The Impact of Female Teachers on Female Students' Lifetime Well-Being*

David Card, UC Berkeley
Ciprian Domnisoru, Aalto University
Seth Sanders, Cornell University
Lowell Taylor, Carnegie Mellon University
Victoria Udalova, US Census Bureau

July 29, 2022

Abstract

It is widely believed that female students benefit from being taught by female teachers, particularly when those teachers serve as counter-stereotypical role models. We study education in rural areas of the US *circa* 1940—a setting in which there were few professional female exemplars other than teachers—and find that female students were more successful when their primary-school teachers were disproportionately female. Impacts are lifelong: female students taught by female teachers were more likely to move up the educational ladder by completing high school and attending college, and had higher lifetime family income and increased longevity.

Keywords: Female Teachers; Education and Gender Equality; Education and Mortality

*The Census Bureau reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY22-POP001-0069. We gratefully acknowledge support from Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD091134-01). Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the US Census Bureau. Similarly, the content does not necessarily represent the views of the NICHD or NIH.

[Kids] don't remember what you try to teach them. They remember what you are.
—Jim Hensen, It's Not Easy Being Green: And Other Things to Consider

Do female students benefit from being taught by female teachers? The recent literature offers surprisingly mixed evidence; some studies find that female teachers improve young women's educational outcomes (e.g., Lim and Meer, 2017, 2020; Paredes, 2014; Carrell et al., 2010; Porter and Serra, 2010), while others do not (e.g., Holmlund and Sund, 2010; Cho, 2012; Antecol et al., 2015). A close reading of the different studies, however, suggests that the effects of gender matching are context-specific. Female teachers appear to matter more in environments where female role models are largely absent, such as science and math classes, or in less-developed countries, where there are few professional women.[1] In such settings female teachers may help overcome the "... powerful and sticky stereotypes about gender-specific skills and gender-specific roles," noted by Bertrand (2020).

In this paper we study the impacts on young women of having female public school teachers in rural areas of the US *circa* 1940. At the time, there were vast gender disparities in US society: women were less likely to enroll in college than men, and college-educated women were far less likely than their male counterparts to hold visible professional positions. There were few high-profile female business leaders or elected politicians at either the national or local level.[2] Especially in the rural areas we study, it is plausible that for many primary-school girls, female teachers served in a "dual role" as both educators and salient female role models.

We focus on women born between 1922 and 1933. These women were on the leading edge of what Goldin (2014) has called "the quiet revolution"—a revolution that ultimately led to profound changes in US society. They were more likely than their mothers to attend college and far more likely to build professional careers. Some went on to break longstanding gender barriers in science, law, and politics.[3]

[1] Among recent studies focusing on developing countries, Lee et al. (2019) provide evidence for Francophone Africa; Muralidharan and Sheth (2016) for India; and Gong, Lu, and Song (2018) for China.

A related literature looks at the benefits of matching the race/ethnicity of students and teachers. Redding (2019) provides a systematic review, and Gershensen et al. (2021) present a carefully designed recent analysis using data from Project Star.

[2] No Fortune 500 company was led by a woman until 1972. In the 76th US Congress, 1939–1941, only eight of the 435 seats in the House and one of 96 seats in the Senate were held by women—not surprising given that women had been granted the right to vote only 20 years earlier.

[3] Examples include Elinor Ostrom (born 1933), the first woman to win the Nobel Prize in Economics; Sandra Day O'Connor (born 1930) and Ruth Bader Ginsburg (born 1933), the first women to serve on the US Supreme Court; Jane Byrne (born 1933), first woman elected mayor of a major US city; and Toni Morrison (born 1927), the only Black American to win the Nobel Prize in Literature.

We show that during this pivotal era, young women were substantially more successful, in terms of school enrollment and grade completion, when taught by female teachers. Positive impacts on educational outcomes are found even *within* households; female teachers improved the outcomes of sisters relative to their brothers. These effects are found for White children, and also for Black children in the segregated South.[4] For the larger nationwide sample of White women, we examine lifetime impacts and find that girls taught by female teachers were more likely to complete high school and attend college, and had higher lifetime family income and increased longevity.

Relative to the extant literature on teacher gender effects, our analysis has two key advantages. First, because we are working with complete-count 1940 Census data, we can control for a very rich set of contextual factors, including detailed characteristics of neighboring families and the education and income of local teachers—allowing *all* these factors to affect girls more or less than boys. In addition, our sample sizes allow us to study the impacts of female teachers on girls with poorly-educated parents—an interaction effect that is strongly suggested by a role model framework. Second, we study lifetime outcomes—completed education, household income, and longevity—in addition to contemporaneous educational process measures (progression in school), as is more typical in the literature. This is valuable because, as is now well known, the impact of one's early-life educational environment can be modest in terms of measured short-term outcomes (grades, test scores, etc.), but nonetheless important for later-life outcomes and behaviors.[5]

We take the following steps in developing our empirical evidence:

First, we construct a variable that measures the gender composition of likely teachers for each child—the local fraction female among teachers (FFT). The 1940 Census Bureau procedure asked enumerators to fill in "data sheets" residence-by-residence as they moved along streets within assigned "enumeration districts" (relatively small geographic areas with counties). We exploit this data structure to identify teachers who are close neighbors for each child—a strategy that makes sense in rural areas, but not in larger cities, where the gender mix of a child's neighboring teachers is less informative about the fraction female

[4]We focus on rural areas because our methods do not allow us to construct teacher gender measures in urban areas. We limit our study of Black students to the segregated South, because we cannot reliably construct teacher gender measures for these students outside of the South.

[5]This point is established in the work of James Heckman and co-authors. For instance, Heckman, Pinto, and Savelyev (2013) estimate positive impacts of the Perry Preschool program on several key lifetime outcomes (e.g., reduced criminal activity and increased employment) even though measured improvements in childhood cognitive tests were short lived. The program is shown to be particularly effective in the development of non-cognitive skills, including enhanced academic motivation for girls.

among teachers in the school attended by that child.

Of course we expect measurement error in our FFT variable. How large is that error, and how does it affect our models? We attack this issue with a detailed case study in Wisconsin, where school-level administrative data allow us to construct an alternative measure of teacher gender for each child. We find strong evidence that our FFT variable has useful empirical content: it is highly significantly correlated with the administrative measure, even within county. Using the two measures we can quantify the expected attenuation bias in the estimated effects of FFT on student outcomes from our various models.

Second, we estimate models of the relationship between the local fraction of female teachers and schooling outcomes for 15–18 year olds who were living with their parents at the time of the 1940 Census. We analyze White and Black families separately. We begin with "county models" that include county fixed effects; here we can estimate the impacts of local differences in FFT *within* counties on male and female students separately. Then we estimate comparable "enumeration district (ED) models" that feature ED-level fixed effects. Since our FFT measure is at the ED level, in these models we can only estimate the *relative effect* of more female teachers on female students compared to males students, i.e., a "differences in differences" style estimate that reflects how changes in FFT affect the relative educational outcomes of girls and boys in the same ED's. Fortunately, our county-level models suggest that the impacts of more female teachers on male students are small, and potentially zero for White students once we account for the impacts of measurement error. Finally, we narrow our focus further, estimating "household models" in which we compare the outcomes of sisters relative to their brothers.

In all three sets of models we include a detailed set of control variables, including child and family characteristics, county characteristics drawn from historical sources, and neighborhood socioeconomic features—variables we construct by averaging the characteristics of families contiguous to a child's housing unit. All these variables, including average education and average earnings of local teachers, are interacted with child gender, thereby allowing family and neighborhood factors to have differing impacts on male and female students. Thus, in the interaction space, *we are not privileging the effect of FFT*.

Third, we evaluate the lifetime impacts of being taught by female teachers. For these analyses we use a sample of White 7–14 year olds in 1940 who can be matched to responses in the 2000 Census (when the target group were age 67–74), and to national death records.

We find that exposure to female teachers substantially improved the educational outcomes of female students, both Black and White, in 1940. Our estimated impacts are highly robust: they stabilize quickly as we add a few key control variables, and then remain stable as we add more and more local contextual variables. They are also remarkably similar across our three main specifications (county-level, ED-level, and within-household). And they easily pass a falsification test derived by assigning opposite-race FFT measures to students in the segregated Southern states.

As for long-term impacts, we find that women who were taught by female teachers are more likely to complete high school and attend college. In old age, they have higher Social Security income—an indication of higher lifetime household income—and have increased longevity. Estimated long-reach impacts are quite large. For example, being taught by all-female teachers (rather than all-male teachers), increases the probability a woman attends college by about 7–8 percentage points (relative to a mean of 33 percent), increases household Social Security income by 6–9 percent, and extends longevity by about 0.5 years.

Our work makes three main contributions. First, we contribute to the literature on the impacts of teacher gender, providing unique evidence on the long-reach impacts of female teachers for female students in a setting with few professional female role models.[6] Second, we contribute to the broader literature on the determinants of educational success, and consequences of that success for lifetime prosperity and longevity. Third, we make a methodological contribution to the estimation of "neighborhood models" with historical US Census data, showing how information on the Census-taking process can be used to define neighbors at a very local level.

Research on teacher gender effects also has potential policy relevance. It informs the debate on merits of single-sex classrooms in which female students are matched with female teachers and male students with male teachers. It also provides guidance concerning the consequences for gender equality of altering the teacher gender mix in developing countries, particularly in places where stereotypical gender norms are strongly held.[7]

We proceed as follows: In Section 1 we set the stage, with an overview of extant research strategies used to study teacher-gender effects; in Section 2 we describe our data; in Sections 3 and 4 we set out research designs and report results; and in Section 5 we conclude.

[6] Research on teacher gender effects can be seen more broadly as contributing to our knowledge of gender norms and stereotypes, which are studied in a wide-ranging set of papers (see, for example, Fortin, 2005 and 2015; Bertrand, 2011 and 2020; and Fernandez, 2011).

[7] See, for example, Lee, et al. (2019), who argue that in Sub-Saharan Africa, an increase in the fraction female among teachers would reduce the gender gap in student achievement without harming boys.

1 The Role of Teacher Gender

The appropriate roles for male and female teachers, and the impacts of teacher gender on students, are old topics.[8] Publicly-supported schools in the US have employed both male and female teachers since their inception, but initially relied disproportionately on male teachers. In the late 19th century, school districts increasingly employed female teachers at the primary school level. Writing in 1935, Rhey Parsons noted, "One of the most interesting chapters in the history of education in the United States concerns the relative feminization of the teaching profession which has taken place since the Civil War. Before the war the typical public school teacher was a man; today the typical public school teacher is a woman" (page 89). Parsons indicates that after the War "hostility" to the employment of female teachers was common, but districts nonetheless increased the employment of female teachers primarily because women could be paid less.[9]

By the early 20th century, the teaching profession was majority female in many school districts, and some educators argued that boys might benefit from an increase in male teachers. For example, David Snedden, Commissioner of Education in Massachusetts, argued, "If the state is willing to pay the price, a certain proportion of men teachers should be assigned to departmental positions, not primarily because they are necessarily better teachers than women, but because it is desirable to introduce, in boys' classes at any rate, the influence of masculine personality" (Snedden, 1916).[10]

As an empirical matter, in 1940 rural America there was substantial variation across public school districts in the fraction female among primary and secondary teachers. Some of this variation was likely driven by district prosperity; it seems that more-prosperous districts were somewhat more likely to prioritize hiring men.

1.1 Empirical Approaches to Studying Impacts of Female Teachers

Over the past 30 years or so, a substantial body of scholarship has emerged on the impact of teacher gender on student achievement. We briefly overview empirical approaches used

[8] Historically, teachers were exclusively male in many societies, and education was often provided only to males. For example, the *New Testament* forbids women to teach (I Timothy 2:12), an injunction that presumably pertains to religious instruction.

[9] As Parsons (1935) puts it, "Even with the meager salaries paid, young women of ability were willing to prepare themselves for the profession of teaching, because it afforded practically the only employment opportunity open to them" (page 90).

[10] Cited in Parsons (1935).

in this research, as a way of placing our work in context.

In many previous papers on teacher gender effects, data are available at the classroom level. Researchers then typically adopt a baseline specification:

$$y_{ij} = \alpha_0 + \alpha_1 D_i + \alpha_2 F^T_{ij} + \alpha_3 D_i F^T_{ij} + e_{ij}, \tag{1}$$

where y_{ij} is an outcome variable for student i in classroom j, D_i is an indicator variable for student's gender (1 if the child is female, 0 for male),[11] and F^T_{ij} is a similar gender indicator variable for the teacher. We suppose here that y is constructed so as to be a favorable (rather than adverse) outcome. Outcomes studied in the literature include assessments by teachers (e.g., grades), standardized test scores or improvements in standardized test scores, and progression to advanced academic coursework.

In specification (1) α_2 is the impact on male students of being taught by a female teacher, and $\alpha_2 + \alpha_3$ is the impact on female students of being taught by a female teacher. Thus, for example, if female teachers are about as effective as male teachers for male students, but are more effective for female students, $\alpha_2 \approx 0$ and $\alpha_3 > 0$. Alternatively, if being matched to a same-gender teacher is beneficial to all students, $\alpha_2 < 0$ and $\alpha_2 + \alpha_3 > 0$.

Of course, there are many ways in which inferences based on regression (1) can go wrong, particularly when students (or students' parents) select classrooms or schools.[12] Researchers therefore often look for settings in which classroom/school assignment is random or quasi-random. Even with random student assignment, though, model (1) may be subject to omitted variable biases. For example, if female teachers in the study population differ from male teachers along some relevant characteristic (e.g., perhaps female teachers have less basic training in the subject of interest), at a minimum the OLS estimate $\hat{\alpha}_2$ will be biased.

The study by Carrell et al. (2010) of grades in STEM courses at the US Air Force Academy illustrates the use of (1) in a setting with random assignment. (Mansour et al. 2022 is essentially a longer term follow-up based on the same design). In their first set of analyses, the outcome variable is a normalized grade. For reasons we have just described, controls are added to regression (1)—a series of professor- and student-level variables. In

[11] We study children in families; D_i is natural notation for designating child i as a *daughter*.

[12] For example, suppose that parents prefer having their child matched to a teacher with the same gender, and suppose that relatively affluent families with academically-strong children are more effective in securing such matches. The result would be a sorting pattern in which academic outcomes would be stronger for students with same-gender teachers.

their various specifications (Table IV), $\hat{\alpha}_2$ is typically close to 0, while $\hat{\alpha}_3$ is positive and statistically significant. For typical female students, being assigned a female professor increases test scores by about 0.10σ, while for academically strong female students the effect is closer to 0.20σ.

A second nice illustration is Lim and Meer (2017), who study academic performance of students who are randomly assigned to classrooms in Korean middle schools. Here again, regression (1) is used, supplemented with a set of controls. Their main results (Table 3) are similar to those in Carrel et al. (2010): $\hat{\alpha}_2 \approx 0$ and $\hat{\alpha}_3 \approx 0.10\sigma$. In a follow-up study, Lim and Meer (2020) find that being taught by female teachers in seventh grade has persistent impacts on female students—improving attendance at STEM-focused high schools, and increasing aspirations for pursuing a STEM degree in college.

Carrell et al. (2010) likewise find persistence in impacts of female teachers on female students at the USAF Academy. When academically-strong female students are taught by female teachers in core STEM courses, their performance in subsequent STEM courses increases, as does their enrollment in higher-level mathematics courses. In drawing these inferences, the researchers employ a revised version of regression (1):

$$y_{ij} = \alpha_0 + \alpha_1 D_i + \alpha_2 F_{ij} + \alpha_3 D_i F_{ij} + e_{ij}, \tag{2}$$

where F_{ij} is the "fraction female" among among the professors a student encounters in core STEM coursework. When y_{ij} is an indicator variable for taking higher-level math coursework (among academically-strong students), $\hat{\alpha}_2 = -0.02$, while $\hat{\alpha}_3 = 0.19$. Taking core STEM courses from all-female teachers (rather than all-male), increases by about 17 percentage points the probability a female student studies advanced math. We note that the use of the "fraction female" variable as an independent regressor in specification (2) is the basic strategy we adopt in our analysis.[13]

Finally, researchers who use specifications (1) or (2) to study the impact of teacher gender often modify the design by including fixed effects. For instance, one popular design is a "within-class design," in which classroom fixed effects are included:

$$y_{ij} = \alpha_0 + \alpha_1 D_i + \alpha_3 D_i F^T_{ij} + \theta_j + e_{ij}. \tag{3}$$

[13]This approach appears in several other studies. For instance, Rothstein (1995), Robst, Keil and Russon (1998), and Price (2010) study the impact of "fraction female" among faculty at the college level on student outcomes, and Nixon and Robinson (1999) and Lindahl (2016) do the same at the high school level.

Here the "direct effect" of teacher gender—α_2 in regression (1)—is absorbed into the classroom fixed effect. Still, if the primary interest is α_3, the within-classroom design can be a compelling way to deal with teacher effects, or other classroom effects that are not controlled for. Carrell et al. (2010) and Lim and Meer (2017, 2020) implement within-classroom designs; in both cases, fixed-effect models reinforce results from the basic design.

1.2 Contexts in Which Female Students Benefit from Female Teachers

In Appendix A we list 24 papers published in the economics journals since 2000 that investigate whether teacher gender impacts outcomes differently for boys than for girls. Of these papers, 16 report generally-positive impacts on female students of being taught by female teachers, while the others typically find little or no impact. A key takeaway is that female teachers are most likely to improve outcomes for female students in contexts in which students have few female role models other than their teachers. For instance, positive impacts of female teachers on female students are often found in developing countries in which social and educational structures generally favor males. Similarly, in developed-country studies, female-teacher effects are often found specifically in STEM-related education—an educational environment in which women are often in a distinct minority.

Lee et al. (2019) provide a salient example. These authors analyze outcomes in ten Francophone African countries—a study area in which only 22% of teachers were female in 2014 (when their data were collected).[14] In this setting, the effect of female teachers on female students' standardized test scores is substantial. One-year exposure to a female teacher in 6th grade increases female students' reading scores by about 0.1σ and math scores by about 0.25σ. It appears that a sustained increase in exposure to female teachers would have substantial effects on the human capital development of girls in this region.

Muralidharan and Sheth (2016) similarly find advantages for female students from being taught by female teachers in the Indian state of Andhra Pradesh, where boys are generally more successful than girls in the educational system.[15] Gong, Lu, and Song (2018) likewise report gains in academic achievement for Chinese female students from being matched to female teachers—impacts that are particularly large for students whose mothers have relatively low education and for ethnic-minority students.[16] While Lim and Meer's (2017,

[14]In Sub-Saharan Africa, more broadly, female teachers are in the minority—less than 46% in primary schools and 32% in secondary schools—and educational outcomes lag for female youth. For instance, 55% of the "youth illiterate population" (aged 15–24) are female (UNESCO, 2022).

[15]The "youth illiterate population" in India as a whole was 55% female as of 2018 (UNESCO, 2022).

[16]Random assignment to a female teacher in middle school increases test scores by 0.27σ for girls whose

2000) research from South Korea is not in a developing-country setting, the authors suggest that female teachers may matter more there than in other developed countries because of persistent stereotypical gender norms that prevail in that country.[17]

In contrast, studies from the US, Chile, Netherlands, Sweden, Germany, and other developed countries often report little or no effect on student outcomes (see Appendix A). Instead, in developed countries evidence of positive impacts of female teachers on female students tend to appear specifically in STEM education.[18]

An ongoing stream of research studies social mechanisms that plausibly underlie the relative efficacy of female teachers for improving female student outcomes. This research focuses on the potential of female exemplars to break down gender-role stereotypes that undermine female students' educational success. The importance of this role model effect is magnified in settings in which such gendered stereotypes are widely and strongly held, which, naturally, tend be settings in which there are relatively few successful women.

A landmark study, from outside of the educational sphere, illustrates. Beaman, et al. (2009) evaluate impacts stemming from a "reservation policy" that increased the number of women in elected village-level political positions in India. The authors show that the policy created random variation across villages in female political leadership. Stronger female political leadership served to "weaken stereotypes about gender roles in the public and domestic spheres." In a remarkable follow-up study Beaman, et al. (2012) show that local female leadership also increased career aspirations among girls, and eliminated the gender educational gap. The authors argue that it was female leaders' "presence as positive role models for the younger generation that seems to underlie observed changes in aspirations and educational outcomes of adolescent girls" (page 586).[19]

Teachers surely play a key role in the social communication of stereotypes to students—stereotypes than can be detrimental specifically to female students. For instance, teachers may convey the stereotypical view that math is easier for boys than for girls, or that educational excellence is more important for boys than for girls. This idea is discussed and evaluated in the insightful work of Carlana (2019), who analyzes the impact on students

mothers have ≤ 9 years of education, and by 0.46σ for those who are ethnic minority. Also, female teacher effects are found to be particularly pronounced in math.

[17] In their study of single-sex schools, Park et al. (2013) make a similar point. They find lasting benefits to female students who are randomly assigned to single-sex schools in Seoul.

[18] Examples include the Carrell, et al. (2010) study and follow-up work by Mansour et al. (2022).

[19] An alternative hypothesis is that female political leaders devote increased resources to the education of girls relative to boys. The authors find no evidence favoring this latter hypothesis.

of exposure to stereotypes held by their teachers. In a study of Italian primary-school students and their teachers, she shows that when students are taught by math teachers with "stronger implicit stereotypes" (e.g., teachers who implicitly associate *male* with *scientific endeavors*), female students tend to exhibit less self-confidence in their math ability, perform less well on math standardized test scores, and subsequently select into less demanding high schools. In many settings female teachers may be much less likely than their male counterparts to communicate such stereotypes.

To the extent that female teachers are effective role models, who increase confidence and strengthen aspirations among female students, this in turn can have lasting consequences for gender differences in educational and occupational choice.[20] There is evidence, indeed, that even modest exposures to female role models can reduce stereotypical views and affect key educational choices.[21]

Naturally, female teachers can improve educational prospects for female students not only by serving as a salient role model, but also in other related ways, such as devoting more effort or attention to female students, or grading female students more favorably (see the work on "grading bias," e.g., Lavy and Sand, 2018, and Lavy and Megalokonomou, 2019). If these latter activities are more prevalent among female teachers—which the existing literature suggests is *not* the case—we would expect that as an empirical matter female students will perform better when taught by female teachers *and* male students will perform less well when taught by female teachers (see Paredes, 2014, for a careful discussion in a context with teacher value-added). In contrast, to the extent that female teachers reduce detrimental impacts of gender-related stereotypes, this improves educational progress for female students without harming educational progress among males. Empirical evidence is necessary to assess effects in specific contexts.

1.3 Our Historical Context

We study primary-school students living in rural America *circa* 1940. In 1940 the US was a wealthy country, with annual per capita (inflation-adjusted) income of more than $10,000. And along some dimensions, the US educational system provided relative gender equality.

[20] See, e.g., Coffman (2014), Buser, et al. (2014), Reuben, et al. (2015), and Kugler, et al. (2021) for evidence about the role of gender differences in confidence for educational and occupational choice.

[21] See, e.g., results from field experiments that increased the percentage of female college students enrolling in economics (Porter and Serra, 2020), and that affected STEM-related educational choices of high-ability female high school students (Breda, et al, 2021).

The majority of primary-school teachers were female, and rates of school attendance were similar for girls and boys. But at the time, American social structures were characterized by vast gender disparities: very few women held high-prestige professional jobs in law, medicine, management or higher education.

Especially in rural areas and small towns, there were very few college-educated professional women other than school teachers. 1940 Census records indicate that among rural White women aged 24–64, only 3.3% had attained at least one year of college *and* were in the labor force. As shown in Table 1, more than half of these women were teachers; the rest were typically employed in non-professional jobs (e.g., clerical or secretarial positions). In contrast, 7.7% of White men had some college and were employed, and a significant number were physicians, lawyers, and judges. As for Southern Black adults in rural areas, only 1.3% of women and only 1.2% of men had some college and were employed. A substantial majority of these Black women and about one third of these Black men were teachers.

For typical school-aged girls in rural America *circa* 1940, female teachers would surely have been the most important exemplars of female professional achievement, and would thus plausibly play a key role in elevating educational and career aspirations of their female students. If so, this is an excellent environment in which to pursue our research objectives: to investigate the impact of female teachers on female students, not only for schooling progress in childhood, but also for life-course outcomes. We have particular interest in effects for female students from disadvantaged households—families in which parents were poorly educated. These students would have been especially unlikely to have professional female role models other than their teacher, and would have had the most to gain from exposure to counter-stereotypical female role models in school. Given gender disparities in professional attainment that differ by race (as shown in Table 1), we are also interested in the possibility of racial differences in the role of teacher gender.

We have already noted one advantage to our focus on tangible lifetime outcomes: early-life educational interventions that have short-lived impacts on student educational metrics may nonetheless have important effects in adulthood (e.g., Heckman et al. 2013). In addition, by focusing on outcomes measured later in life—rather than process measures, such as grades or grade promotion—we sidestep some of the issues raised in the literature on gender-biased favoritism in grading.

2 Data

Our primary data sources are the 100 percent files of the 1940 US Census, and links from these files to 2000/2010 Census files and the Social Security Administration NUMIDENT.[22] We also use complementary historical data sources.

2.1 Target Study Samples

Our study uses two samples drawn from the 1940 Census: White children aged 7–14, and adolescents aged 15–18, both Black and White. Nearly all 7–14 year olds are in school in 1940, while many youth aged 15–18 are not. The latter sample then allows us to examine contemporaneous schooling outcomes, such as current school attendance and educational attainment. We study children living with at least one parent, where all parents present are US born.[23] We restrict our sample of Black youths to those living in the South, where school segregation allows us to more easily assign students to likely teachers (and where 83% of Black youths lived). Our 1940 sample of 7–14 year olds is merged to the 2000/2010 Census and the NUMIDENT using Protected Identification Keys (PIKs). When using the PIKed data we limit the sample to observations with a unique PIK match.

We focus first on White children and adolescents. Table 2 provides relevant summary statistics. The first three columns are for 15–18 year olds. Column (1) has statistics for the full sample, while column (2) reports statistics for the subset of youths who meet the following criterion: they live with parents, they are located in small towns or rural areas, their household has not moved over the previous five years, and they live in places where we are able to construct teacher data. Column (3) further restricts the sample to households that have at least one daughter and one son in the target age range—a subsample we use for our household models. Comparing columns (1) and (2) we see that those living in rural areas and small towns are slightly less likely to be female (because young women tend to leave home at younger ages than young men in rural areas), are more likely to live on a farm, and have parents who are slightly less well educated. A comparison of columns (2)

[22]We employ the version made available to researchers at the Census Bureau on 6/6/2018.

[23]As discussed in Card, Domnisoru, and Taylor (2022), the IPUMS version of the data includes a hierarchy of links between the records of children and potential mothers/fathers in the household. We use the so-called "unambiguous" link, and we therefore likely exclude a relatively small number of children who in fact were living with parents in multi-generational households. In 1940, nearly all children aged 7–14 remained in the household, but modest numbers left at ages 15–18, particularly among females. This creates some selection among remaining children aged 15–18; analyses are thus potentially subject to modest selection bias, as discussed in Card, Domnisoru, and Taylor (2022).

and (3) shows that individuals in the "household sample" are quite similar to other rural families, except, not surprisingly, in these families about half of 15–18 year-old children are daughters.

Columns (4) through (7) in Table 2 pertain to our sample of 7–14 year olds. Column (4) gives statistics for all individuals in this age range, while column (5) shows individuals who meet the additional criteria for inclusion in our sample. As was true for 15–18 year olds, individuals in our target sample of 7–14 years olds are relatively more likely to live on farms, and have parents who are relatively less well educated. Column (6) shows statistics for individuals in the target sample for whom we have a (unique) PIK. Average characteristics in the PIK sample are quite similar to those in the larger sample. We match individuals with PIKs to the 1-in-6 long-form data of the 2000 Census, resulting in a sample of approximately 400,000. Columns (6) and (7) show that matched individuals are very similar to those in the larger sample.[24]

2.2 Measuring Local Contextual Factors Using Census Data

The 1940 Census was conducted by enumerators who physically visited households. Enumerators typically recorded their entries residence-by-residence as they moved along streets within their assigned enumeration districts. Household and person information was recorded on worksheets that would fit information on 40 people. Page information identifying worksheets was recorded when the 1940 full count data was transcribed and is available to us. Our analysis suggests that individuals recorded on the same page almost always lived in close proximity, and that those on consecutive pages also typically lived nearby, so page number data is helpful for identifying individuals who are relatively "close by."[25] This approach extends the work of other scholars, including Logan and Parman (2017), who study neighborhood segregation, and Tan (2022), who studies social networks in neighborhoods.

Using nearby households we construct many neighborhood contextual factors (e.g., the average education of nearby adults, and the fraction of households residing on farms) which we use in our analyses. We also measure the gender composition of likely teachers for each child. This variable, fraction female of nearby teachers (FFT), is constructed as follows.

[24]Household models could be estimated using brothers and sisters who appear in the 2000 Census. However, we find that sample sizes are too small for precise estimation. The reason is that even in households with a brother and sister in the target cohorts, both of whom have a PIK and both of whom survive to 2000, there is a 1-in-6 match rate for individuals, but a 1-in-36 rate for the siblings combined.

[25]One exception is occasional "supplementary pages."

We begin by identifying teachers using the Census occupational code, restricting further to teachers aged 19–70 who are employed. For each teacher we observe gender (as well as other characteristics we use in analyses, including age, annual earnings, and education). Then, for each child i we construct an FFT_i variable. Our workhorse FFT construct is simply the fraction female of teachers in the enumeration district (ED).[26] We also construct FFT using those teachers within the ED who are "closest" based on data-sheet page and line numbers. Using this method, we construct FFT for the "closest k" teachers, for $k = 1$ through $k = 8$. These latter measures are used in robustness analyses, in which we explore the consequences of various choices made in constructing the FFT variable. Of course, measurement error in FFT is a major issue, which we discuss extensively below.

An additional concern is that families are not randomly assigned to locations. Instead, they choose where to live, and possibly sort based on factors related to our outcomes of interest. Our research design is to compare gender differences in educational outcomes across locations that vary in terms of FFT, but are similar on other dimensions. The simplest variant of this plan is a differences-in-differences design comparing female-male outcome differences in high-FFT versus low-FFT places. Such an analysis could give biased estimates of teacher-gender effects if exposure to female teachers is correlated with other characteristics. Suppose, for example, that parents prefer to have their sons educated by male teachers and their daughters by female teachers, but only well-educated/higher-income parents actively relocate when they see the gender of local teachers. Then we would be likely to find that on average boys do better in low-FFT places, while girls do better in high-FFT places, even absent causal teacher-gender effects. Notice that we would also likely be able to spot this problematic sorting pattern in the data, by looking at parental characteristics; for instance, we would expect that in low-FFT places boys' parents would be better educated than girls' parents, while the opposite would be true in high-FFT places.

As a simple check for evidence of such sorting, in Table 3 we compare household characteristics in high-FFT EDs versus low-FFT EDs for our primary sample: 7–14 year olds who are matched to 2000 Census records (those depicted in column 7 of Table 2). Columns (2) and (3) provide statistics for male and female children, respectively, in families that live in EDs with *below-median* FFT, while columns (4) and (5) provide comparable statistics for families that live in EDs with *above-median* FFT. There are some clear differences between the two sets of statistics; for example, families in low-FFT places are less rural (fewer fami-

[26]Rural EDs were to include no more than 1,500 people and 250 farms. (In contrast, modern Census tracts generally have a population size between 1,200 and 8,000 people, with an target size of about 4,000.)

lies live on farms and more families live in counties that contain a large city). Importantly, however, these patterns are the same for sons and daughters; differences in differences, in column (8), are very close to 0 and are not statistically significant. This observation holds also for mother's education, father's education, and single-mother household status.

In Panel B of Table 3 we go further, asking if there is differential sorting of sons and daughters into neighborhoods on the basis of "neighborhood contextual factors," as measured by characteristics of nearby households. Again, differences-in-differences checks do not expose any obvious problems; differences in our contextual factors between low-FFT places and high-FFT places are about the same for female and male children.

Finally, from Panel C we observe that FFT is about 0.7 overall, and averages 0.5 in places with below-median places, and 0.9 in above-median places. Clearly we have a fair amount of variation in FFT. Most importantly, there is no evidence that male and female children sort disproportionately into high- or low-FFT places.

While results in Table 3 do not expose obvious problems in terms of gender-based differential sorting, this issue remains a potential threat to validity. We seek to avoid omitted variable bias by estimating regression models that include extensive controls, and by using within-group designs at the ED and household levels.

3 Impact of Female Teachers on 1940 Schooling Outcomes

Our first set of analyses examine schooling-related outcomes for 15–18 year olds in 1940. We proceed with three study designs: a county design, an ED design, and a household design.

3.1 County Design

We begin with a regression specification common in the literature, equation (2), but including county fixed effects. We do not observe the actual FFT for each student; instead, for student i we use $F_i \equiv F_{d(i)}$, the FFT for enumeration district $d(i)$, in our regression.

Conceptually, it is reasonable to measure students' FFT at the ED level, because these are small contiguous areas; all students within an ED would typically be in the same school catchment areas. Nonetheless, the ED-level teacher gender mix is doubtless a fairly rough measure of the actual FFT in the schools that students attend. To see the consequence of

FFT measurement error, suppose the data generating process (true model) is

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 F^*_{d(i)} + \beta_3 D_i F^*_{d(i)} + \theta_{c(i)} + u_i, \tag{4}$$

where $F^*_{d(i)}$ is the actual FFT for student i in ED $d(i)$, and $\theta_{c(i)}$ is that student's county component. Next, suppose the true ED-level FFT relates to the county FFT as follows:

$$F^*_{d(i)} = F_{c(i)} + \xi_i, \text{ with } \xi_i \perp F_{c(i)}. \tag{5}$$

Also, suppose the measured FFT is related to the true FFT according to

$$F_{d(i)} = F^*_{d(i)} + \phi_i, \text{ with } \phi_i \perp [F_{c(i)}, F^*_{d(i)}]. \tag{6}$$

Then for individuals in county $c(i)$ we have

$$\begin{aligned} F_{d(i)} &= F_{c(i)} + \xi_i + \phi_i && (7) \\ &= F_{c(i)} + \eta_i. && (8) \end{aligned}$$

Note that the within-county variation in observed FFT has variance, $\sigma^2_\eta = \sigma^2_\xi + \sigma^2_\phi$, reflecting a combination of true "signal" (σ^2_ξ) and "noise" (σ^2_ϕ).

To simplify the discussion, assume that half of children are female, and that there is no systematic county-level variation in child gender: $\mathrm{E}[D_i|c(i)] = 0.5$ and $\mathrm{E}[D_i F_{d(i)}|c(i)] = 0.5F_{c(i)}$. Then, when we use $F_{d(i)}$ in place of $F^*_{d(i)}$ in our regression model (4), this is equivalent to deviating the observed variables from county means and running OLS, i.e., a simple OLS model of the form:

$$y_i = \delta_0 + \delta_1 D_i + \delta_2 \left[F_{d(i)} - F_{c(i)}\right] + \delta_3 \left[D_i F_{d(i)} - 0.5F_{c(i)}\right] + v_i. \tag{9}$$

In Appendix B we derive the probability limits of the coefficients in this model. As a baseline we consider the simple case in which $\beta_2 = 0$, i.e., variation in FFT affects female students but not male students (as is found by Carrel et al, 2010, and Meer and Lim, 2017).

In this case,

$$\text{plim}\ [\hat{\delta}_1] = \beta_1 + \frac{\beta_3 \sigma_\phi^2}{\text{var}[F_{d(i)}]}\mu_F \tag{10}$$

$$\text{plim}\ [\hat{\delta}_2] = -\frac{\beta_3}{2}\frac{\text{var}[F_{c(i)}]}{\text{var}[F_{d(i)}]} \times \frac{\sigma_\phi^2}{\sigma_\xi^2 + \sigma_\phi^2} \tag{11}$$

$$\text{plim}\ [\hat{\delta}_3] = \beta_3 \frac{\text{var}[F^*_{d(i)}]}{\text{var}[F_{d(i)}]}, \tag{12}$$

where μ_F is the overall county mean of fraction female teachers, $\text{E}[F_{c(i)}]$.

Despite the modest complexity of model (9), the coefficient $\hat{\delta}_3$ on the interaction between daughter and FFT—our key coefficient of interest—is attenuated relative to the true effect β_3 by a conventional measurement-error correction factor, reflecting the signal-to-total variance ratio in measured FFT. Offsetting this, the coefficient $\hat{\delta}_1$ on the dummy variable for a female student indicator is upward-biased relative to the true effect β_1 (assuming $\beta_3 > 0$) reflecting the "rotation" of the observed regression line around the mean of the data. Less obvious is the *negative* probability limit for the coefficient on the main effect for FFT, $\hat{\delta}_2$. (Recall we are assuming the true main effect is 0.) This bias—which will make it appear that female teachers lead to worse outcomes for boys—becomes larger, the larger is the variation in the county-level FFT relative to the total variation in measured FFT, and the larger is the "noise" component of the within-county variance in measured FFT.[27] We keep these qualitative biases in mind when we evaluate estimates. Also, expressions (11) and (12) will prove helpful for correcting our estimates for measurement-error bias.

Finally, in our actual estimating model we include many control variables, including dummies for the child's age and Hispanic status (imputed from last name); *parental and household characteristics* such as parental education, parental age, number of siblings, and farm status; *characteristics of neighbors*, including the average education of neighbor fathers and mother, average income of fathers, the employment rate of neighbor mothers, the average employment rates of neighbor males and females aged 19–24; *characteristics of nearby teachers,* which might plausibly be related to teacher effectiveness, e.g., average wage, age, and years of schooling; and *county variables,* which are related to county prosperity. All these controls are interacted with child gender.

[27] Reassuringly, in the special case of no measurement error—where $\sigma_\phi^2 = 0$ and $F_i^* = F_i$—our results show that OLS coefficients are unbiased.

Thus we estimate

$$y_i = \delta_0 + \delta_1 D_i + \delta_2 F_{d(i)} + \delta_3 D_i F_{d(i)} + \delta_4 Z_i + \delta_5 D_i Z_i + \theta_{c(i)} + v_i, \tag{13}$$

where Z_i is a vector of included child characteristics, parent characteristics, and other ecological factors, and $D_i Z_i$, interacts these variables also with child gender D_i.[28] This last feature is crucial to our set-up; in the interaction space we are not privileging FFT.

The inclusion of neighborhood characteristics is potentially important. Altonji and Mansfield (2018) demonstrate that in models in which households sort into neighborhoods on the basis of *unobserved* variables, the means of neighbors' *observed* characteristics can, under plausible assumptions, serve as a control function for dealing with bias.

We begin our analysis by estimating our county-model regression for White adolescents using public-use 1940 Census, i.e., the sample depicted in column (2) of Table 2. We are interested in estimates of the main effect of FFT ($\hat{\delta}_2$), and the interaction effect ($\hat{\delta}_3$) in models that have dependent variables related to schooling and labor-force activity variables. In each case our dependent variable is an indicator variable equal to 100 or 0, so a regression coefficient provides the percentage point change in the dependent variable associated with a unit change in the independent variable.

Our first dependent variable is "on track for 9th grade," which is set to 100 if the individual has 9 or more years of schooling, or is currently enrolled and has 7 or 8 years, and is set to 0 otherwise. Key coefficient estimates are in the first row of Table 4, columns (2) and (3). If we were to take our estimates at face value (ignoring measurement-error bias), being taught by all female teachers, rather than male teachers, reduces 9th grade attainment among boys by about one percentage point and increases 9th grade attainment for girls, relative to boys, by more than three percentage points.

We then try several other dependent variables: "On track for 8th grade" indicates that the child has 8 or more years of schooling, or is currently enrolled and has completed 6 or 7 years. As with "on track for 9th grade," the interaction effect is positive, and main effect is negative (and relatively small in absolute value). Similarly, being taught by female teachers increases the likelihood that female students remain enrolled in school, has little effect on working, and reduces the rate of inactivity ("not enrolled or working"). The estimated impacts on female students are notable, especially given that they are biased toward 0. By

[28] Appendix C provides the list of control variables, along with sources. Note that county variables would be fully absorbed in the county fixed effects were it not for interaction terms.

comparison the estimated main effects of FFT are quite small, especially given that they are biased away from 0.[29]

Appendix Table 13 provides estimated coefficients on control variables, for our headline regression ("on track for 9th grade"). Some are quite interesting. For example, the coefficient on "Hispanic" is -3.3 and on the interaction "Hispanic$\times D$" is -5.4; Hispanic children lagged in terms of educational attainment, especially daughters. As another example, children living with a single parent had lower educational attainment than those with two parents; the coefficient on "only mother present" is -6.2 and on "only father present" is -5.6. Interestingly, the interaction effect "mother present$\times D$" is positive (3.2) while the interaction effect "father present$\times D$" is negative (-0.5); apparently, the impact on education of being in a single-parent household depends on the gender of both the single parent and the affected child.

How important are the control variables to our inferences about FFT? When it comes to the key interaction effect, they not as important as we would have thought. As we have noted, male teachers were paid more than female teachers in 1940, so low-FFT locations were generally more prosperous than high-FFT locations. Many control variables relate to local prosperity, and if we omit these controls, FFT main-effect estimates are biased downward (quite aside from measurement-error bias). The same need not be true of the FFT interaction effect. To investigate, we estimate our model with only the limited child controls, and then ask what happens as we add controls. See Figure 1. As expected, in a model with no ecological controls, the main effect is quite negative, but as we progressively add control variables, estimates generally move much closer to 0. On the other hand, estimates of the interaction effect are quite stable—moving from -3.6 to -3.3 overall.

Finally, we re-estimated our models using the subset of the data for which we can assign a PIK, rather than the 100% files. Comparing the estimates in Table 4 (which are based on publicly available 1940 Census data) to those based on the subset of children who can be assigned a unique PIK (and are only available under restrictive conditions to researchers at a Census RDC) provides an indication of whether the PIK process is somehow related to local FFT. Fortunately, as shown in Appendix Table 15, the coefficient estimates are very similar for the two samples, suggesting that selection into the PIK process does not bias the estimates much.

[29]Expression (11) shows that if the true main effect of FFT on outcome y is 0, measurement-error bias results in main-effect estimates with an opposite sign of the interaction effect. This is what we generally observe in column (2).

3.2 Enumeration-District Design

Of course, even with our many controls, there may be omitted highly-local contextual variables that affect girls' educational success *and* that are correlated with FFT. A natural way to deal with this potential problem is to specify fixed effects at a very local level, i.e., at the ED level. In so doing, we parallel the literature that uses within-classroom designs.

ED-level effects have the same granularity as our FFT measure. We thus cannot include an FFT main effect in our regression, as it is absorbed by fixed effects. The assumed data generating process (true model) is now

$$y_i = \beta_0 + \beta_1 D_i + \beta_3 D_i F^*_{d(i)} + v_{d(i)} + e_i, \tag{14}$$

where $F^*_{d(i)}$ is the true FFT and $v_{d(i)}$ is the district component. We estimate this model by substituting observed district FFT, $F_{d(i)}$, for $F^*_{d(i)}$. We show in Appendix B that the effect of measurement error in observed FFT is the same as the effect on the female interaction term in our county design.

As in the county design, we also add a vector of ecological factors, each of which is interacted with child gender. The estimates of the coefficient of interest, β_3, are shown in column (4) of Table 4.

To begin, we note that estimates of the interaction effect from the county design (column 3) and ED design (column 4) are very similar, though for all of the outcome variables the interaction effect is slightly larger when we include ED effects. In column (5) we estimate our ED design with a subset of youth with relatively poorly-educated parents (eight years of schooling or less). Given our interpretation of the FFT effect on girls as a role model effect, we expect the impacts of FFT to be larger for girls from disadvantage families, and indeed they are. The impact of female teachers is especially large for girls with less-educated parents in terms of 9th grade attainment, i.e., completing at least 1 year of high school. We interpret these estimates as showing that for girls from lower SES families, female teachers are important in terms of upward educational mobility.

3.3 Household Design

Finally, we focus on relative outcomes of sisters and brothers within the same family. We note that biases caused by the selective choice of location by parents should be mitigated in this design, since the locations of girls and boys are balanced. Because we are comparing similarly-aged children who live in the same household, we assume they have the same FFT, still measured at the ED level. Now, however, our model has a family-level error component. The assumed data generating process is

$$y_i = \beta_0 + \beta_1 D_i + \beta_3 D_i F^*_{g(i)} + w_{g(i)} + e_i, \tag{15}$$

where sibling i lives in family $g(i)$. Again, we substitute observed FFT for its true value, which results in the same attenuation bias as in our other models (see Appendix B). Also, we continue to include our full list of ecological factors (interacted by child gender) as control variables.

The estimates of the key coefficient β_3, in columns (6) and (7) of Table 4, are very similar to those from the ED design, though somewhat less precise, reflecting the fact that we lose about $\frac{3}{4}$ of our sample when limiting attention to girls and boys in households with at least one girl and one boy between the ages of 15 and 18. As in the ED design, the impacts of female teachers on female students are estimated to be larger in families with relatively poorly-educated parents.

3.4 Measurement Error Corrections

Across our three models, we find that increases in FFT are associated with improved outcomes for female students relative to males. We know we are understating the effect sizes, due to errors-in-variables bias, and would like to get a sense of magnitude involved, i.e., would like to learn about the attenuation factor, $\text{var}[F^*_{d(i)}]/\text{var}[F_{d(i)}]$, highlighted in equation (12).

To proceed, we undertake a case study in which we evaluate teacher gender composition for students in Wisconsin using administrative records, the 1938/1939 *Wisconsin School Directory,* which include the names of all teachers in high schools and "graded schools" (schools with at least two teachers).[30] For nearly all teachers we can assign gender on the basis of names, and can construct an administrative records-based FFT variable for each student in a substantial subset of EDs (see Appendix B). We denote this measure, F^{A}_d.

Recall that with Census data we can construct FFT measures in a number of ways. Our baseline F_d measure (used in the regressions above) uses all teachers in the student's ED. Alternatively, for each student we construct FFT measures using only those teachers within

[30] There appear to be few similar publications for other states.

the ED who are "closest," based on the page/line numbers of Census data sheets—for the "closest k" teachers, from $k = 1$ through $k = 8$. We denote these measures, $F_d^1 \ldots, F_d^8$, respectively. The advantage of setting k to be very small (e.g., 1 or 2) is that included teachers are very likely to be teaching in the child's school; however, these small values of k guarantee that F_d^k will be a noisier measure of the average gender of a student's full set of teachers. As we increase k we of course measure FFT with increased precision as long as the marginal teacher teaches in the student's school. However, as k increases it becomes increasingly likely that the marginal teacher does *not* teach in the student's school.

With this in mind, we proceed as follows: First, using Census data for children living in rural Wisconsin, we estimate a series of models in which we regress F_d^{A} on our Census-based measure using the k nearest teachers, F_d^k. Assuming that any measurement errors in the administratively-based estimate of FFT are uncorrelated with the measurement errors in our Census-based measure, the coefficient on F_d^k (i.e., $\mathrm{cov}[F_d^k, F_d^{\mathrm{A}}]/\mathrm{var}[F_d^k]$) provides an estimate of the associated attenuation factor in using F_d^k as a measure of FFT. The estimates are shown in column (1) of Table 5. Clearly, all of the Census-based FFT measures have useful empirical content. As expected, the size of the attenuation factor is very small when we have only 1 teacher—an indication of severe attenuation bias—but it increases in magnitude as we add more teachers, up to a point. When we include the closest 7 or 8 teachers, the attenuation factor is about the same as with the simple ED-wide measure.

In Appendix B we lay out a simple model of the measurement-error process when k nearby teachers are used to estimate FFT, under the assumption that as k increases, the probability that the marginal teacher is actually a teacher of the focal student declines geometrically. Calibrating this model to best fit the pattern of coefficients in column (1) of Table 5, we show that it can match the estimated attenuation factors remarkably well, providing a simple interpretation of the pattern and increasing confidence in the plausibility of FFT estimates based on nearby teachers.

Armed with the Wisconsin-based estimated attenuation factors, we proceed to estimate our county-design for all children living *outside* Wisconsin—in a series of nine regressions, each of which uses a Census-based FFT measures. In these regressions the dependent variable is "on track for 9th grade" (the same as the dependent variable in row 1 of Table 6). The entries in column (2) of Table 5 provide estimates of the interaction effect, $\hat{\delta}_3$. As expected, the OLS estimate is smallest when we use only the closest single teacher to each student (i.e., set $k = 1$). Moreover, the estimates become progressively larger as we include more nearby teachers in our FFT measure. Estimates are very similar when we use F_d^6, F_d^7 or F_d^8, and these estimates are similar also to the estimate when we use F_d.[31] Then we divide entries in column (2) by the corresponding entries in column (1), giving us error-corrected interaction effect estimates in column (3), with standard errors calculated using the delta method. Inferences, based on attenuation-corrected estimates, are remarkably similar for FFT measures that use at least 5 teachers.

We repeat these steps for our household model, again using 9th grade attainment as the outcome of interest. As with the county model, we find that when we use the nearest 5+ teachers to construct FFT, error-corrected effects are similar regardless of the FFT construct used. Clearly, interaction effect estimates in Table 4 substantially understate actual impacts of FFT; error-corrected estimates are about 2.7 times as large.

Error-corrected estimates in Table 5 indicate that for a female student, being taught by all-female teachers (rather than all-male teachers) increases the probability of being "on track for 9th grade" by about 9 percentage points. When can apply our correction factor—multiplying estimates by 2.7—to other estimates in Table 4, we infer that female teachers have substantial and statistically significant effects on all outcomes we study, except "working at Census date." To put results into perspective and to facilitate comparison with effects in the literature (Appendix A), note that our error-corrected estimates suggest that being taught by female teachers increases "being on track for 9th grade" by 0.22σ overall, and by 0.28σ for those with less-educated parents.[32]

Our measurement-error model shows that the main effects in the county design, reported in column (2) of Table 4, will be biased away from 0, if the true effect is 0. We can use equation (11) to calibrate that the estimated effect will be -0.22 times the estimated interaction effect. Thus, when the dependent variable is "on track for 9th grade," we expect the estimated main effect to be about $-0.22 \times 3.26 = -0.72$, which isn't far off the actual estimate (-1.06). We conclude that the true main effect is likely very close to 0 for this regression, and also for other regressions in Table 4.

We have one final observation about the construction of our workhorse FFT mea-

[31] The estimate in the last row of column (2) corresponds to the interaction estimate reported in Table 4. These estimates differ modestly, which is to be expected given that the samples differ; the sample used for Table 5 excludes children from Wisconsin. Having said that, Table 11 in Appendix B shows that estimates are very similar if we use the Wisconsin data only; the effect of FFT on school progression is very similar in Wisconsin and the rest of the country.

[32] These effects are based on ED model estimates, columns (4) and (5) in Table 4, multiplied by 2.7. Note that for a binary dependent variable, $y = 0$ or 100, the standard deviation is $\sigma_y = \sqrt{\mu_y(100 - \mu_y)}$.

sure, $F_{d(i)}$. Given that the signal-to-noise ratio is very high when measuring FFT with one teacher only, we might reasonably have restricted our analysis to EDs with two or more teachers. As it turns out, only about 8% of children in our sample live in one-teacher EDs, so when we omit these individuals from the sample, this makes only a modest difference in estimates (see Appendix Table 16).

3.5 Results for Black Youth in 1940

We next evaluate impacts of female teachers on educational outcomes of Black students. We focus on the South, where segregation meant that Black students were educated almost exclusively by Black teachers, and White students by White teachers.

Table 6 gives summary statistics similar to those in Table 1. Column (1) provides statistics for all Black 15–18 year olds in the US, and column (2) narrows to Southern states. The sample is not much smaller when we restrict attention to the South, as 83% of Black adolescents resided there. Even so, there are differences in characteristics. For instance, parents in the South have lower levels of education. In column (3) we further narrow our sample to those who live in with parent(s) in rural ares, who have not moved in the past five years, and for whom we can construct suitable teacher data. In the case of Black adolescents, we find that 22% live in EDs with just one Black teacher, so to limit measurement error in FFT introduced by one-teacher EDs, we include only those living in EDs with two or more Black teachers.

Parents of the Black adolescents we study have very low levels of education, about three years less than among parents of White rural adolescents. Similarly, mean grade completion among Black adolescents is much lower than for rural White adolescents (6.1 versus 8.8 years of completed schooling). Nearly three quarters of our sample lives on a farm, and almost one in five live with a mother only. The household sample has characteristics are very similar to the overall sample.

Table 7 reports analyses which parallel those for White adolescents (Table 4). In our headline model, with a dependent variable "on track for 9th grade," we estimate interaction effects that are similar to corresponding estimates for White youth, at least in the county and ED models. Estimated effects in the household model are much less precise than for White youth (not surprisingly, given that the Black household sample is only about one tenth the size of the White sample). When we use the dependent variable "on track for 8th grade," we again estimate statistically significant interaction effects. For this model, the household-model estimate is similar to county- and ED-model estimates, but again inference is imprecise with the household model.

Notably, fewer Black youth than White youth are "on track" for 8th or 9th grade education, primarily because nearly half of Black adolescents are not enrolled in school. Thus, we try two other benchmarks of educational attainment—6th grade completion (achieved by 57% of our sample) and 8th grade completion (achieved by 28%). With these dependent variables, we again estimate large and statistically significant interaction effects. Estimated effect are similar for the full sample and for the sub-sample with less-educated parents, which is not surprising as virtually all Black parents fall in the latter category.

In the preceding section we showed that FFT interaction effect estimates are subject to serious attenuation bias due to measurement error in FFT. On the basis of a study of White children in Wisconsin, we find that to recover unbiased estimates of interaction effects we need to multiply OLS estimates by about 2.7. We have not conducted a comparable exercise for Black FFT, but we suspect that measurement error is, if anything, a larger problem for Black teachers, because there are typically fewer Black teachers than White teachers per ED. Thus, if we multiply interaction effect estimates in Table 7 by 2.7 we likely have conservative estimates. When we do so, we find that for a Black female student, being taught by all-female teachers (rather than male) increases the probability of being on track for 9th grade by about 8.5 percentage points relative to male students. This is a large effect, given that only 42% are on track for 9th grade completion. We estimate similarly large impacts for all educational benchmarks analyzed.

Main effect estimates for Black youth are larger in absolute value than for White youth. For White students, under a null hypothesis that a *true* main effect is 0, the *estimated* main effect is expected to be about -0.22 times the estimated interaction effect. If Black FFT is subject to the same measurement error process, the estimated main effect in the "9th grade attainment" regression should be about -0.7. Instead, it is -3.0, with a standard error of 1.1, suggesting that the true main effect is negative (though probably less than 3 in absolute value, and clearly much smaller than the true interaction effect, which is about 8.5). So, it appears that in the 1940 South, Black female students benefited substantially from being taught by Black female teachers, while Black male students benefited modestly from being taught by Black male teachers.

Overall, our findings are consistent with the literature on teacher role model effects, which shows that students tend to benefit from being taught by same-sex or same-race teachers in contexts in which there are otherwise few professional role models with whom the student identifies. This depiction accurately depicts female students we study, both White and Black. As for male students, White students would have had no shortage of professionals who might serve as exemplars; nearly all high-prestige local professionals (doctors, lawyers, professors, etc.) were White men, as were nearly all nationally prominent professionals. Black male students in the rural South, on the other hand, had little exposure to college-educated professionals of their same sex and race, other than teachers; it follows that that these students might benefit from being taught by Black male teachers.[33]

3.6 A Falsification Test Based on Segregated Schools

We have one additional set of analyses for 15–18 year olds in Southern states—a "falsification exercise" in which we assign Black students the FFT of nearby White teachers, and White students the FFT of Black teachers, while otherwise retaining race-specific control variables. Since education in the South was segregated, the relevant FFT for students of a given race is their own-race FFT. If significant effects appear for opposite-race FFT, this indicates potential omitted variable bias (e.g., unobserved local environmental factors, not entirely race-specific, which drive both female education and the employment of women as teachers). We restrict attention to ED's that contained at least two teachers of each race, thereby avoiding comparisons of local areas that were entirely of one race or the other. Table 8 reports the estimated interaction effects from the ED design using the correct own-race teacher FFT and the incorrect opposite-race FFT. Reassuringly, a comparison of columns (1) and (2) shows that for Black youth, only the same-race FFT affects student outcomes, while a comparison of columns (3) and (4) shows the same for White youth.

4 Empirical Analyses of Lifetime Outcomes

Our final set of results concern impacts of teacher gender on lifetime outcomes, measured in the 2000 Census. Here we study White children aged 7–14 in the 1940 US Census who meet our other study criteria, and who are matched to 2000 Census records (at which

[33]Gershenson et al. (2021) provide a thoughtful overview of research on the value of same-race teachers for minority students, and provide important new empirical evidence. In an evaluation of high school students in North Carolina, they find that the high school dropout rates are lower for economically-disadvantaged Black male students when they have Black teachers. They find suggestive evidence, moreover, that this beneficial impact is larger when the Black teachers are male.

time they are aged 67–74).[34] Of course, our research sample includes only those who are assigned a PIK. Fortunately, as demonstrated in Table 2, the PIK sample is reasonably similar to the broader sample (although it is somewhat more heavily female and more likely to have lived on a farm in 1940). We use both the county and ED designs for our analyses.

Table 9 reports effects of FFT on schooling and income. The first three sets of models evaluate educational outcomes reported in 2000. From the county models, we infer that FFT had little effect on men's education, but substantial positive impacts for women. Estimates of interaction effects are similar in the within-ED design, though typically a little larger (as we saw in Table 4). Impacts are larger yet for female students with less-educated parents in models that evaluate years of schooling and high-school completion, but not college attendance. When we adjust for measurement error, by multiplying coefficients by 2.7, we infer that being taught by all-female teachers (rather than male) in childhood increases schooling by 0.27 to 0.35 years, increases high school completion by as much as 6.5 percentage points (for daughters of less-educated parents), and increases college attendance by 6.5 to 8 percentage points. The latter effect is quite large, given that the college-attendance rate for women in our cohort of interest was only 33 percent.

We have many control variables in our regressions. As noted above, when we omit contextual controls, omitted variable bias will likely bias main effect estimates downward, and might bias interaction effect estimates as well. Panels A of Figure 2 illustrate the effects of including control variables in our county models, for college attendance (pictures for other education variables are very similar). As expected, the main effect is quite negative when we have no ecological controls, but moves closer to 0 as controls are added. Estimates of interaction effects are also affected; they decline by about half as we add controls. Of course, even with the large number of controls, we may have omitted variable bias. Fortunately, however, inferences are very similar in the ED model, which controls for highly local contextual factors in a general way (with ED fixed effects).

We similarly find evidence that being taught by female teachers increase female students' lifetime household income. The evidence comes from models in which the dependent variable is log Social Security Income—a good measure of lifetime income. We estimate two variants of our models, one of which includes allocated data and one that excludes allocated data. Inferences are quite similar in these two analyses. Once we correct estimates for measurement error (multiplying by 2.7), we infer that being taught by all-female

[34]We do not study lifetime impacts for the Black individuals because sample sizes are much smaller and PIK rates are much lower (giving rise to greater concerns about selection bias).

teachers increases income by about 6 to 9 percent in the overall sample. When we focus on daughters of less-educated parents, we estimate that this effect is about 12 percent. We explore also the possibility of effects on personal income, and while interaction effects are positive, they are not statistically significant results (perhaps not surprisingly given that we are studying income of individuals aged 67–74).

Our final set of analyses concern the impact of teacher gender on longevity. There are two obvious reasons to hypothesize that being taught by female teachers improves the longevity of female students relative to male students. First, as we have seen, being taught by female teachers increases women's educational attainment, and for many reasons increased education can improve longevity.[35] Second, female teachers are likely more effective than male teachers in providing girls with relevant health- and hygiene-related education and guidance, and this could have lasting impacts on health.

We provide evidence in Table 10. In the first model we evaluate the impact of FFT on survival to 2000. For this analysis we take the full sample of 7–14 year olds in 1940 for whom we have a PIK, and ask if that individual appears in the 100% files of the 2000 Census or in the records of deaths in 2000 or thereafter (through year 2019). Approximately 72% of individuals in these cohorts survived to 2000 (using this definition). In both the county design and the ED design we estimate a positive interaction effect: being taught by female teachers increased the longevity of women, relative to men in their same county or same ED. The effect is more pronounced for women born to less-educated parents. If we multiply by 2.7 to correct for attenuation bias, we infer that being taught by all-female teachers (rather than male) increases probability of survival to 2000 by about 1.5 percentage points overall and by about 2 percentage points for women with less-educated parents.

Next, we try this same exercise for survival to 2010 (when individuals would be aged 77–84). Here we find statistically significant interaction effects only for women with less-educated parents. Similarly, when we examine survival to 2010, conditional on survival to 2000, we estimate positive interaction effects that are not statistically significant.

Finally, we estimate models in which the dependent variable is "age at death" or "log age at death." We estimate positive statistically-significant interaction effects in both the county and ED models. Focusing on the age of death, the measurement-error corrected interaction effect is about $2.7 \times 0.19 \approx 0.5$ year for the full sample, and about 0.7 for women born to less-educated parents. Clearly, these are substantial effects.

[35] Lleras-Muney (2022) provides an up-to-date discussion of the literature on the topic.

For the “age of death” regression, we provide Panel B of Figure 2, which shows the impact of adding control variables on estimates in our county model. We find that main-effect estimates change substantially when we add household variables. The main reason is that our household control variables include “living on a farm,” which is associated with a higher age of death *and* is positively correlated with FFT. Once we add household variables interacted with child gender, estimates of both the main effect and interaction effect stabilize somewhat. Importantly, estimates of the interaction effect are nearly identical in the county design and the ED design, with its highly-local fixed effects.

5 Conclusion

We study child educational development in a very interesting historical setting. In rural US communities *circa* 1940 there were virtually no college-educated professional women other than female teachers. The young women in our analysis ultimately attained higher education in far larger numbers than their mothers or grandmothers, and entered a wider variety of professions, ushering in what Goldin (2014) has termed the quiet revolution. We show that some of this success was driven by exposure to female teachers. Being taught by female teachers significantly increased young women’s educational attainment as measured at the time of the 1940 Census. This is true for Black and White female students. At the same time, it appears that Black male students fared modestly better when matched to Black male teachers, while teacher gender was irrelevant to the educational progress of White male students.

Linking respondents of the 1940 Census forward to the 2000 Census and mortality records, we also estimate important long-reach impacts of female teachers on female students. Women who were taught by female teachers were ultimately more likely to complete high school and attend college. Also, they had higher lifetime family income, and experienced increased longevity. Our work thus fits into the growing literature on the importance of a child’s educational environment for lifelong well-being. There are doubtless many elements crucial to the formation of a nurturing educational environment. Our research is consistent with the body of evidence indicating that for many students one such element is being taught by teachers who serve as counter-stereotypical role models.

References

Alan, Sule, Seda Ertac, and Ipek Mumcu. 2018. "Gender Stereotypes in the Classroom and Effects on Achievement." *Review of Economics and Statistics* 100(5): 876-890.

Altonji, Joseph G., and Richard K. Mansfield. 2018. "Estimating Group Effects using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects." *American Economic Review* 108(10): 2902-2946.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2015. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics* 33(1): 63-89.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124(4): 1497-1540

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science* 335(6068): 582-586.

Bertrand, Marianne. 2011. "New Perspectives on Gender," in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics* Vol. 4B, Amsterdam: Elsevier Science, 1545-1592.

Bertrand, Marianne. 2020. "Gender in the Twenty-First Century." *AEA Papers and Proceedings,* 110: 1–24.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clementine Van Effenterre. 2021. "Do Female Role Models Reduce the Gender Gap in Science? Evidence from French High Schools." Paris School of Economics Working Paper (halshs-01713068).

Card, David, Ciprian Domnisoru, and Lowell Taylor. 2022. "The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility." *Journal of Labor Economics* 40(S1): S39-S95.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics* 134(3): 1163-1224

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125(3), 1101-1144.

Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *Quarterly Journal of Economics* 129: 1625-1660.

Fernandez, Raquel (2011). "Does Culture Matter?" in J. Benhabib, A. Bisin, and M. O. Jackson (eds.), *Handbook of Social Economics.* Amsterdam: Elsevier Science, 481-510.

Fortin, Nicole. 2005. "Gender Role Attitudes and the Labour Market Outcomes of Women across OECD Countries." *Oxford Review of Economic Policy* 21: 416-438.

Fortin, Nicole. 2015. "Gender Role Attitudes and Women's Labor Market Participation: Opting-Out, AIDS, and the Persistent Appeal of Housewifery" *Annals of Economics and Statistics* 117-118, Special Issue on the Economics of Gender: 379-401.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4): 1091-1119.

Gong, Jie, Yi Lu, and Hong Song. 2018. "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes." *Journal of Labor Economics* 36(3): 743-778.

Gershenson, Seth, Cassandra M. D. Hart, Joshua Hyman, Constance, Lindsay, and Nicholas W. Papageorge. 2021. "The Long-Run Impacts of Same-Race Teachers." NBER Working Paper 25254.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes" *American Economic Review* 103(6): 2052-2086.

Kugler, Adriana D., Catherine H. Tinsley, and Olga Ukhaneva. 2021. "Choice of Majors: Are Women Really Different from Men?" *Economics of Education Review* 81: 102079.

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92(10): 2083-2105.

Lavy, Victor, and Rigissa Megalokonomou. 2019. "Persistency in Teachers' Grading Bias and Effects on Longer-Term Outcomes: University Admissions Exams and Choice of Field of Study." NBER working paper 26021.

Lavy, Victor, and Edith Sand. 2018. "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases." *Journal of Public Economics* 167: 263-279.

Lee, Jieun, Dong-Eun Rhee, and Robert Rudolf. 2019. "Teacher Gender, Student Gender, and Primary School Achievement: Evidence from Ten Francophone African Countries." *The Journal of Development Studies* 55(4): 661-679.

Lim, Jaegeum, and Jonathan Meer. 2020. "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea." *Journal of Human Resources* 52(4): 979-997.

Lim, Jaegeum, and Jonathan Meer. 2020. "Persistent Effects of Teacher-Student Gender Matches." *Journal of Human Resources* 55(3): 809-835.

Lindahl, Erica. 2016. "Are Teacher Assessments Biased? Evidence from Sweden." *Education Economics* 24(2): 224-238.

Lleras-Muney, Adriana. 2022. "Education and Income Gradients in Longevity: The Role of Policy." *Canadian Journal of Economics/Revue canadienne d'economique* 55(1): 5-37.

Logan, Trevon D., and John M. Parman. 2017. "The National Rise in Residential Segregation." *Journal of Economic History* 77(1): 127-170.

Mansour, Hani, Daniel I. Rees, Bryson M. Rintala and Nathan N. Wozny. 2022. "The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students." *Industrial and Labor Relations Review* 75(3): 693-715.

Muralidharan, Karthik, and Ketki Sheth. 2016. "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers." *Journal of Human Resources* 51(2): 269-297.

Nixon, Lucia A., and Robinson, Michael D. 1999. "The Educational Attainment of Young women: Role Model Effects of Female High School Faculty." *Demography* 36(2): 185-194.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi. 2013. "Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools." *Demography* 50(2): 447-69.

Parsons, Rhey Boyd. 1935. "Men and Women Teachers in Tennessee." *Peabody Journal of Education* 13(2): 89-97.

Price, Joshua. 2010. "The Effect of Instructor Race and Gender on Student Persistence in STEM fields" *Economics of Education Review* 29(6): 901-910.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics* 12(3): 226-254.

Robst, John, Jack Keil, and Dean Russo. 1998. "The Effect of Gender Composition of Faculty on Student Retention." *Economics of Education Review* 17(4): 429-439.

Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review* 48(3): 515-530.

Snedden, David. 1916. "Reorganization of Education for Children from Twelve to Fourteen Years of Age." *Educational Administration and Supervision* September Issue.

Tan, Hui Ren. 2022. "How Widespread Are Social Network Effects? Evidence from the Early Twentieth-Century United States." *Journal of Labor Economics* 40(1): 187-237.

UNESCO Institute for Statistics. 2022. *Educational Statistics: Classroom Teachers and Academic Staff by Sex.* Accessed 2/18/2022.

Winters, Marcus A., Robert C. Haight, Thomas T. Swaim, and Katarzyna 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: Occupations of Working Adults with Some College Education, Rural US in 1940

White women		White men	
Percent who had some college and were employed	3.28	Percent who had some college and were employed	7.68
Percent in occupation:		Percent in occupation:	
Teachers	55.74	Farmers (owners and tenants)	14.47
Stenographers, typists, secretaries	5.89	Managers, officials, proprietors	14.05
Clerical and kindred workers	3.49	Teachers	10.67
Managers, officials, proprietors	3.46	Salesmen and sales clerks	5.15
Bookkeepers	3.24	Clerical and kindred workers	3.51
Salesmen, sales clerks	3.20	Clergymen	3.14
Nurses, professional	2.64	Physicians and surgeons	2.98
Farmers (owners and tenants)	1.83	Lawyers and judges	2.63
Musicians and music teachers	1.74	Bookkeepers	2.51
Operative and kindred workers	1.68	Operative and kindred workers	1.98
Black women in the South		**Black men in the South**	
Percent who had some college and were employed	1.30	Percent who had some college and were employed	1.23
Percent in occupation:		Percent in occupation:	
Teachers	61.94	Teachers	31.65
Private household workers	12.01	Farmers (owners and tenants)	19.68
Service workers, not private household	1.69	Clergymen	6.92
Stenographers, typists, secretaries	1.53	Laborers	6.43
Laundresses, private household	1.49	Farm laborers and wage workers	4.71
Operative and kindred workers	1.42	Professors and instructors	2.46
Nurses, professional	1.37	Managers, officials, proprietors	2.41
Professors and instructors	1.35	Mine operatives, laborers	1.82
Clerical and kindred workers	1.32	Farm laborers, family workers	1.77
Beauticians, manicurists	1.09	Private household workers	1.40

Notes: 1940 IPUMS data, individuals aged 24–64 living in rural areas and small towns (less than 1,000 inhabitants), with at least one year of college education.

Table 2: Characteristics of White Children and Adolescents in the 1940 Census

	Adolescents Aged 15–18			Children Aged 7–14			
		Non-movers, living with parent(s), in rural areas, with teacher data			Non-movers, living with parent(s), in rural areas, with teacher data		
	All (1)	All (2)	Household sample (3)	All (4)	All (5)	Has PIK (6)	Has PIK, in 2000 Census (7)
Fraction living with parent(s)	0.887	1.000	1.000	0.967	1.000	1.000	1.000
Fraction with mother only	0.110	0.094	0.085	0.083	0.066	0.056	0.049
Fraction female	0.499	0.469	0.498	0.492	0.490	0.475	0.514
Fraction enrolled at Census	0.652	0.651	0.628	0.935	0.923	0.939	0.947
Mean grade completed	9.136	8.849	8.688	4.238	4.131	4.154	4.104
Fraction in small place	0.551	1.000	1.000	0.567	1.000	1.000	1.000
Fraction living on farm	0.313	0.571	0.599	0.313	0.550	0.521	0.551
Fraction non-mover	0.750	1.000	1.000	0.765	1.000	1.000	1.000
Fraction with teacher data	0.906	1.000	1.000	0.902	1.000	1.000	1.000
Assigned PIK	–	–	–	–	–	1.000	1.000
Mean mother's education	8.407	7.861	7.680	8.650	8.101	8.342	8.505
Mean father's education	8.070	7.315	7.119	8.244	7.460	7.702	7.833
Observations*	6,714,346	2,326,299	534,995	12,971,012	4,940,564	3,000,000	403,000

Notes: Columns (1) through (5) are calculated using the 1940 IPUMS. Columns (6) and (7) use internal Census data. *Approximate sample sizes provided for columns (6) and (7).

Table 3: Characteristics of 1940 Census Enumeration Districts (EDs), White Households

		EDs with below-median FFT		EDs with above-median FFT		Mean differences, high − low FFT		Diff in diff
	All	Males	Females	Males	Females	Males	Females	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Household characteristics								
Mother's education	8.50	8.50	8.45	8.57	8.49	0.06	0.04	-0.03
						(0.03)	(0.03)	(0.04)
Father's education	7.83	7.91	7.87	7.82	7.73	-0.08	-0.13	-0.05
						(0.03)	(0.03)	(0.04)
Only mother present in household	0.05	0.05	0.05	0.04	0.05	-0.01	-0.01	0.00
						(0.001)	(0.001)	(0.001)
Living on farm	0.55	0.50	0.48	0.62	0.59	0.12	0.11	-0.01
						(0.01)	(0.01)	(0.01)
County contains a larger city	0.22	0.24	0.24	0.21	0.21	-0.03	-0.03	0.00
						(0.01)	(0.01)	(0.01)
B. Characteristics of closest neighbors								
Mother's education	8.29	8.27	8.25	8.33	8.29	0.06	0.04	-0.02
						(0.03)	(0.03)	(0.04)
Father's education	7.69	7.75	7.75	7.64	7.61	-0.11	-0.13	-0.02
						(0.03)	(0.03)	(0.04)
Living on farm	0.55	0.50	0.48	0.62	0.59	0.12	0.12	0.00
						(0.01)	(0.01)	(0.01)
C. Fraction female of nearby teachers (FFT)								
Closest 8	0.72	0.51	0.51	0.92	0.91	0.40	0.40	0.00
						(0.003)	(0.002)	(0.003)
ED-wide	0.72	0.51	0.51	0.92	0.91	0.40	0.40	0.00
						(0.002)	(0.002)	(0.003)
Observations*	405,800	95,800	102,500	101,100	106,300	–	–	–

Note: Statistics constructed using internal Census files. *Sample sizes are approximate. FFT is the fraction female among nearby teachers. Column (8) reports the difference between columns (6) and (7). Standard errors for differences and differences-in-differences, in parentheses, are clustered by county.

Table 4: Effect of Local FFT on Educational Attainment and Activities in 1940, White Adolescents

		Estimated effect of local fraction of female teachers (FFT)					
		County FE model		ED FE model, interaction effects		Household FE model, interaction effects	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)	Overall sample (6)	Less-Educ. parents (7)
On track for 9th grade	75.40	-1.06	3.26	3.57	4.52	3.42	4.50
		(0.26)	(0.28)	(0.26)	(0.34)	(0.41)	(0.52)
On track for 8th grade	85.82	-0.50	1.56	1.63	2.04	1.60	2.27
		(0.21)	(0.23)	(0.21)	(0.29)	(0.35)	(0.46)
Enrolled at Census date	65.08	-1.01	3.25	3.59	4.31	3.41	4.53
		(0.27)	(0.28)	(0.27)	(0.34)	(0.46)	(0.55)
Working at Census date	18.82	-0.34	-0.54	-0.55	-0.84	-0.36	-0.75
		(0.32)	(0.37)	(0.34)	(0.43)	(0.53)	(0.65)
Not enrolled or working	19.18	0.98	-2.44	-2.65	-3.21	-2.56	-3.31
		(0.27)	(0.40)	(0.36)	(0.46)	(0.57)	(0.72)
Counties/EDs		3,047		61,669	60,016	3,025	3,007
Observations		2,290,628		2,290,628	1,477,385	523,575	351,161

Notes: Table contains estimated coefficients for models fit to the dependent variable in the row heading, using data for White children age 15–18 in 1940 IPUMS. In all cases the dependent variable is a binary variable equal to 100 or 0. Models reported in columns (2) and (3) include family controls, family controls interacted with female, neighbor controls, neighbor controls interacted with female, county fixed effects, and county-level variables interacted with female (see Appendix for full list of controls). Models reported in columns (4) and (5) contain these same controls plus fixed effects for Enumeration Districts (ED), which absorbs the main effect in the model. Models reported in columns (6) and (7) contain the same controls but also household fixed effects (which again absorb main effects of family, and also neighbor variables). Models in columns (5) and (7) are fit to the subset of families in which maximum parental education is less than 8 years. Standard errors clustered by county or ED.

Table 5: Attenuation in the Estimated Interaction Effect, Models with "On Track for 9th Grade" as the Dependent Variable

		County model, estimated for all states except Wisconsin		Household model, estimated for all states except Wisconsin	
Teachers used to form FFT	Estimated attenuation factor in Wisconsin (1)	Baseline estimates (2)	Attenuation-corrected estimates (3)	Baseline estimates (4)	Attenuation-corrected estimates (5)
Closest 1	0.078	1.14	14.58	1.27	16.28
	(0.021)	(0.14)	(4.30)	(0.21)	(5.14)
Closest 2	0.132	1.81	13.71	1.92	14.55
	(0.032)	(0.18)	(3.60)	(0.29)	(4.15)
Closest 3	0.181	2.25	12.45	2.60	14.36
	(0.040)	(0.22)	(3.01)	(0.33)	(3.66)
Closest 4	0.249	2.55	10.23	2.81	11.29
	(0.047)	(0.25)	(2.17)	(0.36)	(2.57)
Closest 5	0.304	2.76	9.07	3.06	10.07
	(0.054)	(0.26)	(1.83)	(0.38)	(2.18)
Closest 6	0.338	2.93	8.66	3.08	9.11
	(0.058)	(0.27)	(1.68)	(0.40)	(1.96)
Closest 7	0.355	3.06	8.62	3.16	8.90
	(0.057)	(0.27)	(1.59)	(0.41)	(1.84)
Closest 8	0.354	3.08	8.71	3.23	9.12
	(0.060)	(0.28)	(1.67)	(0.42)	(1.95)
All in ED	0.368	3.37	9.17	3.64	9.90
	(0.061)	(0.30)	(1.73)	(0.44)	(2.04)

Notes: Data are from the 1940 IPUMS, and Wisconsin administrative records (see Appendix B). Each entry in column (1) is a coefficient from a regression of a student-level Census-based FFT on the corresponding FFT constructed from Wisconsin administrative records; the definition of "nearby teachers" used to construct the Census-based FFT varies across rows. Entries in column (2) are the estimated interaction effect in models that have "on track for 9th grade" as the dependent variable, using the county design. These results correspond to Table 4, but are estimated for White adolescents aged 15–18 *not* living in Wisconsin. Entries in column (3) are attenuation-corrected estimates, formed by dividing the entry in column (2) by the entry in column (1); standard errors are calculated by the delta method. Columns (4) and (5) provide corresponding estimates for the household design.

Table 6: Characteristics of Black Adolescents Aged 15–18 in the 1940 Census

	Adolescents aged 15–18			
			In small towns/rural areas, non-movers with teacher data	
	All	In Southern states	All	In household sample
	(1)	(2)	(3)	(4)
Fraction living with parent(s)	0.795	0.788	1.000	1.000
Fraction with mother only	0.198	0.186	0.197	0.170
Fraction female	0.517	0.516	0.482	0.500
Fraction enrolled at Census date	0.512	0.481	0.555	0.566
Mean grade completed	6.733	6.299	6.074	6.076
Fraction in small place	0.593	0.689	1.000	1.000
Fraction living on farm	0.437	0.519	0.721	0.763
Fraction non-mover	0.768	0.815	1.000	1.000
Fraction with teacher data*	0.521	0.564	1.000	1.000
Mean mother's education	5.501	5.116	4.883	4.938
Mean father's education	4.533	4.133	3.877	3.919
Sample Size	1,011,416	839,866	190,138	58,245

Notes: Data are from the 1940 IPUMS. *Individuals are recorded as "having teacher data" only if they are in an enumeration district with with at least two Black teachers.

Table 7: Effect of Local FFT on Educational Attainment and Activities in 1940, Southern Black Adolescents

		Estimated effect of local fraction of female teachers (FFT)					
		County FE model		ED FE model, interaction effects		Household FE model, interaction effects	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)	Overall sample (6)	Less-educ. parents (7)
On track for 9th grade	42.3	-2.98	3.21	2.86	3.28	1.14	1.50
		(1.09)	(1.17)	(1.11)	(1.19)	(1.73)	(1.82)
On track for 8th grade	51.9	-2.81	3.05	2.67	3.30	2.49	3.16
		(1.19)	(1.17)	(1.14)	(1.22)	(1.64)	(1.73)
Completed 8th grade	27.5	-2.51	3.71	3.39	3.63	2.02	2.04
		(0.89)	(1.05)	(0.96)	(1.00)	(1.42)	(1.47)
Completed 6th grade	56.7	-2.07	3.06	2.95	3.06	2.07	2.03
		(1.01)	(1.10)	(1.07)	(1.14)	(1.64)	(1.73)
Enrolled at Census date	55.5	-3.19	2.25	1.82	2.34	1.22	1.42
		(1.33)	(1.25)	(1.16)	(1.22)	(1.69)	(1.76)
Working at Census date	34.2	1.57	-2.13	-1.95	-1.89	-2.15	-1.97
		(1.56)	(1.50)	(1.38)	(1.46)	(1.78)	(1.90)
Not enrolled or working	16.4	0.20	0.89	1.21	0.78	2.98	2.65
		(0.82)	(1.14)	(1.16)	(1.22)	(1.46)	(1.51)
Counties/EDs		925		4,765	4,745	873	865
Observations		187,160		187,160	168,053	57,730	51,782

Notes: Table contains estimated coefficients for models fit to the binary dependent variable in row headings, using data for Black children age 15–18 in 1940 IPUMS. In all cases the dependent variable is a binary variable equal to 100 or 0. Models include controls as reported in the notes to Table 4. Models in columns (5) and (7) are fit to subset of families in which maximum parental education is less than 8 years. Standard errors clustered by county or ED.

Table 8: "Falsification Exercise" for Southern Black and White Adolescents: Interaction Effect of FFT (in the ED Model) with Own-Race and Opposite-Race FFT Measures

	Black students		White students	
	Black FFT	White FFT	Black FFT	White FFT
	(1)	(2)	(3)	(4)
On track for 9th grade	2.72	-1.25	0.78	4.21
	(1.37)	(1.62)	(0.95)	(1.28)
On track for 8th grade	2.73	1.40	0.42	3.92
	(1.37)	(1.61)	(0.87)	(1.20)
Completed 8th grade	3.65	0.20	0.15	3.37
	(1.19)	(1.46)	(0.94)	(1.31)
Completed 6th grade	3.28	1.82	-0.67	1.81
	(1.31)	(1.54)	(0.70)	(0.94)
Observations	117,118	117,118	158,724	158,724

Notes: Notes: Table contains estimated coefficients for models fit to the dependent variable in row headings, using data for Southern Black and White children age 15–18 in 1940 IPUMS. In all cases the dependent variable is a binary variable equal to 100 or 0. Control variables are as reported in the notes to Table 4. The sample is restricted to enumeration districts that have at least two White and two Black teachers, and non-missing teacher wage, age, and teacher education information.

Table 9: Effect of Local FFT on Educational Attainment and Income in 2000, White Individuals who were Aged 7–14 in 1940

		Estimated effect of local FFT			
		County FE model		ED FE model interaction effects	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)
Years of schooling reported in 2000	12.08	0.021	0.103	0.125	0.130
		(0.034)	(0.039)	(0.046)	(0.070)
Completed high school	75.18	0.36	1.40	1.27	2.42
		(0.51)	(0.64)	(0.76)	(1.18)
Some college	33.11	-0.48	2.41	2.95	1.91
		(0.52)	(0.68)	(0.79)	(1.03)
Log of SS income, including allocated data	9.022	-0.015	0.024	0.033	0.045
		(0.004)	(0.006)	(0.009)	(0.012)
Log of SS income	9.037	-0.015	0.022	0.028	0.045
		(0.005)	(0.010)	(0.009)	(0.015)
Log personal income	9.734	-0.006	0.018	0.018	0.025
		(0.008)	(0.011)	(0.015)	(0.020)
Counties/EDs*		3,000		55,000	48,000
Observations*		387,000		387,000	213,000

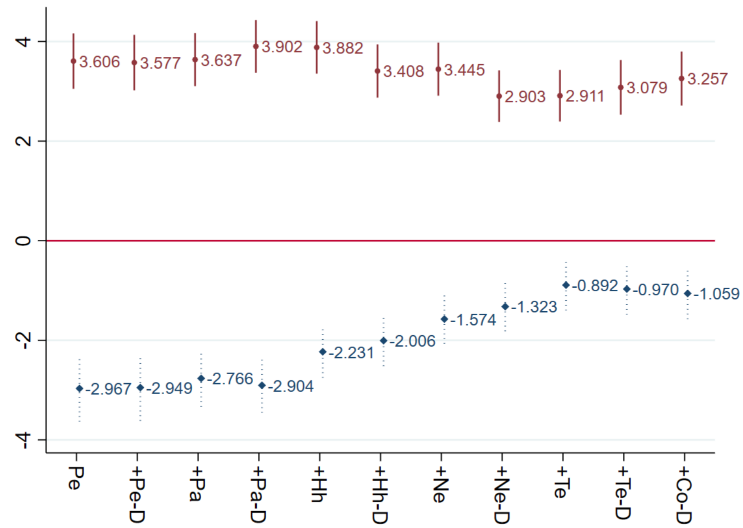
Notes: Data are from 1940 Census internal records matched to 2000 Census internal long-form records. *Sample sizes are approximate. Coefficient estimates are for models fit to dependent variable in row headings; binary variables (in the first three regression models) are coded as 100 or 0. Controls are as reported in notes to Table 4. Results in column (5) are for models fit to families in which maximum parental education is less than 8 years. Standard errors clustered by county or ED. We excludes cases with allocated values, except as indicated.

Table 10: Effect of Local FFT on Survival, White Individuals who were Aged 7–14 in 1940

		Estimated effect of local FFT			
		County FE model		ED FE model, interaction effects	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)
Alive in 2000: Appeared in the 2000 Census or died thereafter	72.05	-0.466	0.540	0.610	0.811
		(0.169)	(0.215)	(0.222)	(0.291)
Alive in 2010: Appeared in the 2010 Census or died thereafter	48.38	-0.329	0.305	0.364	0.684
		(0.178)	(0.236)	(0.244)	(0.316)
Survived to 2010 conditional on being alive in 2000	67.15	-0.172	0.086	0.143	0.456
		(0.207)	(0.262)	(0.275)	(0.370)
Age at death (imputed to 101 if alive in 2010)	79.55	-0.143	0.187	0.194	0.255
		(5.438)	(0.068)	(0.071)	(0.091)
Log age at death (imputed to 101 if alive in 2010)	4.358	-0.204	0.283	0.292	0.392
		(0.078)	(0.094)	(0.097)	(0.126)
Counties/EDs*		3,000		61,000	59,000
Observations*		2,997,000		2,997,000	1,699,000

Notes: Data are from 1940 Census internal records matched to 2000 Census internal long-form records and death records. *Sample sizes are approximate. Coefficient estimates are for models fit to dependent variable in row headings; binary variables (in the first three regression models) are coded as 100 or 0. Controls are as reported in notes to Table 4. Results in column (5) are for models fit to families in which maximum parental education is less than 8 years. Standard errors clustered by county or ED.

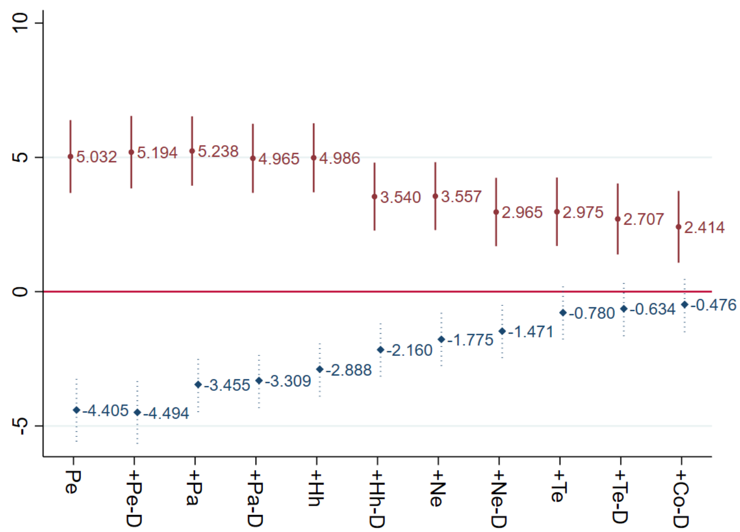
Figure 1: Effect of Adding Controls on Key Coefficient Estimates: Analysis of 9th Grade Attainment (1940)



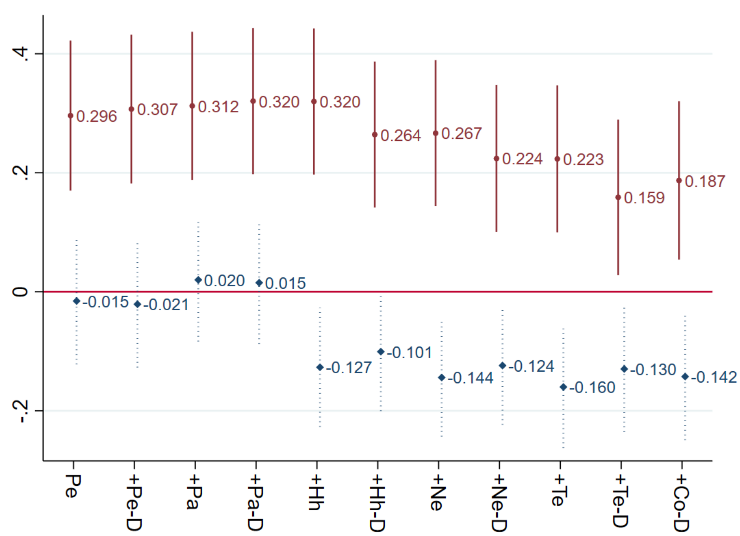
Data: 1940 PUMS. Point estimates and 95% confidence intervals of *main effects* are in blue; *interaction effects* are in red. On the far left of the graph, we provide estimates for a model with minimal control variables—only variables related to the child (age and Hispanic status). Then we add those same variables but with gender interactions. Then moving to the right we add parental variables; then parental variables interacted with child gender; then household controls; then household controls interacted with gender; then neighborhood controls; then neighborhood controls interacted with gender; then teacher controls; then teacher controls interacted with gender; and finally also county control variables interacted with gender.

Figure 2: Effect of Adding Controls on Key Coefficient Estimates: Outcomes in 2000

A. College attendance



B. Age of death



Data: Matched 1940 and 2000 Census internal files. See Figure 1 for regression specifications.

APPENDICES: NOT FOR PUBLICATION

Appendix A Literature Review

To help place our findings into intellectual context, we briefly review the recent literature in economics on the effects of teacher gender on student outcomes. We summarize findings from 24 articles, published since 2000 in economics, which provide empirical evidence on this matter. To keep the review manageable we ignore papers that were published earlier or in non-economics journals. Of these 24 papers:

- Six papers study students in low-income countries or study students from low-SES families in medium-income countries: children in rural India (Muralidharan et al. 2016); students in China (Gong et al. 2018, and Xu and Li 2018, and Eble and Hu 2020); students in ten African countries (Lee et al. 2019); and girls with with poorly-educated mothers in Chile (Paredes 2014). *All of these papers show that female teachers improve outcomes for female students, while having little or no adverse impacts on male students.* Several of them show also that beneficial effects are larger for girls from lower-SES backgrounds.

- Ten papers report largely beneficial impacts of female teachers for female students in medium- and high-income countries: Dee (2005), Dee (2007), Carrell et al. (2010), Winters et al. (2013), and Egalite et al. (2018), and Mansour et al. (2022) in the US; Lindahl (2016) in Sweden; Paredes (2014) in Chile; and Lim and Mear (2017 and 2020) in Korea. Many of these papers find beneficial impacts of female teachers for female students *specifically in science and math.*

- Six papers report weak evidence, mixed evidence, or no evidence of beneficial impacts of female teachers for female students—all in high-income countries: Bettinger et al. (2005), Krieg (2005), Price (2010), and Gershenson et al. (2016) in the US; Holmlund et al. (2008) in Sweden; and Cho (2013) in 15 OECD countries.

- Only two papers feature negative impacts of female teachers for female students. Antecol et al. (2015) find that female students in math to tend not fare well when matched to female teachers in the Teach for America organization, which places volunteer teachers in low-income neighborhoods. As the authors emphasize, this effect is driven primarily by female teachers who lack a strong math background. Gørtz et al. (2018) study Danish preschools, where teachers are overwhelmingly female (91% female at the median school). In that environment there appears to be a small benefit (in terms of test scores) to having at least one male teacher in the preschool.

In addition, we include in our review three relevant papers which focus primarily on the extent to which teachers' behaviors, expectations, and attitudes affect outcomes among female and male students: Sansone (2017), Alan et al. (2018), and Carlana (2019). The latter two studies specifically evaluate gender stereotypes held by teachers, and ask how they differentially impact male and female students.

Finally, listed separately from the other articles (at the end of the list), are five related papers that focus primarily on the impacts of teacher and student gender on grading and assessment.[36] These papers find evidence of consequential grading bias among teachers—by comparing systematic gender differences in boys' and girls' non-blind versus blind student test scores. Importantly, *none of the papers show that female teachers are biased in favor of female students in terms of grading.*

We draw several lesson from this literature, relevant to our own work:

First, female teachers seem to matter most for female students in educational situations in which those teachers are plausibly counter-stereotypical, in particular in less developed countries, where women often are marginalized in society, and female students lag male students in achievement.[37]

Second, the research on the consequences of teacher gender stereotyping (Alan et al. 2018, and Carlana, 2019) provides further direct evidence on a mechanism by which female teachers can play a key role in improving outcomes for female students without causing any harm to male students.

Third, research on gender biases in grading is important because it demonstrates a key mechanism whereby teacher/student genders can be consequential for students. One might hypothesize that female teachers would be more biased in favor of female students than male teachers—which of course might benefit female students while harming male students. Importantly, however, no evidence in the literature favors this hypothesis.

Finally, while none the research we review examines the specific outcomes in our study—progression in school, completed education, and other lifetime outcomes—the literature nonetheless provides us with a sensible way of tracing out likely impacts of female teachers in environments in which those teachers are counter-stereotypical role models. Consider, for instance, the consequences of female teachers found in the clear and persuasive analysis of young students in Andra Pradesh by Muralidharan et al. 2016. This study shows that each year of student exposure to a female teacher increases female academic achievement by 0.034σ, while boys suffer no adverse outcomes. Thus, female students who have female teacher during their first six years of school plausibly increase achievement by more that 0.2σ, and plausibly have stronger non-cognitive skills as well (as shown in many papers reviewed below). Many of these students would then be better positioned and motivated

[36] Mechtenberg (2009) provides an interesting theoretical contribution.

[37] Even the two papers finding negative effects of female teachers on female students could be construed to be consistent with the importance of counter-stereotypical role models. In Danish preschools, studied by Gørtz et al. (2018), *male* teachers are rare, and might be viewed as counter-stereotypical. The authors of that study note that the positive impact of male teachers seems to be largest for boys who lack male role models at home. As for work by Antecol et al. (2015), we find plausible the authors' suggestion that "... math anxiety among primary school female teachers in conjunction with female student endorsement of gender stereotypes may be leading to poorer math achievement among female students but not male students."

to progress to higher grades.[38] Hopefully, researchers will examine this important question; our work does so, though of course in a different educational context.

[38]If so, this would lower the male/female ratio of students in upper grades found in rural Andra Pradesh, which now exceeds 1 for students in most castes (see Bagde, et al. 2022).

Recent Papers on Effects of Teacher Gender on Student Outcomes*

Paper	Summary, with estimated effects (where applicable)
Bettinger et al. (2005)	At public four-year colleges in Ohio, female faculty in freshman classes are found to impact subsequent course selection and major choice by female students. Female faculty have mixed statistically significant effects on female students—increasing engagement in mathematics and statistics, geology, sociology, and journalism and communications, while reducing engagement in biology, physics, economics, and political science.
Dee (2005)	An analysis of data from the National Education Longitudinal Study of 1988 (NELS88) shows that 8th graders are perceived by opposite-gender teachers to be more inattentive, more disruptive, and more likely to "rarely complete homework." Similar results pertain also for students taught by other-race teachers.
Krieg (2005)	Being matched to teachers by gender is shown to have no effect on achievement—as measured by the Washington Assessment of Student Learning—among 3rd/4th graders in Washington.
Dee (2007)	Using data from the NELS88, the author finds that assignment to a same-gender teacher improves achievement by about 0.05σ for both boys and girls. Teacher-gender impacts are also found along other dimensions, e.g., boys are more likely to be seen as disruptive when the teacher is female, and girls are about 0.1σ less likely to report "science is not useful for their future" when their teacher is female.
Holmlund et al. (2008)	In a study of upper-secondary students in Stockholm, the authors find little evidence of teacher-student gender match effects.
Hoffmann et al. (2009)	In a study of University of Toronto students, same-gender instructor effects are found to be small—increasing grade performance by 0.05σ or less, and lowering the likelihood of dropping a class by 0.04σ.
Carrell et al. (2010)	When female students at the US Air Force Academy are assigned to female professors, grades in mandatory math and science courses increase by about 0.1σ, while having little impact on male students. For female students with high SAT scores, the effect is closer to 0.2σ. Among these latter female students, having female professors increases the likelihood of follow-on STEM coursework, and increases graduation with a STEM degree by nearly 0.3σ.

*Papers published in economics over the past 20 years (since 2002). Results are reported as statistically significant only if the p value is ≤ 0.05. Papers on grading bias are featured separately (below).

Recent Papers on Effects of Teacher Gender on Students (Continued)

Price (2010)	An OLS analysis of students in first-semester STEM courses at public four-year universities in Ohio shows that each additional female STEM instructor reduces persistence in a STEM major by about one percentage point (about 0.04σ) among males after the first semester, while having no statistically significant effect for female students. (A complementary IV design yields statistically insignificant estimates.)
Winters et al. (2013)	The authors examine the effect of female teachers on standardized math and reading scores of students in Florida—grades 4 through 10. No effects of teacher gender are detected in grades 4 and 5. In grades 6 through 10, however, having a female teacher improves math achievement among female students (by 0.02σ to 0.04σ), while improving male achievement by slightly less (0.01σ to 0.03σ).
Cho (2013)	The author estimates the effects of teacher-student gender matching on achievement, measured by the Trends in International Mathematics and Science Study (TIMSS) in 15 OECD countries. As indicated in the paper's abstract: "The results provide little support for the conjecture that students benefit from teacher-student gender matching." Statistically significant teacher gender-match effects are as follows: Female teachers improved female student scores in math in France (0.09σ) and Greece (0.06σ), and in science in Sweden (0.03σ). Male teachers improved male student math scores in Spain (0.12σ), and in science in Canada (0.06σ) and Spain (0.13σ).
Paredes (2014)	For female 8th graders in Chile, having a female teacher increase scores in math (by 0.03σ to 0.05σ), in natural sciences (by 0.02σ to 0.03σ) and in social sciences (by 0.02σ to 0.03σ). Estimated effects are larger for girls with poorly-educated mothers, and are close to 0 among girls with well-educated mothers. Teacher gender does not generally impact achievement among male students.
Antecol et al. (2015)	The authors study impacts of teachers in the Teach for America corps, who teach primary school students in disadvantaged neighborhoods. Having a female teacher does not affect male student test scores (math or reading) and does not affect female student reading test scores, but has a *negative* impact of female student math test scores (about -0.1σ). This outcome is driven primarily by female teachers who do not have a strong math background. The authors suggest that "… math anxiety among primary school female teachers in conjunction with female student endorsement of gender stereotypes may be leading to poorer math achievement among female students but not male students."

Recent Papers on Effects of Teacher Gender on Students (Continued)

Gershenson et al. (2016)	Using data from the Education Longitudinal Study of 2002 (ELS), the authors evaluate teacher expectations of their 10th grade students. They find little effect of teacher gender matching on teacher expectations. (However, they do find that non-Black teachers have lower expectations of Black students.)
Muralidharan et al. (2016)	The authors study education in public primary schools in rural Andhra Pradesh (India). In this setting, girls and boys have similar levels of knowledge in both math and language at the start of primary school (girls indeed have a slight advantage in language), but then girls lose ground relative to boys—losing 0.01σ/year in language and 0.02σ/year in math assessments. Against this backdrop, female students taught by female teachers see gains of 0.034σ/year, relative to boys, in test scores (math and language combined). Boys suffer no adverse effects from being taught by female teachers.
Lim and Meer (2017)	Female Korean middle school students are found to have higher standardized test scores when taught by female teachers (impacts are near 0.1σ). In contrast, teacher gender has little effect on test scores of male students. When female students are taught by female teachers, they are more likely to report that they have an equal chance to participate in class (6 pp), that their teacher encourages expression (6 pp) and that the subject they are studying is their favorite subject (4 pp).
Sansone (2017)	In an analysis of data from the High School Longitudinal Study of 2009, the author finds significant effects of teacher gender on students' interest and self-confidence in math and science (e.g., female teachers raise self-confidence in female students by 0.05σ). However, effects are not significant in specifications that control also for a set of variables measuring teacher behaviors, expectations and attitudes. This leads the author to suggest, "Teacher beliefs about male and female ability in math and science—as well as how teachers treat boys and girls in the classroom—matter more than teacher's own gender."
Alan et al. (2018)	In a low income area in Istanbul, female primary students (but not their male counterparts) are found to have lower performance on math and verbal tests when taught by teachers who hold traditional views on gender roles. Given a four-year exposure to a teacher, a one standard deviation increase in the teacher gender stereotype construct has a -0.21σ effect on female student math scores and a -0.16σ effect on verbal scores.

Recent Papers on Effects of Teacher Gender on Students (Continued)

Egalite et al. (2018)	Using data from six school districts across the US, the authors study effects of student-teacher gender matches for students, grades 4–8, in terms of student perceptions (student feels cared for by their teacher, student interest and enjoyment of classwork, quality of teacher-student communication, clarity in teaching style and methods, etc.). Student perceptions and attitudes are generally more favorable when they have a same-gender teacher: effect sizes (when significant) are in the range of 0.04σ to 0.09σ. (Similar effects pertain for student-teacher race matching.)
Gong et al. (2018)	In an analysis of data from the 2014 China Education Panel Survey (CEPS), a nationally representative survey covering middle schools, the authors find that for female students, having a female teacher improves test scores by about 0.2σ, relative to boys, and increases self-assessed learning performance by more than 0.3σ. Impacts on test scores are larger for girls whose mothers have ≤ 9 years of education (0.27σ) and ethnic minorities (0.46σ). Also, many non-cognitive outcomes are studied: for example, when taught by a female teacher, girls are less likely to report feeling depressed, blue, or unhappy at school, and report stronger social acclimation and satisfaction.
Gørtz et al. (2018)	Using administrative data, the authors study Danish children enrolled in preschool (2006–2007)—a setting in which very few teachers are male (at the median preschool, 91% of teachers are women). Here, an increase in the fraction *male* among teachers increases student test scores at age five, e.g., a 10 pp increase in fraction male increases reading test scores by 0.035σ for boys and 0.026σ for girls.
Xu and Li (2018)	In a study utilizing two waves of the CEPS, the authors study effects of teacher-student gender matches on students. The authors find that boys are little affected by teacher gender. Having a female teacher improves girls' self-reported interactions with students, and increases exam scores in math (0.15σ) and Chinese (0.10σ)

Recent Papers on Effects of Teacher Gender on Students (Continued)

Carlana (2019)	The author measures gender stereotypes using an implicit associaton test (IAT) among teachers in Northern Italy and then examines impacts on students of exposure to teacher stereotypes. A one standard deviation increase in a teacher's IAT score is found to decrease females' 8th-grade math performance by more than 0.03σ relative to male counterparts. Math teacher stereotypes also induce female students to self-select into less demanding academic tracks, and have a negative impact on their self-confidence in math.
Lee et al. (2019)	This paper studies 2nd and 6th grade students in ten African countries. Teacher gender has no effect on boys, but being taught by female teachers is beneficial for girls—increasing girls' 2nd grade reading and math scores by 0.1σ, and 6th grade reading by 0.1σ and math by 0.2σ.
Eble and Hu (2020)	In a study of Chinese students using data from the CEPS, the authors examine the impact of teacher gender on widely-held gender stereotypes. When girls who believe themselves to have low math ability are assigned to female math teachers, they score 0.45σ higher on math exams relative to other children who perceive themselves to have low ability. They are also less likely to perceive math as difficult, and less likely to aspire to traditionally-female jobs.
Lim and Meer (2020)	Using data from the Seoul Education Longitudinal Study (2010), which tracks students from grades 7–12, the authors find that female 7th graders taught by a female teacher score higher on a 7th grade standardized tests (by 0.19σ). These effects are persistent, resulting in increases in test scores that range from 0.16σ in grade 8 to 0.25σ in grade 12. There is also evidence that having a female 7th grade math teacher increases the likelihood female students attend a STEM-focused high school, take advanced math, and aspire to a STEM degree.
Mansour, et al. (2022).	The authors study post-graduation outcomes for students at the US Air Force Academy. Female students with more female professors are more likely to complete a masters degree in STEM and less likely to have a professional degree. Effects on male students are in the opposite direction and about half as big.

Recent Papers that Focus on Grading/Assessment Bias

Paper	Summary, with estimated effects (where applicable)
Lavy (2008)	The author exploits blind and non-blind exam score data on matriculation exams taken by Israeli high school students. Bias in grading—the difference between blind and non-blind scores—on average disadvantages boys. The size of the disadvantage ranges from 0.05σ to 0.18σ in nine academic subjects. However, no clear associations were found between *teacher gender* and this grading bias.
Lindahl (2016)	This work evaluates impacts of teacher gender on student outcomes among 9th graders in Sweden. When female students are found to have higher math test scores when taught by female teachers; we infer that being taught by all-female teachers (rather than all-male) increases math test scores by 0.21σ. Furthermore, there is evidence of gender bias, favoring girls, in terms of being "graded up" (receiving a relatively favorable teacher-assigned score on the "School Leaving Certificate"). However, if anything, female teachers are less likely than male teachers to be biased in terms of "grading up."
Puhani (2018)	This paper studies 5th to 9th graders in the German state of Hesse (2007–2012), where only 10% of teachers are male. The author examines the impact of teacher gender on the teachers' recommendations for middle school type, and also on the actual middle school type attended. There are "virtually no teacher gender effects;" the only exception is an increase in high-track recommendations for boys by male teachers (0.07σ), which did not translate to increased attendance at that higher track.
Lavy and Sand (2018)	The authors find lasting consequences of teacher gender biases in grading on 6th grade students in Tel-Aviv. For example, they show that teacher gender bias—measured as the difference between boys' and girls' average gaps in non-blind and blind test scores—has a positive effect on subsequent enrollment in advanced high school math courses for boys relative to girls. A one SD decrease in teacher gender bias is estimated to reverse the gender gap in advanced course participation, from 3.2 pp in favor of boys to 4.6 pp in favor of girls.
Terrier (2020)	The author utilizes a data set of French students (in grades 6–11), which includes both blind and non-blind test scores, to explore teacher bias in grading. Many middle school teachers tend to bias grading in favor of female students, and these biases are found to affect educational outcomes. For example, being exposed to a math teacher with a one standard deviation bias in favor of girls, compared to one with no bias, increases by 0.10σ the likelihood a female student chooses a scientific track in high school.

Additional References for Appendix A

Bettinger, Eric P. and Bridget T. Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review* 95(2): 152-157.

Canaan, Serena and Pierre Mouganie. forthcoming. "The Impact of Advisor Gender on Female Students' STEM Enrollment and Persistence." *Journal of Human Resources.*

Cho, Insook. 2012. "The Effect of Teacher-Student Gender Matching: Evidence from OECD Countries." *Economics of Education Review* 31(3): 54-67.

Cornwell, Christopher, Mustard, David B., and Van Parys, Jessica. 2013. "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School." *Journal of Human Resources* 48(1): 236-264.

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95(2): 158-165.

Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources* 42(3): 528-554.

Eble, Alex, and Feng Hu. 2020. "Child Beliefs, Societal Beliefs, and Teacher-Student Identity Match" *Economics of Education Review* 77: 101994

Egalite, Anna J. and Brian Kisida 2018. "The Effects of Teacher Match on Students' Academic Perceptions and Attitudes." *Educational Evaluation and Policy Analysis* 40(1): 59-81.

Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2016. "Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations." *Economics of Education Review* 52: 209-224.

Gørtz, Mette, Eva Rye Johansen, and Marianne Simonsen. 2018. "Academic Achievement and the Gender Composition of Preschool Staff." *Labour Economics* 55: 241-258.

Griffith, Amanda L. 2014. "Faculty Gender in the College Classroom: Does it Matter for Achievement and Major Choice?." *Southern Economic Journal* 81(1): 211–231.

Holmlund, Helena, and Krister Sund. 2008."Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics* 15(1): 37-53.

Hoffmann, F. and P. Oreopoulos. 2009. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources* 44(2), 479-494.

Krieg, John M. 2005. "Student Gender and Teacher Gender: What Is the Impact on High Stakes Test Scores?" *Current Issues in Education* 8(9).

Lindahl, Erica. 2016. "Are Teacher Assessments Biased? Evidence from Sweden." *Education Economics* 24(2): 224-238.

Mechtenberg, Lydia. 2009. "Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages." *Review of Economic Studies* 76(4): 1431-1459.

Morgenroth, Thekla, Michelle K. Ryan, and Kim Peters. 2015. "The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals." *Review of General Psychology* 19(4): 465-483.

Paredes, Valentina. 2014. "A Teacher Like Me or a Student Like Me? Role Model versus Teacher Bias Effect." *Economics of Education Review* 39: 38-49.

Puhani, Patrick A. 2018. "Do Boys Benefit from Male Teachers in Elementary School? Evidence from Administrative Panel Data." *Labour Economics* 51: 340-354.

Sansone, Dario. 2017. "Why does Teacher Gender Matter?" *Economics of Education Review*, 61: 9-18

Terrier, Camille. 2020. "Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement." *Economics of Education Review* 77: 101981

Xu, Di, and Qiujie Li. 2018. "Gender Achievement Gaps among Chinese Middle School Students and the Role of Teachers' Gender." *Economics of Education Review* 67: 82-93.

Appendix B Measurement Error: Theory and Evidence

B.1 Measurement Error in our Research Designs

County Design

Recall that the data generating process in the county FE design is assumed to be

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 F^*_{d(i)} + \beta_3 D_i F^*_{d(i)} + \theta_{c(i)} + u_i, \tag{16}$$

where D_i is a dummy for student female gender, F^*_d is the true fraction female among teachers in enumeration district d, $d(i)$ is an index function giving the enumeration district of individual i, and $c(i)$ is another index function giving her county. We assume that true FFT, $F^*_{d(i)}$, relates to the average county FFT, $F_{c(i)}$ as follows:

$$F^*_{d(i)} = F_{c(i)} + \xi_i, \text{ with } \xi_i \perp F_{c(i)},$$

and that measured FFT relates true FFT according to

$$F_{d(i)} = F^*_{d(i)} + \phi_i, \text{ with } \phi_i \perp [F_{c(i)}, F^*_{d(i)}].$$

Thus,

$$\begin{aligned} F_{d(i)} &= F_{c(i)} + \xi_i + \phi_i \\ &= F_{c(i)} + \eta_i. \end{aligned}$$

Within-county variation in observed FFT has variance, $\sigma^2_\eta = \sigma^2_\xi + \sigma^2_\phi$, a combination of "signal" (σ^2_ξ) and "noise" (σ^2_ϕ).

To simplify the discussion, assume that half the students in every county are female (i.e., $\mathrm{E}[D_i|c(i)] = 0.5$) and that half the students in every enumeration district are also female, so $\mathrm{E}[D_i F_{d(i)}|c(i)] = 0.5F_{c(i)}$. Let $\mathrm{E}[F_{c(i)}] = \mu_F$. We use $F_{d(i)}$ in place of $F^*_{d(i)}$ when estimating model (16) and include county fixed effects. This is equivalent to running a simple regression of the outcome y_i on three variables—the female gender dummy, observed FFT in the enumeration district, and the interaction of female gender and observed FFT—*all deviated from their county-specific means.* Under our simplifying assumptions this amounts to estimating the sample version of the following population regression model:

$$y_i = \delta_0 + \delta_1 D_i + \delta_2 \left[F_{d(i)} - F_{c(i)}\right] + \delta_3 \left[D_i F_{d(i)} - 0.5F_{c(i)}\right] + v_i. \tag{17}$$

To derive expressions for the coefficients $(\delta_1, \delta_2, \delta_3)$ in this population model, we project each of the three terms on right-hand side of (16) on the the terms on the right-hand side of (17). Ignoring constants, these projections can be expressed as

$$D_i = \pi_{11}D_i + \pi_{12}(F_{d(i)} - F_{c(i)}) + \pi_{13}(D_iF_{d(i)} - 0.5F_{c(i)}) + v_{1i}, \tag{18}$$
$$F^*_{d(i)} = \pi_{21}D_i + \pi_{22}(F_{d(i)} - F_{c(i)}) + \pi_{23}(D_iF_{d(i)} - 0.5F_{c(i)}) + v_{2i}, \tag{19}$$
$$D_iF^*_{d(i)} = \pi_{31}D_{d(i)} + \pi_{32}(F_{d(i)} - F_{c(i)}) + \pi_{33}(D_iF_{d(i)} - 0.5F_{c(i)}) + v_{3i}. \tag{20}$$

where the π_{ij} coefficients are to be determined. We can then express the coefficients in equation (17) as

$$\begin{aligned}
\delta_1 &= \beta_1\pi_{11} + \beta_2\pi_{21} + \beta_3\pi_{31}, \\
\delta_2 &= \beta_1\pi_{12} + \beta_2\pi_{22} + \beta_3\pi_{32}, \\
\delta_3 &= \beta_1\pi_{13} + \beta_2\pi_{23} + \beta_3\pi_{33}.
\end{aligned}$$

Deriving the projection coefficients

To begin, note that $\pi_{11} = 1$, $\pi_{12} = 0$, $\pi_{13} = 0$ in equation (18). Next, we assume that $\beta_2 = 0$, i.e., that female teachers have no causal effect on male students (i.e., no "main effect"). In this case, the only unknowns are $(\pi_{31}, \pi_{32}, \pi_{33})$. To find these coefficients, let X represent the vector of variables on the RHS of equations (18), (19), and (20):

$$\begin{aligned}
x_1 &= D_{d(i)}, \\
x_2 &= F_{d(i)} - F_{c(i)}, \\
x_3 &= D_{d(i)}F_i - 0.5F_{c(i)}.
\end{aligned}$$

To calculate the coefficients $(\pi_{31}, \pi_{32}, \pi_{33})$ we need to calculate $\text{var}[X]^{-1}$ and $\text{cov}[X, D_{ij}F^*_{ij}]$.

The diagonal terms in $var[X]$ are

$$\begin{aligned}
v_1 = \text{var}[x_1] &= 0.25, \\
v_2 = \text{var}[x_2] &= \sigma_n^2, \\
v_3 = \text{var}[x_3] &= E[(D_iF_{d(i)} - 0.5F_{c(i)})^2] \\
&= 0.5E[F^2_{d(i)}] - 0.25E[F^2_{c(i)}] \\
&= 0.5(E[F^2_{c(i)}] + \sigma_n^2) - 0.25E[F_j^2] \\
&= 0.25\text{var}[F_{c(i)}] + .25\mu_F^2 + 0.5\sigma_n^2.
\end{aligned}$$

where μ_F is the mean fraction of female teachers.

The covariance (off-diagonal) terms in var$[X]$ are

$$
\begin{aligned}
c_{12} = \text{cov}[x_1, x_2] &= 0, \\
c_{13} = \text{cov}[x_1, x_3] &= E[(D_i - 0.5)(D_i F_{d(i)} - 0.5F_{c(i)})] \\
&= E[(D_i F_{d(i)} - 0.5D_{d(i)}F_{c(i)} - 0.5D_i F_{d(i)} + 0.25F_{c(i)}] \\
&= 0.25\mu_F, \\
c_{23} = \text{cov}[x_2, x_3] &= E[(F_{d(i)} - F_{c(i)})(D_i F_{d(i)} - 0.5F_{c(i)})] \\
&= E[D_i(F_{d(i)} - F_{c(i)})F_{d(i)} - 0.5(F_{d(i)} - F_{c(i)})F_{d(i)}] \\
&= 0.5\sigma_n^2.
\end{aligned}
$$

So var$[X]$ is

$$
\begin{bmatrix} v_1 & 0 & c_{13} \\ 0 & v_2 & c_{23} \\ c_{13} & c_{23} & v_3. \end{bmatrix}.
$$

The inverse is then

$$
\frac{1}{\Delta} \begin{bmatrix} v_2 v_3 - c_{23}^2 & c_{13}c_{23} & -v_2 c_{13} \\ c_{13}c_{23} & v_1 v_3 - c_{13}^2 & -v_1 c_{23} \\ -v_2 c_{13} & -v_1 c_{23} & v_1 v_2 \end{bmatrix},
$$

where

$$
\begin{aligned}
\Delta &= v_1 v_2 v_3 - v_1 c_{23}^2 - v_2 c_{13}^2 \\
&= \frac{\sigma_n^2}{16}\left(\text{var}[F_{c(i)}] + \sigma_n^2\right).
\end{aligned}
$$

Finally, the elements of the vector cov$[X, D_{ij}F_{ij}^*]$ are:

$$
\begin{aligned}
E[(D_i - 0.5)D_i F_{d(i)}^*] &= E[(D_i - 0.5)D_i]E[F_{d(i)}^*] = 0.25\mu_F, \\
E[(F_{d(i)} - F_{c(i)})D_i F_{d(i)}^*] &= E[D_i]E[F_{d(i)} - F_{c(i)}F_{d(i)}^*] = 0.5\sigma_\xi^2, \\
E[(D_i F_{d(i)} - 0.5F_{c(i)})D_i F_{d(i)}^*] &= E[D_i F_{d(i)} F_{d(i)}^*] - 0.5E[D_i F_{c(i)} F_{d(i)}^*] \\
&= 0.5E[F_{d(i)}^{*2}] - 0.25E[F_{c(i)}^2] \\
&= 0.25\text{var}[F_{c(i)}] + 0.25\mu_F^2 + 0.5\sigma_\xi^2.
\end{aligned}
$$

Using these terms we can derive the coefficients $(\pi_{31}, \pi_{32}, \pi_{33})$.

A series of substitutions and simplifications then allow us to solve for the coefficients in

(17). Specifically, we obtain:

$$\delta_1 = \beta_1 + \frac{\beta_3 \sigma_\phi^2}{\mathrm{var}[F_{d(i)}]} \mu_F, \tag{21}$$

$$\delta_2 = -\frac{\beta_3}{2} \frac{\mathrm{var}[F_{c(i)}]}{\mathrm{var}[F_{d(i)}]} \times \frac{\sigma_\phi^2}{\sigma_\xi^2 + \sigma_\phi^2}, \tag{22}$$

$$\delta_3 = \beta_3 \frac{\mathrm{var}[F^*_{d(i)}]}{\mathrm{var}[F_{d(i)}]}. \tag{23}$$

As we note in the text, δ_3, the population regression coefficient of the interaction between female and measured fraction of female teachers, is attenuated relative to the true causal effect β_3 by a conventional measurement error attenuation factor. Similarly, the coefficient δ_1 on the female dummy variable has a familiar upward bias relative to the true effect β_1, reflecting the "rotation" of the observed regression line around the mean of the data. Finally, if $\beta_3 > 0$, we obtain a *negative* coefficient on the main effect of FFT, δ_2 (recall that we are assuming $\beta_2 = 0$). This bias becomes larger, the larger is the variation in the county-level mean relative to the total variation in observed fraction of female teachers, and the larger is the "noise" component of the within-county variance in measured fraction female teachers.

Enumeration-District Design

In our ED design the assumed data generating process is

$$y_i = \beta_0 + \beta_1 D_i + \beta_3 D_i F^*_{d(i)} + v_{d(i)} + e_i,$$

where F^*_d is the true FFT in the enumeration district and $v_{d(i)}$ is the district component. Ignoring differences in the sizes of EDs, the same coefficients β_1 and β_3 appear in a model relating the difference in mean outcomes between females and males in the same ED to the true FFT in the ED:

$$E[y_i|D_i = 1, d(i) = d] - E[y_i|D_i = 0, d(i) = d] = \beta_1 + \beta_3 F^*_d + \xi_d.$$

If instead we relate the difference in means to the *observed* FFT in the district, F_d, we obtain a population regression model

$$E[y_i|D_i = 1, d(i) = d] - E[y_i|D_i = 0, d(i)] = \gamma_1 + \gamma_3 F_d + \psi_d.$$

Given our assumption that $F_i \equiv F_{d(i)} = F^*_{d(i)} + \phi_i$, with $\phi_i \perp F^*_i$, we have conventional attenuation bias in the estimation of our key parameter of interest:

$$\gamma_3 = \beta_3 \frac{\mathrm{var}[F^*_d]}{\mathrm{var}[F_d]}.$$

This attenuation bias is the same as in the county model.

Within-Family Design

With the family model the assumed data generating process is

$$y_i = \beta_0 + \beta_1 D_i + \beta_3 D_i F^*_{g(i)} + w_{g(i)} + e_i$$

where $g(i)$ is the index of the family for individual i. Assuming for simplicity that each family has one girl and one boy, and noting that both siblings are in the same enumeration district $d(g)$, estimating this model is equivalent to estimating a model for the difference in outcomes between a female sibling (denoted by superscript f) and her brother (denoted by superscript m):

$$\Delta_g = y^f_g - y^m_g = \beta_1 + \beta_3 F^*_{d(g)} + \tau_g.$$

If we replace the true FFT with the observed value in the appropriate ED we obtain the following population model

$$\Delta_g = \pi_1 + \pi_3 F_{d(g)} + e_g.$$

We assume the same measurement-error structure as in the previous two models, and observe the same attenuation bias,

$$\pi_3 = \beta_3 \frac{\text{var}[F^*_d]}{\text{var}[F_d]}.$$

B.2 Wisconsin Case Study

To get a sense of the plausible level of measurement-error bias in our estimates, we undertake a "case study," in which we evaluate the teacher gender composition for students in Wisconsin using 1940 Census data *and* administrative data recorded in the 1938/1939 *Wisconsin School Directory.* These latter records are valuable for our purposes because they provide the names of all teachers in high schools and "graded schools" (i.e., schools with at least two teachers). There appear to be few similar publications for other states.

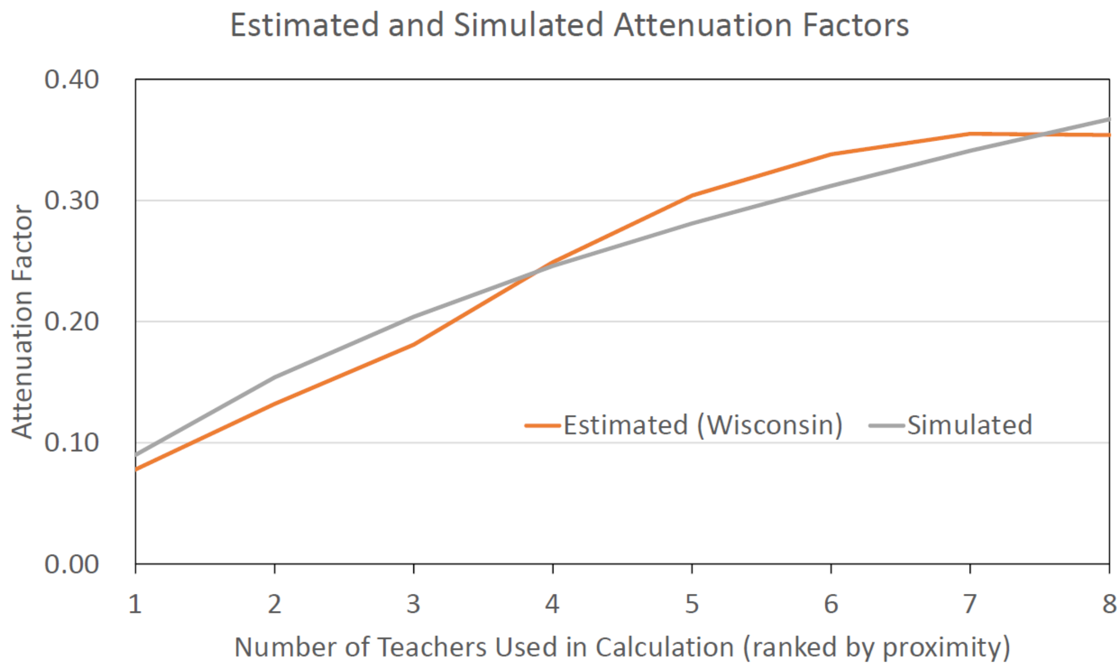
Before proceeding, we estimate our core set of models—of FFT on educational attainment and activities in 1940—for the state of Wisconsin only. The goal is to check if Wisconsin is a representative state in terms of these estimated relationships. See Table 11, which appears at the end of this appendix. The top row, with a dependent variable "on track for 9th grade" reports estimates for Wisconsin that are similar to the rest of the country (see the top row of Table 4). It appears that the impact of FFT on female students is indeed similar in Wisconsin to other states.[39]

[39] This same is not true for "on track for 8th grade," because in Wisconsin 8th grade completion was near-universal (so there was little variation left to analyze). Estimates for other outcomes are somewhat noisy, but broadly in line with the rest of the country.

We begin by using the administrative records to construct a new FFT variable for students in rural Wisconsin, for a substantial set of enumeration districts (rural enumeration districts and small towns with $<$ 2000 inhabitants). We matched the city names listed in the directory as the location of high schools, and graded schools to the names of townships/towns provided in Census descriptions of enumeration districts. Two research assistants then independently recorded the number of female teachers in each school. In nearly all instances they were able to deduce teacher gender from the teacher's name. We restricted the sample to enumeration districts where both research assistants were able to locate the enumeration district; we dropped a few enumeration districts for which there were name/location ambiguities. We denote our administrative records-based FFT, F_d^A.

As described in Section 4.4 of the text, we return to our Census data, which we use to construct alternative FFT measures, based on the "closest k" teachers. We use $k = 1$ through $k = 8$. Intuitively, if we let k be very small, it is likely that included teachers indeed teach at the child's school, but a small k guarantees that F_d^k will be a noisy measure of the student's full set of teachers. As we increase k, the precision of FFT improves *if* marginal teachers indeed teach in the student's school; unfortunately, though, as k increases, it becomes increasingly unlikely that the marginal teacher does teach at the student's school.

Next we estimate a series of regressions in which we regress each Census-based measure on F_d^{A}. These coefficients can be interpreted as estimated attenuation factors. Results are reported as entries in column (1) of Table 5, and are graphed (in orange) in the Figure below. As expected, the attenuation factor is smallest when we have only 1 teacher (attenuation bias is severe), and increases as we add more teachers, up to a point.



Note: Simulation assumes true fraction female teachers in ED is the sum of a county effect and an ED effect; that the relevant attenuation effect is from a model with county fixed effects; and that there is declining probability that teachers who are further from student actually teaches in the student's ED.

We set out a simple model of the measurement error process depicted in the preceding paragraphs. We start with 3000 counties, each of which has 20 districts. We suppose that true FFT varies across counties, and across districts within counties as follows: in county c, FFT is μ_c, drawn from $\mu_c \sim N(0.70, 0.10)$, and in district d within county $c(d)$, true FFT is $\pi_d = \mu_{c(d)} + e_d$, where $e_d \sim N(0, 0.14)$.[40]

Teachers typically live in their district, but some live nearby outside of their districts; "close by" teachers are more likely to be district teachers than those further away. With this in mind, we suppose teacher j in district d has gender F_{dj} (which is 1 if female, and 0 if male) according to

$$
\begin{aligned}
F_{dj} &\sim \text{Bernoulli}(\mu_{c(d)}) \quad \text{with probability } \theta_j, \text{and} \\
F_{dj} &\sim \text{Bernoulli}(\mu_{\pi_d}) \quad \text{with probability } \theta_j, \text{where} \\
\theta_j &= \kappa(1 - \alpha^{j-1}),
\end{aligned}
$$

with $0 < \kappa < 1$ and $0 < \alpha < 1$. Notice that with probability θ_j the teacher's gender is drawn from the county-wide distribution, and with probability $1 - \theta_j$ it is a drawn from

[40]These numbers accord roughly with our data.

the district-specific distribution. In our set-up, teachers who are closest are most likely to teach in the district.

For our simulations we use $\kappa = 0.75$ and $\alpha = 0.80$. Thus, for the closest teacher ($j = 1$), $\theta_j = 0$, and for the 2nd closest it is 0.15, while for the 8th closest it is 0.59. We construct simulated "observed FFT" measures for the k closest teachers as

$$m_{dk} = \frac{1}{k}\sum_{j=1}^{k} F_{dj}.$$

We try $k = 1, \ldots, 8$.

Finally, we run a OLS regression with county fixed effects:

$$\pi_d = \lambda_J m_{dk} + \text{county FE}$$

The (simulated) attenuation factor for k teachers is $\hat{\lambda}_k$, the within-county estimate of the coefficient on the "observed FFT" based on the closest k teachers. We run regressions for all 8 values of k, and plot the resulting attenuation factors in gray in the Figure above.

In our empirical analyses we find that even in the best-case situations (e.g., when we use $k = 8$ or when we use the ED-wide average) the attenuation factors are quite low (less than 0.4). This means that our measurement-corrected estimates are more than 2.5 times their OLS counterparts. Our simulation exercise provides a helpful guide to understanding how this substantial measurement error arises, and it gives us increased confidence in the approach we use to correct for this measurement error.

Table 11: Effect of Local FFT on Educational Attainment and Activities in 1940, Wisconsin

		Estimated effect of local fraction of female teachers (FFT)					
		County FE Model		ED FE model, interaction effects		Household FE model interaction effects	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)	Overall sample (6)	Less-educ. parents (7)
On track for 9th grade	72.9	1.33	3.45	3.91	4.81	3.24	3.51
		(1.36)	(1.68)	(1.72)	(2.11)	(2.14)	(2.68)
On track for 8th grade	94.4	1.35	-0.89	-1.15	-1.10	0.05	-0.78
		(1.04)	(1.19)	(0.94)	(1.24)	(1.90)	(2.50)
Enrolled at Census date	61.4	1.89	1.08	0.91	1.32	1.03	2.93
		(1.50)	(1.52)	(1.71)	(2.07)	(2.39)	(2.36)
Working at Census date	22.4	-3.81	2.11	2.29	1.57	6.94	6.09
		(1.72)	(2.20)	(2.17)	(2.58)	(3.15)	(3.75)
Not enrolled or working	19.3	0.94	-2.83	-2.67	-2.99	-7.33	-9.12
		(1.06)	(1.94)	(2.15)	(2.53)	(2.59)	(3.27)
Counties/EDs		71		1,702	1,689	71	71
Observations		67,883		67,883	47,322	16,605	11,790

Notes: Table contains estimated coefficients for models fit to the dependent variable in row headings, using data for White adolescents age 15–18 in 1940 IPUMS, for Wisconsin only. Otherwise this table reports models exactly as in Table 4.

Appendix C Supplementary Material

We provide additional details and analyses, in support of the research presented above.

C.1 Control Variables Included in Regressions

Our regressions include many control variables. We provide the list here:

- Child Characteristics. We include a gender indicator ($D = 1$ if the child is a daughter), and also age indicators (e.g., for 15, 16, 17, or 18) and a Hispanic indicator, which are also interacted with child gender.
- Parental and Household Characteristics. We have indicator variables for parental education, parental age, number of siblings, farm status, only mother present, only father present, Hispanic status, farm residence, small town residence, and residence outside a metro area. Also included is the difference between the mother's and father's education. In addition, we include interactions of these variables with child gender.
- Characteristics of Neighbors. Using the closest 50 neighboring households, we construct neighbors' average fathers' education, mothers' education, fathers' income, mothers' employment, employment rate of men aged 19–24, employment rate of women aged 19–24, percent residing on a farm, and emergency employment among fathers. Also included are interactions of these variables with child gender.
- Characteristics of Nearby Teachers. We include teacher variables that are plausibly be related to teacher effectiveness: average wage, age, and years of schooling, and interactions of these variables with child gender.
- County Variables. We include variables related to county prosperity (Table 12). Our model has county fixed effects, which absorb the main effect of these variables; we are including the child-gender interaction effects only.

C.2 Additional Tables

Table 12 provides sources for county-level variables used in our analyses.

Table 13 provides estimated coefficients for the county fixed effects model, with "on track for 9th grade" as the dependent variable, estimated for White adolescents in 1940. This is the model from the first row of Table 4, columns (2) and (3). Table 14 similarly provides estimated coefficients for the county fixed effects model, with "on track for 8th grade" as the dependent variable, estimated for Black adolescents in 1940. This is the model from the second row of Table 7, columns (2) and (3).

Table 15 repeats the analysis from Table 4, but only for individuals for whom a PIK was assigned. Table 16 repeats Table 4 but only for adolescents who lived in an enumeration district with at least two teachers.

Table 12: County-Level Variables, with Sources

County-level variable	Source
Average occupational score, men aged 35–54 Average occupational score, women aged 35–54 Average number of children, women aged 21–30 Indicator, city with population 25,000+ in county Doctor/population ratio for Whites	1940 IPUMS, 100 percent sample
Minimal distance to a normal school	Author's calculations, using locations listed in *The Factors Operating in the Location of State Normal Schools* by Harry C. Humphreys (Columbia University, 1923)
Minimal distance to a 4-year college	Author's calculations, using data for institutions founded before 1940, *College Scorecard,* US Department of Education
Fraction of White births delivered in hospital Infant mortality for Whites, 1941	Data provided by Michael Haines
Population density (persons per square mile), 1940 % of rural farm dwelling units with electric lighting, 1940 % of rural farm dwelling units with running water, 1940 % of workers employed in agriculture, 1940 % of dwelling units with private bathtub or shower, 1940 Sum of major war supply contracts and projects ($)	*County and City Data Book*[a]

[a]US Bureau of the Census, *County and City Data Book,* Consolidated File: County Data, 1947–1977. Distributed by Inter-University Consortium for Political and Social Research, 2012-09-18. https://doi.org/10.3886/ICPSR07736.v2

Table 13: Estimated Coefficients for Control Variables, Regression with Dependent Variable, "On Track for 9th Grade," White 15–18 Year Olds in 1940

Effect of FFT	
FFT ("main effect")	-1.060
	(0.262)
FFT × D ("interaction effect")	3.260
	(0.277)
Individual, parental and household characteristics	
Daughter (D)	-8.050
	(4.410)
Only mother present	-6.170
	(0.170)
Only mother present × D	3.210
	(0.182)
Only father present	-5.550
	(0.225)
Only father present × D	-0.527
	(0.296)
Hispanic	-3.340
	(1.080)
Hispanic × D	-5.350
	(0.780)
Farm	0.230
	(0.184)
Farm × D	1.330
	(0.170)
Small town	0.109
	(0.247)
Small town × D	-0.453
	(0.182)
Difference, mother's and father's education	-2.060
	(0.031)
Difference, mother's and father's education × D	0.608
	(0.023)
Outside metro area × D	0.352
	(0.342)

Continued on next page

Table 13 – *Continued from previous page*

Neighborhood characteristics	
Neighboring fathers' education	0.723
	(0.071)
Neighboring fathers' education × D	-0.701
	(0.002)
Neighboring mothers' education	1.420
	(0.075)
Neighboring mothers' education × D	0.809
	(0.082)
Neighboring fathers' income	-0.052
	(0.027)
Neighboring fathers' income × D	-0.046
	(0.017)
Neighboring mothers' employment	-6.630
	(1.000)
Neighboring mothers' employment × D	4.580
	(0.951)
Employment rate, neighboring men aged 19–24	-3.030
	(0.267)
Employment rate, neighboring men aged 19–24 × D	1.630
	(0.276)
Employment rate, neighboring women aged 19–24	1.520
	(0.267)
Employment rate, neighboring women aged 19–24 × D	-1.420
	(0.278)
Wage of neighboring men aged 19–24	0.252
	(0.020)
Wage of neighboring men aged 19–24 × D	-0.157
	(0.023)
Wage of neighboring women aged 19–24	0.099
	(0.019)
Wage of neighboring women aged 19–24 × D	-0.024
	(0.023)
Percent of neighbors residing on farm	8.740
	(0.300)
Percent of neighbors residing on farm × D	2.110
	(0.276)

Continued on next page

Table 13 – *Continued from previous page*

Emergency employment among neighboring fathers	-0.943
	(1.010)
Emergency employment among neighboring fathers $\times$ D	-4.680
	(0.902)
Neighborhood-county imputed data	-1.200
	(0.105)
Other teacher characteristics	
Teacher wages	0.089
	(0.026)
Teacher wages $\times$ D	0.109
	(0.023)
Teacher age	-0.069
	(0.011)
Teacher age $\times$ D	0.024
	(0.011)
Teacher years of schooling	0.459
	(0.055)
Teacher years of schooling $\times$ D	0.067
	(0.057)
Gender interactions on county characteristics	
Occupational score of men in county $\times$ D	0.150
	(0.085)
Occupational score of women in county $\times$ D	0.211
	(0.060)
Percent of births in hospital $\times$ D	0.597
	(0.620)
Doctor-population ratio $\times$ D	0.539
	(0.132)
Infant mortality $\times 100 \times D$	-0.241
	(0.476)
Fraction employed in agriculture $\times$ D	0.025
	(0.012)
Population density $\times 1000 \times D$	0.101
	(0.125)
Percent of households with bathtub $\times$ D	-0.075
	(0.015)

Continued on next page

Table 13 – *Continued from previous page*

Percent of farms with water $\times$ D	0.021
	(0.009)
Percent of farms with electricity $\times$ D	0.029
	(0.006)
War economy spending $\times 1000 \times D$	0.037
	(0.109)
Distance to normal school $\times 100 \times D$	-0.280
	(0.155)
Distance to college $\times 100 \times D$	0.002
	(0.308)
Number of children, women aged 21–30 $\times$ D	-1.730
	(0.655)
Indicator, city in county $\times$ D	0.082
	(0.314)
County-state imputed data	-5.74
	(20.70)
Constant and household/child fixed effects included	
Constant	0.191
	(0.029)
Parental education FE	
Parental education FE $\times$ D	
Parental age FE	
Parental age FE $\times$ D	
Child age FE	
Child age FE $\times$ D	
Number of siblings FE	
Number of siblings FE $\times$ D	

Notes: Table shows all coefficient estimates for the regression presented in the first row of Table 4, columns (2) and (3). See Tables 2 and 4 for additional detail on the samples and models. See Table 12 (this Appendix) for a description of county variables and sources. The regression includes county fixed effects, which absorb the main effects of county variables; we include these variables interacted with a daughter indicator. Standard errors are clustered at the county level.

Table 14: Estimated Coefficients for Control Variables, Regression with Dependent Variable, "On Track for 8th Grade," Black 15–18 Year Olds in 1940

Effect of FFT	
FFT ("main effect")	-2.809
	(1.189)
FFT $\times$ D ("interaction effect")	3.048
	(1.171)
Individual, parental and household characteristics	
Daughter (D)	27.48
	(12.88)
Only mother present	-12.14
	(0.417)
Only mother present $\times$ D	1.136
	(0.593)
Only father present	-7.010
	(0.688)
Only father present $\times$ D	-0.878
	(0.978)
Hispanic	-4.318
	(3.558)
Hispanic $\times$ D	0.928
	(5.372)
Farm	-1.661
	(0.652)
Farm $\times$ D	3.540
	(0.827)
Small town	5.356
	(1.176)
Small town $\times$ D	-4.763
	(1.107)
Difference, mother's and father's education	-2.156
	(0.078)
Difference, mother's and father's education $\times$ D	0.424
	(0.100)
Outside metro area $\times$ D	1.399
	(1.540)

Continued on next page

Table 14 – *Continued from previous page*

Neighborhood characteristics	
Neighboring fathers' education	1.376
	(0.213)
Neighboring fathers' education × D	-1.293
	(0.256)
Neighboring mothers' education	1.554
	(0.229)
Neighboring mothers' education × D	1.108
	(0.252)
Neighboring fathers' income	0.003
	(0.001)
Neighboring fathers' income × D	-0.003
	(0.001)
Neighboring mother's employment	-2.694
	(1.481)
Neighboring mothers' employment × D	3.021
	(1.482)
Employment rate, neighboring men aged 19–24	-8.894
	(1.097)
Employment rate, neighboring men aged 19–24 × D	3.983
	(1.255)
Employment rate, neighboring women aged 19-24	-1.410
	(0.862)
Employment rate, neighboring women aged 19–24 × D	-1.013
	(1.029)
Wages of neighboring men aged 19–24	0.001
	(0.001)
Wages of neighboring men aged 19–24 × D	-0.000
	(0.001)
Wages of neighboring women aged 19–24	0.003
	(0.001)
Wages of neighboring women aged 19–24 × D	-0.001
	(0.002)
Percent of neighbors residing on farm	-5.253
	(1.088)
Percent of neighbors residing on farm × D	0.056
	(1.214)

Continued on next page

Table 14 – *Continued from previous page*

Emergency employment among neighboring fathers	-4.574
	(4.295)
Emergency employment among neighboring fathers × D	7.865
	(4.298)
Neighborhood-county imputed data	0.793
	(0.380)
Other teacher characteristics	
Teacher wages	0.577
	(0.162)
Teacher wages × D	-0.350
	(0.131)
Teacher age	0.050
	(0.036)
Teacher age × D	-0.069
	(0.037)
Teacher years of schooling	0.214
	(0.185)
Teacher years of schooling × D	0.182
	(0.169)
Gender interactions on county characteristics	
Occupational score of men in county × D	-0.747
	(0.279)
Occupational score of women in county × D	-0.568
	(0.191)
Doctor-population ratio × D	0.991
	(0.866)
Infant mortality × D	0.009
	(0.008)
Fraction employed in agriculture × D	-0.079
	(0.031)
Population density × 1000 × D	0.893
	(0.488)
Percent of households with bathtub × D	-0.120
	(0.056)
Percent of farms with water × D	-0.0103
	(0.062)

Continued on next page

Table 14 – *Continued from previous page*

Percent of farms with electricity × D	-0.032
	(0.016)
War economy spending × D	0.003
	(0.004)
Distance to HBCU ×D	-0.005
	(0.008)
Number of children, women aged 21–30 × D	1.046
	(0.531)
Indicator, city in county × D	0.190
	(1.610)
Constant and child/household fixed effects included	
Constant	17.05
	(7.64)
Parental education FE	
Parental education × D	
Parental age FE	
Parental age FE × D	
Age FE	
Age FE × D	
Number of siblings FE	
Number of siblings FE × D	

Notes: Table shows all coefficient estimates for the regression presented in the second row of Table 7, columns (2) and (3). See Tables 6 and 7 for additional detail on the samples and models. See Table 12 (this Appendix) for a description of county variables and sources. The regression includes county fixed effects, which absorb the main effects of county variables; we include these variables interacted with a daughter indicator. Standard errors are clustered at the county level.

Table 15: Effect of Local FFT on Education Attainment and Activities in 1940, White Adolescents Assigned PIKs

		Estimated effect of local fraction of female teachers (FFT)					
		County FE model		ED FE model, interactions only		Household FE model, interactions only	
	Mean of dep. var. (1)	Main effect (2)	Interaction effect (3)	Overall sample (4)	Less-educ. parents (5)	Overall sample (6)	Less-Educ. parents (7)
On track for 9th grade	78.37	-1.31	3.85	4.18	5.40	3.14	3.89
		(0.28)	(0.32)	(0.31)	(0.42)	(0.52)	(0.70)
On track for 8th grade	88.41	-0.43	1.67	1.82	2.38	1.30	1.64
		(0.21)	(0.24)	(0.24)	(0.34)	(0.42)	(0.58)
Enrolled at Census date	67.25	-1.42	3.72	4.12	4.84	3.56	3.80
		(0.30)	(0.33)	(0.33)	(0.44)	(0.60)	(0.78)
Working at Census date	19.38	-0.35	-0.61	-0.72	-0.74	-0.20	-0.42
		(0.32)	(0.38)	(0.37)	(0.47)	(0.65)	(0.81)
Not enrolled or working	16.79	1.16	-2.66	-2.88	-3.59	-2.45	-2.51
		(0.27)	(0.42)	(0.38)	(0.52)	(0.65)	(0.85)
Counties/EDs		3,000		60,000	60,000	3,000	2,900
Observations*		1,347,000		1,347,000	836,000	209,000	132,000

Notes: This table replicates Table 4 but includes only individuals assigned a PIK (using Census internal records). *Sample sizes are approximate. See notes for Table 4 for other details.

Table 16: Effect of Local Fraction of Female Teachers on Education Attainment and Activities in 1940, Whites with At Least Two Nearby Teachers

		Estimated effect of local fraction of female teachers (FFT)					
		County FE model		ED FE model, interaction effects		Household FE model, interaction effects	
	Mean of Dep. Var. (1)	Main Effect (2)	Interaction w/ Female (3)	Overall Sample (4)	Less-Educ. Parents (5)	Overall Sample (6)	Less-Educ. Parents (7)
On track for 9th Grade	75.93	-1.42	3.73	4.17	5.45	4.18	5.66
		(0.31)	(0.31)	(0.30)	(0.40)	(0.47)	(0.60)
On track for 8th Grade	86.16	-0.50	1.65	1.78	2.34	1.86	2.71
		(0.24)	(0.25)	(0.24)	(0.33)	(0.41)	(0.53)
Enrolled at Census date	65.54	-1.35	3.62	4.11	5.16	3.74	5.14
		(0.33)	(0.32)	(0.32)	(0.40)	(0.54)	(0.64)
Working at Census date	18.52	-0.36	-1.06	-1.16	-1.70	-0.85	-1.29
		(0.36)	(0.43)	(0.40)	(0.50)	(0.62)	(0.76)
Not enrolled or working	19.06	1.14	-2.31	-2.57	-3.26	-2.28	-3.26
		(0.32)	(0.46)	(0.41)	(0.53)	(0.65)	(0.82)
Counties/EDs		3046		52,806	51,725	3,020	2,991
Observations		2,097,652		2,097,652	1,342,482	478,028	318,480

Notes: This table replicates Table 4 but excludes individuals in EDs with one teacher only.
1940 IPUMS data. See notes for Table 4 for other details.