher Characteristics" are salary, age, experience, teacher absence, class enrollment size and indicators for caste, teacher status, education, training, native to school location, marital status, union status, and a multigrade class. (3) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (4) Significance levels are as follows: *10%, **5%, and ***1%.

Appendix: Data and Attrition

The data used in this paper were collected over 5 school years from 2005-06 to 2009-10 from a representative sample of government-run rural primary schools in the Indian state of Andhra Pradesh (AP). Since primary school consists of grades 1 to 5, a total of nine cohorts of students are present in our data (with the oldest cohort being in grade 5 in Year 1 of the project, and the youngest cohort being in grade 1 in Year 5). Appendix Table 1 shows the number of student observations by grade and year in our core estimation sample.

Appendix Table 1: Estimating Sample by Year and Grade					
	Year 1	Year 2	Year 3	Year 4	Year 5
Grade 1	14011	13030	11332	11150	9194
Grade 2	10286	8021	8322	6778	6162
Grade 3	11496	10381	10372	9757	8276
Grade 4	14119	11430	10702	11010	9711
Grade 5	15415	14024	11801	11295	10473

For students in grades 2 through 5, the estimating sample includes only those who have a test score in the current grade/year *and* have a test score from the previous grade/year (which is needed to estimate value-added). For grade 1 students, we include all those who have a test score in Grade 1 and set the normalized lagged test score to zero since there is no previous test (the estimates in Tables 5-8 are unchanged even if we exclude Grade 1). For grades 2-5, field teams conducted two rounds of testing at the end of each year (the first test covered competencies from the previous year, while the second test covered current year competencies). Since student attendance rates are around 70% having two rounds of testing helps considerably with reducing attrition from the sample.¹ However, there is only one round of testing at the end of grade 1 (since there are no previous grade competencies to be covered). Thus, the grade 2 sample in any year is smaller than the other grades.²

¹ Student scores are first normalized with respect to each test and then averaged across the two tests, and so we have a valid normalized test score for any student who took at least one of the two tests.

² Grade 1 has the highest number of missing students in the end-line, but does not require a baseline; and grades 3-5 have the benefit of fewer missing data points since they are less likely to have missing test score data from the previous year (where there would have been 2 rounds of testing).

This sample is further limited to observations for which we have student gender data (97%) and for specifications that include the teacher characteristics, the sample is restricted to cases where teacher interviews were conducted (which is 88% of the sample conditional on having student test data and student and teacher gender data for the year).

Moving across a row in Appendix Table 1 (over years), we observe a reduction in student observations. This is because the share of private school enrollment is growing considerably in rural Andhra Pradesh (Pratham 2012) and fewer students are entering the public school system over time. Appendix Table 2 tests whether entering cohorts over time differ in relative ability by student gender. We find no differences in Grade 1 test scores over time by student gender suggesting that the ability of girls relative to boys is not changing over time for the later entering cohorts. Thus, our estimates of the gender gap or of the impact of students sharing a teacher's gender are unlikely to be affected by the changing cohort sizes and composition over time.

Appendix Table 2: Entering Cohorts by Gender		
Dependent Variable: Normalized Test Score		
	(1)	(2)
Female Student	-0.00135 (0.0261)	-0.0000674 (0.0210)
Year	-0.00532 (0.0120)	-0.00540 (0.00480)
Female Student * Year	0.0102 (0.00832)	0.00725 (0.00665)
Number of Observations	66660	66660
School Fixed Effects	No	Yes
Notes: (1) Sample limited to students in Grade 1. (2) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (3) Significance levels are as follows: *10%, **5%, and ***1%.		

We next review how attrition from the sample will affect our estimates and interpretation of the gender gap and the effect of 'gender matching'. Attrition is defined as the fraction of students in a given year who are in the potential estimation sample (which comprises of all students who have a valid test score for the previous year), but are not in the final sample because they were absent from the end of year test (i.e., have no recorded test score for the current year). Grade 1 students are not included in the attrition analysis because they do not have a test-score from the

previous year, and we therefore cannot define attrition for grade 1. As mentioned earlier, all the results in Tables 5-8 are robust to excluding grade 1.

From our analysis on student attendance (Table 8), we know that girls are less likely to be absent from school on any given school day. Similarly, we find that girls have lower attrition (of 3%) in the sample used for the value-added calculations (Appendix Table 3). But we also see that there is no effect of a student having the same gender as the teacher on the probability of attrition. Thus, our main estimates (presented in Tables 5-7) are unlikely to be biased due to the lower attrition of girls from our estimation sample. Furthermore, the differential attrition by student gender will only change our interpretation of the gender matching effect if the students who attrite are *differentially affected* by shared teacher gender, which is unlikely given the lack of any effect of gender matching on either student attendance (Table 8 - columns 5 and 6) or on the probability of taking an end of year test conditional on having taken the test at the end of the previous school year (Appendix Table 3 - columns 5 and 6).

	Dependent Variable: Indicator of Attrition					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Student * Female Teacher	-0.0067 (0.0053)	-0.0062 (0.0045)	-0.0087** (0.0044)	-0.0029 (0.0045)	-0.0054 (0.0044)	-0.0049 (0.0048)
Female Student	-0.0320*** (0.0038)	-0.0334*** (0.0030)	-0.0308*** (0.0029)			
Female Teacher	0.0089 (0.0069)	0.0156*** (0.0037)	0.0047 (0.0041)	-0.0055 (0.0037)	0.0030 (0.0041)	0.0050 (0.0047)
Number of Observations	131585	131585	131585	131585	131585	115592
Male Student Attrition Mean	0.227	0.227	0.227	0.227	0.227	0.227
Female Student Attrition Mean	0.193	0.193	0.193	0.193	0.193	0.193
Teacher Characteristics	No	No	No	No	No	Yes
School Fixed Effects	No	Yes	No	Yes	No	No
School*Grade Fixed Effects	No	No	Yes	No	Yes	Yes
Grade Fixed Effects by Student Gender	No	No	No	Yes	Yes	Yes

Notes: (1) Student Attrition is calculated as an indicator for being absent for the test in a given year and having taken the test the preceding year. (2) Grade 1 students are excluded because they do not have a test score prior to enrollment in school. (3) Year 1 students who drop out of the sample in the first year are excluded. (4) Standard errors (in parentheses) are clustered at the school level for OLS regressions not including fixed effects, and are clustered at the student level for OLS regressions including fixed effects. (5) Significance levels are as follows: *10%, **5%, and ***1%.