NBER WORKING PAPER SERIES

TEACHER EXPECTATIONS MATTER

Nicholas W. Papageorge
Seth Gershenson
Kyung Min Kang

Working Paper 25255
http://www.nber.org/papers/w25255

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2018

We gratefully acknowledge helpful comments from conference participants at the North American Meetings of the Econometric Society and the IZA Junior-Senior Labor Economics Symposium. Stephen B. Holt provided excellent research assistance. For helpful suggestions we thank Tim Bartik, Barton Hamilton, Robert Moffitt, Robert Pollak, Yingyao Hu, Victor Ronda, Richard Spady, and seminar participants at Washington University in St. Louis, the W.E. Upjohn Institute and Urban Institute. The usual caveats apply. Papageorge gratefully acknowledges that this research was supported in part by a grant from the American Educational Research Association (AERA). AERA receives funds for its "AERA Grants Program" from the National Science Foundation under NSF Grant #DRL-0941014. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Teacher Expectations Matter
Nicholas W. Papageorge, Seth Gershenson, and Kyung Min Kang
NBER Working Paper No. 25255
November 2018
JEL No. D84,I2,J15

ABSTRACT

We develop and estimate a joint model of the education and teacher-expectation production functions that identifies both the distribution of biases in teacher expectations and the impact of those biases on student outcomes via self-fulfilling prophecies. Our approach leverages a unique feature of a nationally representative dataset: two teachers provided their educational expectations for each student. Identification of causal effects exploits teacher disagreements about the same student, an idea we formalize using lessons from the measurement error literature. We provide novel, arguably causal evidence that teacher expectations affect students' educational attainment: Estimates suggest an elasticity of college completion with respect to teachers' expectations of about 0.12. On average, teachers are overly optimistic about students' ability to complete a four-year college degree. However, the degree of over-optimism of white teachers is significantly larger for white students than for black students. This highlights a nuance that is frequently overlooked in discussions of biased beliefs: less biased (i.e., more accurate) beliefs can be counterproductive if there are positive returns to optimism or if there are socio-demographic gaps in the degree of teachers' optimism; we find evidence of both.

Nicholas W. Papageorge
Department of Economics
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
and IZA
and also NBER
papageorge@jhu.edu

Seth Gershenson
School of Public Affairs
American University
4400 Massachusetts Avenue NW
Washington, DC 20016-8070
IZA
gershens@american.edu

Kyung Min Kang
Johns Hopkins University
Department of Economics
3400 N. Charles St.
544E Wyman Bldg.
Baltimore, MD 21218
kmkang@jhu.edu

A data appendix is available at http://www.nber.org/data-appendix/w25255

1 Introduction

At least since Becker (1964) cast schooling as an investment in human capital, economists have sought to understand the factors that drive variation in educational outcomes. Socio-demographic gaps in educational attainment have received particular attention, as education facilitates upward economic and social mobility across generations (Bailey and Dynarski, 2011) and increased earnings (Card, 1999).[1] Attainment gaps are especially concerning if they reflect sub-optimal investments in human capital by under-represented or historically disadvantaged groups (e.g., racial minorities).

Teacher expectations constitute one potentially important, but relatively understudied, educational input that might contribute to socio-demographic gaps in educational attainment. Despite pervasive views that teacher expectations matter, however, it is difficult to credibly identify their causal effects on student outcomes (Brophy, 1983; Jussim and Harber, 2005; Ferguson, 2003). The reason is that observed correlations between teacher expectations and student outcomes could reflect accurate forecasts or a causal relationship. In the first case, expectations do not drive student outcomes, but instead reflect the same factors that drive educational attainment. In the second case, a causal impact arises if incorrect (i.e., biased) teacher expectations create self-fulfilling prophecies in which investments made in or by students are altered, thereby leading to outcomes that resemble teachers' initially incorrect beliefs (Loury, 2009; Glover et al., 2015).[2]

In this paper, we provide evidence that teacher expectations affect educational attainment. To identify causal effects, we exploit a unique feature of a nationally-representative longitudinal dataset: two teachers provide their educational expectations for each student.[3] When teachers disagree about a particular student, which they frequently do, they provide within-student, within-semester variation in expectations, which we argue below is conditionally random. We leverage this variation to identify the impact of expectations on educational attainment. Intuitively, our analysis uses one teacher's expectation to control for unobserved factors that both teachers use to form expectations and which affect student educational at-

[1] Among other outcomes, education also affects civic engagement (Dee, 2004a; Milligan et al., 2004), health (Grossman, 2006) and crime (Lochner and Moretti, 2004; Machin et al., 2011).

[2] We are not the first to examine bias and self-fulfilling prophecies in the classroom. Rist (1970) provides a rather harrowing account of how subjective teacher perceptions, driven largely by social class, affected how both teachers and students behaved in the classroom. Eventually, these behaviors produced student outcomes that corresponded to the teachers' initial and negative beliefs about students from lower social classes. Both Jussim and Eccles (1992) and Jussim and Harber (2005) recognize how accuracy and self-fulfilling prophecies could contribute to a correlation between expectations and outcomes.

[3] Previous research has leveraged this feature to estimate the effect of student-teacher racial match on teachers' perceptions and expectations via student-fixed effects models (Dee, 2005; Gershenson et al., 2016).

tainment.[4] Our approach addresses the fundamental endogeneity problem that arises if teacher expectations reflect omitted variables that also drive education outcomes (Gregory and Huang, 2013; Boser et al., 2014).

The current study begins by documenting several interesting patterns in the teacher expectations data. First, we find that teacher expectations predict student outcomes, though teachers appear to be optimistic.[5] Second, teacher expectations respond in expected ways to information that would presumably affect college-going, such as family income, standardized test scores, and ninth-grade GPA. Third, teachers frequently disagree about how far a given student will go in school, which is key to our identification strategy.

Next, we provide evidence that teacher expectations causally affect student outcomes. Our main analyses use OLS regressions and condition on both teachers' expectations. Using this strategy, our estimates provide consistent and compelling evidence of a causal impact of teachers' expectations on the likelihood of college completion. The estimates suggest that the elasticity of the likelihood of college completion with respect to teachers' expectations is about 0.12. In an earlier contribution, Rosenthal and Jacobson (1968) report effects of informing teachers that some randomly selected students are high-aptitude. These students eventually perform better on tests, which provides some evidence for the view that teacher expectations matter in the sense that randomly assigned biases can become self-fulfilling prophecies. Our findings using a longitudinal data set show that these so-called "Pygmalion Effects" can have long-run impacts on educational attainment and that non-experimentally induced teacher expectations can generate self-fulfilling prophecies.

A possible concern is that high teacher expectations not only raise the probability of a college degree, but also induce some students to enroll in post-secondary education without completing a degree. Given low returns to "some college," this would suggest a possible downside to high expectations for some students. Instead, we find that the high expectations raise college graduation rates by lowering the likelihood of dropping out of high school, of obtaining only a high school degree and, moreover, of attaining some post-secondary education without a four-year degree. In other words, high expectations reduce the probability of educational investments with low returns in the labor market. We also provide evidence that the positive impacts of high expectations extend to longer-run outcomes, inducing for

[4]Our identification strategy relates intuitively to the strategy in Ashenfelter and Krueger (1994), who study identical twins with diverging schooling levels to mitigate bias due to omitted variables bias (e.g., ability) in estimating returns to schooling. See Bound and Solon (1999) for further discussion of the use of twins to address endogeneity problems in labor and education economics.

[5]This does not necessarily imply that teachers are on average wrong at the time that expectations information was collected. It is possible that unexpected aggregate shocks occurring after expectations data were collected shifted behavior (Van der Klaauw and Wolpin, 2008).

example higher rates of employment and lower usage of public benefits.

Using a similar research design, we also provide suggestive evidence of mechanisms underlying the impact of high expectations. A possibility is that high expectations do not lead to better teaching or learning, but simply mean that teachers ease students' pathway to college, for example, by writing better college recommendation letters. Our evidence suggests that this is not the sole mechanism driving our results. We show that high expectations translate to changes in three key 12th grade outcomes. Higher 10th grade expectations lead to higher 12th grade GPA's and more time spent on homework for students. Together, these findings suggest that students put forth greater effort in response to high teacher expectations. Moreover, we show that high teacher expectations raise *students' own* expectations in the 12th grade about their educational attainment. This finding is consistent with the view that teacher expectations matter in part because they raise students' own views about their performance and outcomes. Directions of causality among these factors is, however, unclear. Higher GPA or effort could lead to higher student expectations or result from them (or both). Nonetheless, that all three of these 12th-grade outcomes are influenced by 10th-grade teachers' expectations suggests that increased effort, engagement, and aspirations are channels through which teachers' expectations increase long-run educational attainment.

Motivated by our main findings linking teacher expectations to college completion, we develop an econometric model to jointly estimate the production of teacher expectations and student outcomes. The model formalizes the idea that teacher disagreements can be used to identify causal estimates of the impact of teacher expectations on student outcomes. We draw upon lessons from the measurement error literature (Hu and Schennach, 2008) and treat teacher expectations as forecasts, possibly with error, of the objective probability that students will graduate college. This objective probability is treated as a latent factor. The forecast errors (biases) not only explain why teachers disagree, but can also affect student outcomes by initiating self-fulfilling prophecies.[6]

The model serves two key purposes. First, we estimate the impact of expectations on outcomes in a framework that explicitly incorporates how the same factors — some of which are observed by teachers, but not by the econometrician — jointly generate teacher expectations and student educational outcomes. This means we can validate the reduced-form estimates using a different approach. Second, we use the model to recover the distribution of teacher biases, which are treated as forecast error, and defined as the difference between teacher expectations and the latent objective probability of school completion absent the impact of

[6]The techniques used in this literature draw upon the psychometric literature (see e.g., Goldberger (1972) and Jöreskog and Goldberger (1975)), where an aim is to separate measurement error from an underlying latent factor (e.g., depression) captured imperfectly by a set of measurements.

expectations. Recovering the distribution of bias is in part motivated by earlier findings that white teachers' expectations are systematically lower than are black teachers' expectations when both are evaluating the same black student (Gershenson et al., 2016). Yet, it is unclear whether white teachers are too pessimistic, black teachers are too optimistic, or both.

Model estimates corroborate our results on the size of the impact of teacher expectations. Moreover, using the estimated model — in particular, computing differences between teachers' reported expectations and the latent probability of college completion we recover — we examine the distribution of bias for different teacher- and student-race pairs. We show that teachers are on average optimistic: their expectations are higher than the student's objective probability of college completion. However, white teachers are systematically less optimistic about black students. An overlooked nuance is therefore that white teachers are more accurate about black students in that their expectations are closer to the objective probability. However, since higher expectations lead to better outcomes, "accuracy" in this context amounts to a selective lack of optimism that puts black students at a disadvantage.

Our study contributes to research examining how teachers affect student outcomes. A number of studies have shown that teachers are important inputs in the education production function (Chetty et al., 2014b; Hanushek and Rivkin, 2010). Other studies have recognized that same-race teachers are more effective, especially for racial minorities (Fairlie et al., 2014; Dee, 2004b; Gershenson et al., 2018). However, it remains unclear what specific behaviors and characteristics make teachers effective (Staiger and Rockoff, 2010). Our study suggests one possible mechanism: teachers' expectations that potentially affect student outcomes. Teachers might directly impart biased expectations to students or do so indirectly by modifying how they teach stigmatized students (Burgess and Greaves, 2013; Ferguson, 2003; Mechtenberg, 2009). In either case, biased expectations function like a self-fulfilling prophecy that perpetuates educational attainment gaps.[7] These effects might be particularly salient for relatively disadvantaged students who rarely interact with college-educated adults outside of school settings (Jussim and Harber, 2005; Lareau, 2011; Lareau and Weininger, 2008), since a model of costly information acquisition would predict that such students rely on teacher expectations as a primary source of information.

We also contribute to the economic literature on biased information sources and self-fulfilling prophecies. Several studies examine biases in beliefs and human capital investments (Cunha et al., 2013; Wiswall and Zafar, 2015). Regarding teacher biases, Lavy and Sand (2015) identify primary school teachers in Israel who have "pro-boy" grading bias by

[7] Fortin et al. (2015) and Jacob and Wilder (2010) examine how students' expectations evolve over time and might explain demographic gaps in achievement.

comparing students' scores on "blind" and "non-blind" exams.[8] More similar to us, Jones and Hill (2018) provide evidence that teacher expectations improve student outcomes, in their case, test scores. Their findings use a different data set and identification strategy and thus complement ours. A key difference is that we examine older students and longer-run outcomes such as college completion and labor supply. In another related study, but in a different context, Glover et al. (2015) study managers' perceptions of cashiers and examine how negative perceptions (implicit biases) against certain demographic groups can lead to lower performance.[9] A key difference in our study is that, instead of negative perceptions, we observe actual expectations over an economic outcome along with realized outcomes. Thus, we are able to explicitly define bias as a deviation of expectations from objective probabilities and then to identify the impact of such bias on outcomes through self-fulfilling prophecies.

More broadly, our paper contributes to research on the importance of subjective expectations. The idea that subjective beliefs (rather than objective probabilities) drive individual behavior is not new (Savage, 1954; Manski, 1993). However, despite mounting evidence that subjective expectations can affect important economic outcomes, they are only recently entering into economic analyses of decision-making (Manski, 2004; Hurd, 2009).[10] One reason is that data on subjective expectations have rarely been collected. Another reason is that it is difficult to assess whether beliefs have causal effects on outcomes absent experimentally-induced exogenous variation. We examine the conditions under which multiple subjective expectations about a single objective probability can be used both to recover a distribution of bias and to identify causal effects of expectations on outcomes. We thus offer a methodology for using observational data to identify biased expectations and self-fulfilling prophecies.

Section 2 describes the data set used in the project and documents some basic facts about the information contained in teacher expectations and how and why teachers disagree. Section 3 presents our main reduced form evidence that teacher expectations affect student outcomes. Section 4 develops and estimates a structural model that formalizes identification off disagreements and can also be used to examine the distribution of bias, including differences by teacher and student race. Section 5 concludes.

[8] Terrier (2015) finds similar effects of gender-based grading bias on short-run achievement and subsequent course-taking in France.

[9] Similarly, Loury (2009) develops an informal model in which taxi drivers incorrectly believe that black passengers are more likely than white passengers to rob them. This belief leads drivers to avoid black passengers. In response, black passengers with no criminal intent find other forms of transportation. This affects the composition of black passengers waiting for a cab so that in equilibrium the original biases become true.

[10] Contributions to this line of work include the studies cited above along with numerous papers linking beliefs to economic behaviors such as voting (Chiang and Knight, 2011; DellaVigna and Kaplan, 2007; Gentzkow and Shapiro, 2006), risky sexual behavior (Delavande and Kohler, 2016), and financial decisions (Hudomiet et al., 2011).

2 Data

In this section, we discuss the data set used in the project and conduct a preliminary data analysis of teacher expectations and student educational attainment. Section 2.1 introduces the 2002 Education Longitudinal Study (ELS 2002). Section 2.2 establishes some key patterns exhibited by teacher expectations variables.

2.1 The ELS 2002

The ELS 2002 is a nationally representative survey of the cohort of U.S. students who entered 10th grade in 2002.[11] The ELS data contain rich information on students' socio-demographic backgrounds as well as secondary and postsecondary schooling outcomes (including educational attainment through 2012, or within 8 years of an "on time" high school graduation). Students were sampled within schools and school identifiers facilitate within-school (school fixed effects (FE)) analyses. The data also contain a number of observed school and teacher characteristics, including teachers' experience, demographic background, credentials, and expectations and perceptions of specific students.

The main analytic sample is restricted to the 6,060 students for whom the above-mentioned variables are observed.[12] Because there are two teacher expectations per student, the analytic sample contains 12,130 student-teacher pairs.[13] Table 1 summarizes the students who compose the analytic sample. Column (1) does so for the full sample and columns (2)-(5) do so separately by student race and sex. The outcome of interest, students' educational attainment, is summarized in three ways: percentage of students who earn a four-year college degree (or more), percentage of students who fail to complete high school, and average years of schooling. About 45% of students in the sample completed a four-year degree, though whites and females were significantly more likely to do so than blacks and males, respectively.[14] This is consistent with demographic gaps in educational attainment observed in

[11] The ELS data are collected, maintained, and made available to researchers by the National Center for Education Statistics. See `https://nces.ed.gov/surveys/els2002`.

[12] All sample sizes are rounded to nearest ten in accordance with NCES regulations for restricted data. The instrumental variables analysis described below uses a further restricted sample, for whom a wider range of teacher-perception variables are observed *and* in which teachers taught multiple students.

[13] This does not mean that there are 12,130 different teachers, as some students share one or both teachers. As best we can tell, the data set contains approximately 3,000 unique teachers. However, this total is likely mis-measured because the data set lacks teacher identifiers. As explained below, we use a probabilistic matching procedure to determine which teacher observations come from the same teacher.

[14] The college completion rate in the analytic sample (45%) is larger than that in the full ELS sample (30%). The latter is consistent with national estimates from other data sources for this cohort (Bailey and Dynarski, 2011). This positive selection into the analytic sample is mirrored in other student attributes, such as test scores and socioeconomic background. It is driven by the restriction that two teachers' expectations

other datasets (Bailey and Dynarski, 2011; Bound and Turner, 2011; Cameron and Heckman, 2001). The racial gaps in educational attainment are particularly stark, as whites were about 20 percentage points (69%) more likely to graduate from college than blacks while blacks were twice as likely as whites to fail to complete high school. Racial differences in educational attainment are also apparent in Figure 1, which provides a histogram for educational attainment categories for the full sample and then separately for blacks, whites, males, and females.

The key teacher-expectation variable is based on teachers' responses to the following question: "*How far do you think [STUDENT] will go in school*?" Teachers answered this question by selecting one of seven mutually exclusive categories.[15] In most of our subsequent analysis, we exploit a unique feature of the ELS 2002's design: two teachers, one math and one ELA, provided their subjective expectations and perceptions of each student.[16] Teachers' expectations are summarized in the next section of Table 1. Overall, about 64% of teachers expected the student to complete a four-year college degree. This suggests that teachers, on average, are too optimistic about students' college success, as only 45% of students complete a four-year degree. This over-optimism is apparent in each demographic group, though teachers' expectations for black students are significantly lower than for white students, as are expectations for male students relative to females. This points to an interesting feature in the data that foreshadows our results: teacher expectations for black students are not necessarily low relative to observed outcomes. Rather, they are less inflated relative to observed outcomes compared to expectations for white students. Still, observed racial and sex gaps in expectations are consistent with the patterns in actual educational attainment

are observed for each student. This necessarily excludes students in remedial or special education tracks who do not have distinct math and reading teachers. Thus, the positive selection observed in the analytic sample arguably yields a sample of students for whom teacher expectations about college completion are most relevant. About half of the ELS respondents are missing either educational outcomes or at least one teacher expectation. A break down of sample selection is reported in Table S1 in Appendix A. Sample means for the variables used in the analysis are reported separately for individuals in and out of the analytic sample, are are often significantly different. However, point estimates for the baseline specifications reported in Table 5 are robust to using either the full or restricted sample. Table S2 in Appendix A shows point estimates using the full sample. This, together with the similarity of IV and OLS estimates, suggests that selection occurs primarily on observables. Importantly,then, our preferred specification includes a rich set of teacher controls, student SES and ability measures, and school FE. Indeed, in Table S3 in Appendix A, we show that characteristics for students in our main analytic sample and for students who do not have both teachers' expectations are similar on most dimensions if we control for 9th grade GPA, reading and math test scores, and school fixed effects.

[15]Options were Less than high school graduation; High school graduation or GED only; Attend or complete 2-year college/school; Attend college, 4-year degree incomplete; Graduate from college; Obtain Master's degree or equivalent; Obtain PhD, MD, other advanced degree.

[16]Students do not directly observe teachers' responses to this survey question. However, there are numerous mechanisms through which teachers both directly and indirectly transmit their expectations to students (Mechtenberg, 2009; Gershenson et al., 2016).

described above, suggesting that teachers' expectations are informative. However, while math and ELA teachers' expectations are similar on average, ELA teachers' expectations tend to be slightly higher, particularly among black students. This shows that teachers occasionally disagree about how far a particular student will go in school. Specifically, teachers disagree on slightly more than 20% of students, with math teachers having higher expectation in slightly less than half of those cases. Below, we further investigate the sources of teacher disagreements and consider how such disagreements can be leveraged to identify the impact of expectations on student outcomes.

The final two panels of Table 1 summarize students' academic and socioeconomic characteristics. A comparison of columns (2) and (3) shows that white students have significantly higher test scores, GPAs, and household incomes than black students, as well as better educated mothers, all of which is consistent with longstanding racial disparities in academic performance and socioeconomic status (Fryer, 2010). Another notable difference by student race is in their assigned teacher's race: black students are four to five times as likely as white students to be assigned a black teacher, which is due to non-white teachers being more likely to teach in majority non-white schools (Hanushek et al., 2004; Jackson, 2009). Nonetheless, the majority of students, white and black, have white teachers. Columns (4) and (5) of Table 1 show that girls have higher GPAs and perform better on reading assessments than boys, while boys perform better on math assessments. This is again consistent with the extant literature (Jacob, 2002). Unsurprisingly, there are no significant differences in SES by sex, since boys and girls live in the same neighborhoods and attend the same schools.

Table 2 similarly summarizes the teachers represented in the analytic sample. Overall, 11% of teachers are nonwhite and nonwhite teachers are evenly represented across subjects and sex. The average teacher has about 15 years of experience though 16% of teachers have $\leq$ 3 years of teaching experience. Math teachers are more experienced than ELA teachers, on average, as are male teachers relative to female teachers, and white teachers relative to black teachers. Almost half of teachers have an undergraduate degree in the subject they teach. A similar percentage hold a graduate degree. The bottom panel of Table 2 confirms that black teachers are significantly more likely to teach black students than are teachers from other racial backgrounds. Looking further into racial differences between teachers, columns (4) and (5) show that white teachers, compared to black teachers, are more likely to be male, experienced, and hold teaching certificates, and these differences are statistically significant.[17]

In this section, and for most of the remainder of our study examining teacher expectations

[17] The finding that white teachers are more experienced and more likely to have a teaching certificate is robust across subjects.

and student outcomes, we focus on the college-completion margin because recent research explicitly notes that individuals with some college, but less than a four-year degree, have socioeconomic trajectories that closely resemble those of high school graduates (Lundberg et al., 2016). This choice is also due to the striking patterns observed in Figure 1: blacks are significantly more likely than whites to only complete "some college." This suggests that college completion, relative to college entrance, is an important margin to consider in the analysis of racial attainment gaps. Thus, we define students' educational attainment and teachers' educational expectations for the student in the same way: the student outcome of interest in the primary analyses is an indicator for "student completed a four-year college degree or more" (as of 2012, 8 years removed from an on-time high-school graduation) and the independent variables of interest are indicators for "teacher expects a four-year college degree or more."

2.2 Key Patterns in Teacher Expectation Data

This section establishes three empirical patterns regarding teacher expectations. Section 2.2.1 shows that teacher expectations are predictive of student outcomes (rather than being pure noise). In Section 2.2.2, we focus on the production function of teacher expectations, establishing that teachers respond to student-level characteristics, which are likely to lead to higher educational attainment. Omitting these variables would thus lead to omitted variables bias. Moreover, the fact that teacher expectations respond to observable student-level information suggests that they likely also respond to unobservable factors, so that even after controlling for student characteristics, omitted variables bias remains a concern when relating expectations to outcomes. In Section 2.2.3, we show that teachers frequently disagree when evaluating the same student. As we discuss in the following section, this goes a long way towards alleviating concerns about omitted variables bias and is the basis of our identification strategy.

2.2.1 Teacher Expectations Are Predictive

Figure 2 plots the percentage of students who complete a four-year college degree for each category of teacher expectations, separately for math and ELA teachers. According to the figure, higher expectations are associated with a higher probability of college completion. Interestingly, however, teacher forecasts are subject to error. For example, of students for whom ELA teachers expect some college (but not college completion), roughly 15% go on to obtain a 4-year degree. Forecast errors tend to be in the opposite direction, however: fewer than 60% of students whose math or ELA teachers expect a 4-year degree actually obtain one. This pattern extends to students for whom teachers expect a Masters or other higher degree, who obtain at least a 4-year degree roughly 80% and 85% of the time, respectively. In other words, though teacher expectations are predictive of student outcomes, on average teachers over-estimate educational attainment, which is consistent with patterns in Table 1.

2.2.2 The Production of Teacher Expectations

Understanding the determinants of teacher expectations is a precursor to credibly identifying the impact of those expectations on student outcomes. However, previous analyses of the association between teacher expectations and student outcomes generally pay short shrift to the formation of teacher expectations. Thus one contribution of the current study is a systematic analysis of the teacher expectation production function. We show that factors that would presumably affect educational attainment also produce teacher expectations. We do so by estimating equations of the form

$$T_{ij} = X_i\beta_j + \nu_{ij}, j \in \{M, E\}. \tag{1}$$

The usual suspects are predictive of teacher expectations. Results are found in Table 3. Columns (1)-(3) show that higher income, being white, and high GPA are associated with higher teacher expectations. Notice, when we evaluate these factors jointly in Column (4), we find higher expectations for Asians and lower expectations for Hispanics. Interestingly, if we adjust for parental income and GPA, black students do not face lower expectations. Similar patterns are found for math teacher expectations, but expectations for black students are lower even after we have controlled for 9th grade GPA and household income (Gershenson et al., 2016).

These estimates highlight how the correlation between teacher expectations and student outcomes may reflect how teachers respond to information about students that could affect their educational attainment. If we regress educational attainment onto one teacher's expectation, a positive estimated coefficient is unlikely to be appropriately interpreted as a causal effect. For example, omitting income would lead to an upwardly biased estimate since income presumably drives educational attainment, but is also associated with higher expectations. More generally, results from the production function estimates show that the same factors that drive higher teacher expectations are also likely to drive educational attainment. Omitting such factors thus leads to biased coefficients. Moreover, there are likely to be other factors that teachers observe and which we do not observe that also affect teacher expectations and student outcomes, which would lead to omitted variables bias despite adjusting for observable student characteristics.

2.2.3 Teacher Disagreements

A key pattern in the data is that teachers frequently disagree about a particular student's educational prospects, which we leverage in our identification strategy. The transition matrices reported in Table 4 document the frequency of such disagreements. The modal disagreement is over whether or not students who enter college will earn a 4-year degree, rather than more substantial disagreements. This suggests that disagreements are often subtle, and might hinge on arbitrary factors that do not directly affect student outcomes. For example, all else equal, some teachers might have higher baseline levels of optimism than others.[18] This turns out to be useful and also key to our identification strategy. We explore further how teacher expectations arise in our section on identification.

3 Main Results

A preliminary analysis of the data shows that higher teacher expectations are associated with factors that would presumably also predict higher educational attainment, such as 9th grade GPA and parental income. This pattern is consistent with the idea that, when forming expectations, teachers respond to information about factors that generate higher educational attainment. If omitted from the analysis, some of these factors could generate correlations between expectations and outcomes that should not be assigned a causal interpretation. We have also shown that even though teacher expectations predict educational attainment teachers also disagree a fair amount about a given student.

In this section, we provide arguably causal evidence that higher teacher expectations lead to higher educational attainment. Our main identification strategy leverages teacher disagreements. The reasoning is that disagreements generate within-student, within-semester variation in teacher expectations. To the degree that these disagreements are conditionally random, we can use this variation to estimate causal effects. Section 3.1 presents our main results. Section 3.2 discusses additional student outcomes. Section 3.3 provides evidence of possible mechanisms explaining our results. Section 3.4 discusses threats to identification, which relies on teacher disagreements being conditionally random.

[18]This idea is similar in spirit to Kling (2006), who exploited exogenous variation in judges' baseline sentencing propensities to estimate the impact of incarceration length on labor market outcomes. Specifically, the author used judges' other sentences to instrument for the actual sentence length, as sentence lengths might reflect omitted variables (e.g., the severity of a crime) that presumably affect post-incarceration outcomes. We return to this point when discussing robustness checks in Section 3.4.

3.1 Evidence that Teacher Expectations Matter

Table 5 presents OLS estimates of linear regressions of the form:

$$y_i = \gamma_E T_{Ei} + \gamma_M T_{Mi} + X_i\beta + \epsilon_i, \tag{2}$$

where the T's denote teacher expectations, y denotes student outcomes, and i indexes students.[19] Either γ_E or γ_M can be restricted to equal zero, where E and M index ELA and math teachers, respectively. The vector X includes a progressively richer set of statistical controls, up to and including school or teacher fixed effects (FE). Standard errors are clustered by school, as teachers and students are nested in schools (Angrist and Pischke, 2008).

Columns (1) and (2) of Table 5 report simple bivariate regressions of y on the ELA and math teachers' expectations, respectively. The point estimates are nearly identical, positive, and strongly statistically significant. Of course, these positive correlations cannot be given causal interpretations because there are many omitted factors that jointly predict student outcomes and teachers' expectations (e.g., household income). In subsequent columns of Table 5 we attempt to reduce this omitted-variables bias by explicitly controlling for such factors. In column (3), we simultaneously condition on both teachers' expectations. Interestingly, though both estimates of γ decrease in magnitude, they remain nearly identical to one another and both remain individually statistically significant. The decline in magnitude suggests that one teacher's expectation can be viewed as a proxy for factors that both teachers observe and which could generate a correlation between expectations and outcomes. That both teachers' expectations remain individually significant indicates that there is substantial within-student variation in teacher expectations (i.e., teachers frequently disagree).

It is possible that teacher disagreements are not fully random if we fail to condition on additional information. Therefore, we would expect expectations to become less predictive of outcomes once we control for factors that potentially affect both. Thus, subsequent columns of Table 5 continue to add covariates to the model, which lead to a similar pattern in the estimated γ: the estimated effects of expectations decrease in magnitude, but remain positive, similar in size to one another, and individually statistically significant. The largest

[19]To allay concerns that these results are driven by students with extreme levels of attainment, Table S4 in Appendix B reports OLS estimates of equation (2) for the restricted sample that excludes students who either did not complete high school or who earned a graduate degree. We present OLS estimates of these linear probability models (LPM) for ease of interpretation and to facilitate the inclusion of school and teacher fixed effects. However, estimates of analogous logit and probit models yield similar patterns. Logit estimates are reported in Table S5 in Appendix B. Probit estimates are reported in Table S6 in Appendix B.

drop in the size of the coefficient occurs when we adjust for 9th grade GPA, which suggests that teacher expectations, in particular, disagreements, might exhibit different patterns depending on a student's earlier grades. We return to this point when discussing threats to identification in Section 3.4. One consequence is that we control for 9th grade GPA in our subsequent analyses.[20]

Our preferred specification, which conditions on students' socio-demographic background, past academic performance, and school FE, is reported in column (7). These estimates suggest that conditional on the other teacher's expectation and a rich set of observed student characteristics including sex, race, household income, mother's educational attainment, 9th grade GPA, and performance on math and ELA standardized tests, the average marginal effect of changing a teacher's expectation that a student will complete college from zero to one increases the student's likelihood of earning a college degree by about 15 percentage points.

Column (8) shows that the preferred point estimates are robust to controlling for teacher FE. Specifically, this model controls for ELA-math teacher dyad FE. That is, we compare students who had the same pair of math and ELA teachers.[21] Two caveats to this analysis are of note. First, this approach can only be applied to the subsample of math-ELA teacher dyads that taught multiple students in the ELS 2002 analytic sample. To verify that the teacher-FE results are not driven by this necessary sample restriction, in column (9) we estimate the preferred school-FE specification using the restricted teacher-FE sample, and see that the point estimates are similar. Second, the ELS 2002 does not provide actual teacher identifiers, so we create teacher identifiers using a probabilistic matching process, which is necessarily prone to measurement error. This procedure makes within-school matches based on teachers' race, sex, subject, educational attainment, experience, and college majors and minors. The algorithm is likely to perform well given the relatively large number of observable teacher characteristics and the fact that the sample is limited to teachers of tenth graders; still, the possibility remains that teachers with identical observable profiles are incorrectly coded as being the same teacher. For these reasons, we take the school-FE estimates in column (7) as the preferred baseline estimates, though it is reassuring that the teacher-FE estimates are remarkably similar. Finally, Columns (10) and (11) show that the point estimates are similar in magnitude for white and black students, though the black-sample estimates are

[20] 9th grade GPA is predetermined in the sense that it is fixed before 10th grade teachers form expectations about 10th grade students. Moreover, it is determined prior to student-teacher classroom assignments in the 10th grade, which is important given that most sorting into classrooms is driven by past achievement (Chetty et al., 2014a).

[21] We obtain similar point estimates if we instead condition on two-way teacher-specific FE (one for each subject's teacher) rather than on ELA-math teacher dyad FE. The difference between the teacher-FE strategies occurs when there are two math (or ELA) teachers in a given school in the ELS analytic sample.

less precise, likely due to the smaller sample size.

To interpret the preferred point estimate of 0.14 reported in Column (7), consider that this reflects a change in expectation from 0% chance of completing a college degree to 100% chance of completing a college degree.[22] Such a drastic change in expectations is unlikely to be of policy interest and likely to be an "out-of-sample" change. Rather, the policy-relevant change in teachers' expectations is more likely in the range of a 10 or 20 percentage point increase in the probability that a teacher places on a student completing college, which corresponds with the unconditional black-white gap in expectations shown in Table 1. The corresponding marginal effects of these changes on the likelihood that the student graduates from college are about 1.4 and 2.8 percentage points, respectively. From the base college-completion rate of 45%, these represent modest, but nontrivial, increases in the graduation rate of 3.1 to 6.2%. These effect sizes are remarkably similar to those found in other evaluations of primary-school inputs' impacts on post-secondary outcomes. For example, Dynarski et al. (2013) find that assignment to small classes in primary school increased the probability that students earned a college degree by 1.6 percentage points. Similarly, Chetty et al. (2014b) find that a one-SD increase in teacher effectiveness increases the probability that a student attends at least four years of college between the ages of 18 and 22 by about 3.2%.[23] Still, even with these rich controls and conditioning on the other teacher's expectation, the threat of omitted-variables bias remains. We discuss alternatives to OLS estimation of equation (2) that address this concern in Section 3.4.

3.2 Additional Outcomes

A possible downside of high expectations is an increase in the number of students who enroll in college, but do not obtain a college degree. This could occur if expectations encourage a subset of students to attempt college even though they are unprepared for it. Given relatively low returns to "some college," high expectations could potentially lead to a waste of resources, including students' opportunity costs of time and their financial resources.

To investigate this possibility, we estimate a multinomial logit model (MNL) with three mutually exclusive outcomes: a high school degree or less, college enrollment without a degree, and completion of a college degree. Average partial effects (APE) are reported in Table 6. Similar to estimates presented in Table 5, specifications include increasingly rich sets of controls as we move from the left to the right.[24] Consider estimates in column (6),

[22] 0.13 is the point estimate for the math teacher. The coefficient for the ELA teacher is 0.14.

[23] Chetty et al. (2014b) do not observe actual college completion and instead use this as a proxy.

[24] Ordered-logit models yield similar results. We omit school FE from these models to avoid the incidental parameters problem and computational issues in the MNL. We feel comfortable making this trade-off, as the

which condition on teacher characteristics, student SES, and 9th grade GPA. Consistent with earlier results, we find that higher teacher expectations increase the probability of college completion by 13 percentage points. The concern is that higher expectations also cause an inefficient increase in college enrollments for students who fail to complete college. On average, we find no evidence that this is the case, as we find *declines* of about 6 percentage points in both the probability of obtaining a high school degree or less and of enrolling in but failing to complete college. This suggests that the group of students being induced into enrolling in, but failing to complete, college is small.

We also consider the impact of teacher expectations on additional, longer-run outcomes including employment, marital status, and measures of financial well-being (e.g., home ownership and use of public benefits). These variables are measured 12 years after the baseline survey.[25] For each outcome, we use the preferred specification corresponding to column (7) of Table 5, which conditions on teacher controls, student SES, 9th-grade GPA, and school FE. Moreover, the table includes mean values of each outcome variable along with a joint significance test of the two teachers' expectations. Coefficient estimates tend to be noisy and only marginally statistically significant. However, they are in the expected sign and provide some evidence that the positive impacts of teacher expectations on educational attainment extend to associated longer-run socioeconomic outcomes. For example, high ELA teacher expectations lead to a 5 percentage point increase in the probability of being employed (either full or part time) and a 7 percentage point drop in using public benefits. High expectations also lead to lower probabilities of being married and having children, which suggests that high expectations may lead some individuals to postpone starting a family in order to invest more in their education. In general, these results show that, in addition to educational attainment, high teacher expectations in the 10th grade also have positive impacts on economic outcomes over the life cycle. Conversely, these results underscore concerns about low expectations, which can harm students for years to come.

3.3 Mechanisms

Having shown that higher teacher expectations raise educational attainment — and may have additional impacts on later outcomes — we now turn to a discussion of mechanisms that could explain how. One possibility is that high expectations have no direct impact on student behavior or learning, but function solely through changes in how teachers perceive students. This could affect a student's chances of successfully completing college if, for

results in Table 5 are quite robust to adding school FE to a model that controls for these covariates.

[25]Results are presented in Table S7 in Appendix B.

example, teachers write stellar recommendations or otherwise ease students' pathway to college. Alternatively, teachers with high expectations might modify how they interact with a student or how they allocate their time and effort, which could affect student learning more concretely. Yet another possibility is that teachers' expectations shift students' own expectations about their ultimate educational attainment, which can translate to shifts in their own behavior.

While it is difficult to pinpoint the precise mechanisms since we do not observe teacher-student interactions directly, the ELS provides some information on 12th grade outcomes, which can help to shed some light on why teacher expectations matter. In particular, we examine 12th grade GPA, 12th grade time spent on homework and 12th grade student expectations. We use the same basic research design as in our main results. One difference is that, for each outcome, we present two specifications. The first does not control for the lag (10th-grade value) of the outcome variable, while the second one does. We prefer the lag-score specifications reported in even-numbered columns, as these estimates are more robust and capture the *growth* in the intermediate outcome attributable to teacher expectations. Results are reported in Table 7.

We first examine 12th grade GPA. Columns (1) and (2) provide evidence that teacher expectations lead to a higher GPA. For math teachers, the coefficient is 0.11 and for ELA teachers is 0.16, where mean GPA is 3.04. This change could reflect better student performance due to changes in teacher or student effort decisions. It could also reflect easier grading (or easier classes), which could facilitate a student's path to college if higher a GPA increases the set of colleges to which a student is accepted.

Thus, we ask if there is more direct evidence of changes in student behavior. While teacher effort and time allocations are not observed, we do observe student time investments. In particular, we examine how many hours students spend on homework. We find that higher teacher expectations in the 10th grade lead to increases in time spent doing homework in the 12th grade of roughly 1/3 to 1/2 of an hour. Scaling these coefficients to reflect a more reasonable 10 to 20% change in expectations suggest that a 20% change in the math teacher's expectations would lead to a rise of about 7 minutes per week spent on homework. While modest, this result provides evidence of changes in student behavior, which could explain higher grades and which is inconsistent with the idea that GPA merely reflects easier grading or easier classes.

To examine these shifts a bit further, we conclude by asking if high teacher expectations affect students' own expectations about their future. This would suggest that teacher expectations matter in part through their impact on how students view their educational pathways and futures. We find strong evidence that high 10th grade teacher expectations shift students' expectations upward. For example, adjusting for 10th grade expectations, high 10th grade teacher expectations lead to a rise of 8-10 percentage points in the probability that a student believes he or she will attain a college degree.

It is difficult to identify how the factors examined in Table 7 interact, in part because they are likely to be jointly determined and mutually reinforcing. For example, if a teacher allocates more time to a student due to high expectations, a possible response is that the student puts forth more effort and thus earns a higher GPA, leading to higher expectations. Alternatively, a teacher with high expectations could grade more easily, which might lead a student to have higher expectations and to thus put forth more effort. Still, the results in this section suggest that high expectations do lead to observable changes in student behaviors, performance in school, and to a broader shift in students' own expectations about their future. Together, these mechanisms shed light on how teacher expectations can become self-fulfilling prophecies and buttress a causal interpretation of the main results.

3.4 Identification and Robustness of Estimates

Identification of the causal impact of teacher expectations on educational attainment requires that teacher disagreements be conditionally random and thus generate within-student exogenous variation in teacher expectations. If so, OLS estimates of models that condition on two teachers' expectations can be given a causal interpretation. Intuitively, one teacher's expectation "controls" for omitted factors that might jointly predict the student's educational attainment and the other teacher's expectation. First, in Section 3.4.1, we explicitly test the key threat to identification: that teacher disagreements arise because one teacher sees non-excludable information about a student that is relevant to the student's educational attainment, but is not seen by the other teacher. Second, in Section 3.4.2, we use instrumental variables (IVs) to estimate equation (2) by 2SLS. While each set of IVs we use provides a different source of arguably exogenous variation in teacher expectations, each set has certain drawbacks, which we discuss below. Still, each set answers different possible critiques to our main research design. Thus, we do not present IV results as our main specifications, but as robustness tests that generate results similar to our main OLS estimates.

3.4.1 Falsification Test

In this section, we consider a particular threat to identification: that differences in teacher expectations are due to factors that are not observed by both teachers, but that do matter for college going. For example, consider a student who is exceptionally strong in math, but mediocre in English. A math teacher may recognize this skill when the English teacher does not. This would lead to variation in teacher expectations that is based upon differences in teacher observations of skills that might matter for college. However, the data suggest that this is not true: Figure 3 shows that the expectation gradients with respect to test scores for both teachers (ELA and math) are nearly identical for both ELA and math tests, even though these tests were not administered by teachers and the teachers did not see the students' scores. If teacher disagreements were explained by subject-specific skills differences, we would expect math teachers to respond to reading test scores less strongly than would ELA teachers, and vice versa.

We formally test whether differences in students' subject-specific skills predict teacher disagreements by estimating linear probability models of the form

$$1\{T_{Ei} \neq T_{Mi}\} = \delta_1|S_{Ei} - S_{Mi}| + \delta_2 G_i + X_i\delta_3 + e_i, \tag{3}$$

where S_j are subject-j test scores, $1\{\cdot\}$ is the indicator function, G is 9th-grade GPA, and X is the vector of socio-demographic controls and school fixed effects from equation (2). Estimates of δ_1 and δ_2 are reported in the top rows of Table 8. Row 1, which restricts δ_1 to equal zero, shows that disagreements are decreasing in 9th-grade GPA. This is intuitive, since there is more ambiguity regarding the future outcomes of moderate and low-performing ninth graders. This is also reflected in our main results in Table 5, where controlling for 9th grade GPA affects the size of coefficients even after we control for both teachers' expectations, which leads us to include G in our subsequent analysis. Thus, it is not a threat to identification. However, rows 2 and 3 of Table 8 show that subject-specific skill differences, whether included in levels or a quadratic, do not significantly predict teacher disagreements. This is consistent with the nearly overlapping plots in Figure 3 and reinforces the idea that teacher disagreements are not driven by actual differences in students' subject-specific aptitudes, which might directly enter the education production function.

Another possibility is that variation in expectations is due to large shocks that might eventually affect college completion, but that only one teacher observes. For example, one teacher may learn that a student has a learning disability and revise her expectations accordingly. If this information is not known by the other teacher, then it is not controlled for by including the other teacher's expectation, which means it is an omitted variable correlated with expectations. Of course, if both teachers are aware of the learning disability, then that information is captured by controlling for a second teacher's expectation.

To assess whether relevant information known to only one teacher drives differences in teacher expectations, we estimate variants of equation (3) that replace $|S_{Ei} - S_{Mi}|$ with student-specific information about problems, skills, and inputs that might (i) affect college completion and (ii) only be known by one teacher. These factors include: whether the student is being bullied, has been in a fight, participated in the science fair, finds classes interesting, participated in a "test prep" course for college applications, and whether the parent thinks the student might have an un-diagnosed learning disability.[26]

In general, we show that there are few disagreements. An important exception is the set of variables measuring teachers' perceptions about subject-specific student characteristics, such as attentiveness, passiveness or whether the student likes the subject. This would violate our identifying assumption if these variables directly affected educational attainment. We test this by repeating the analysis in Table 5 including some of these variables, and find that OLS coefficients on teacher expectations do not change. This is perhaps because these types of factors, conditional on student performance, may lead to random variation in teacher expectations without being inputs into the education production function. This suggests that these factors could be used as instruments for teacher expectations, a possibility we explore next.

3.4.2 Instrumenting for Teacher Expectation Disagreements

In our second robustness test, we estimate equation (2) by 2SLS using two distinct sets of instrumental variables. The first stage is thus a modification of equation (1), the teacher expectations production function augmented to include a set of variables Z, which are excluded from the education production function. Z includes variables that could lead to disagreements, but should not affect student outcomes once we have controlled for a sufficient number of student and teacher characteristics.

The first set of instruments leverages the fact that many teachers in our sample are observed multiple times. Thus, for each student, we can use as instruments the average of his or her teachers' expectations for other students (Kling, 2006). The intuition here is that conditional on past achievement, students are as good as randomly assigned to teachers with different propensities for having high expectations (Kane and Staiger, 2008; Chetty et al., 2014a).[27] The second set of instruments uses student-teacher specific data on transi-

[26] Summary statistics for these variables are found in Table S8 in Appendix B.

[27] Kling (2006) relies on random assignment to judges to identify the impact of sentencing on post-incarceration labor market outcomes. In our case, students are not randomly assigned to teachers. However, we rely on a robust result from the teacher value-added literature: conditional on lagged achievement, student-teacher matches are as good as random (Chetty et al., 2014a; Kane and Staiger, 2008). Accordingly, all models explicitly condition on the student's cumulative grade point average (GPA) in the previous year. Moreover, Gershenson et al. (2016) provide evidence using the same data set that systematic racial

tory factors, which are arguably excluded from the education production function, such as teacher disagreements about whether the same student is "passive" in class. Notice, each set of instruments relies on a different source of identifying variation and thus necessitates a different exclusion restriction. The first set are robust to the main critique of the second set of instruments: that disagreements about student demeanor are due to factors that enter the education production function, but are only seen by one teacher. The second set of instruments are robust to the main critique of the first set of instruments: that teachers with high baseline expectations are more effective teachers who affect student outcomes via other practices.

Columns (1) and (2) of Table S9 in Appendix B report baseline estimates of equation (1) using only the Kling-style instruments and X. After conditioning on X, teachers' average expectations for other students, which are arguably excluded from student i's education production function, are statistically significant predictors of the expectations facing student i. Columns (3) and (4) of Appendix Table S9 report estimates of equation (1) using the second set of instruments. These perception variables tend to be individually significant and intuitively signed. For example, column (3) shows that being perceived as passive in English class significantly reduces the likelihood that the English teacher expects a college degree, but has no effect on the math teacher's expectation. The reverse is true for being perceived as passive in math class (column (4)). Moreover, these perception variables are strongly jointly significant. These results are fascinating in their own right, as they imply that teachers are not responding to the student's steady-state (underlying) demeanor, but rather that teachers are forming expectations based on within-semester, within-student, between-class variation in students' passiveness. Similar differences are observed in teachers' perceptions of students' "attentiveness." Most remarkable are English teachers' *negative* responses to whether students "find math fun." The Z_j are arguably excluded from equation (2), because they should not directly affect college completion.

Finally, columns (5) and (6) of Appendix Table S9 report estimates of the first-stage regressions that include both sets of candidate instruments. Once again, the instruments tend to be intuitively signed and jointly and individually statistically significant.[28] These results suggest that (2) can be estimated by 2SLS using different sets of instruments that rely on different sources of identifying variation. We can then formally test for differences using standard over-identification tests.

2SLS estimates and results from over-identification tests are reported in Table S10 in Appendix B, which is organized in the same fashion as the first-stage results in Appendix

differences in teacher expectations are not due to differential sorting to teachers.

[28]The results are qualitatively similar when the school FE are replaced by teacher dyad FE.

Table S9: columns (1) and (2) use the Kling (2006)-type measure of teachers' expectations for other students in the sample to instrument for their expectation for student i, columns (3) and (4) use teachers' perceptions of students' attitudes and dispositions as instruments, and columns (5) and (6) use both sets of instruments simultaneously. Panels A and B of Appendix Table S10 estimate models that condition on school and teacher dyad FE, respectively. We also report OLS estimates for the 2SLS analytic samples, which are smaller than the baseline samples, due to missing values of some instruments.

Three patterns emerge from this analysis, which are particularly striking. First, except for one case, control-function Hausman tests fail to reject that the OLS and 2SLS estimates are equivalent. This suggests that the main OLS estimates presented in Table 5, which are identified off of teacher disagreements, can be given a causal interpretation. Moreover, the 2SLS estimates are quite similar to those for the baseline sample reported in Table 5, which suggests that the 2SLS results are not driven by selection into the 2SLS analytic samples (i.e., by non-randomly missing data). Second, the estimates in panels A and B of Appendix Table S10 are quite similar to each other: the baseline point estimates in panel A are not significantly different from their analogs in panel B. This is consistent with patterns observed in the OLS estimates reported in Table 5 and suggests that the baseline school-FE estimates are not biased by students sorting to teachers who have high expectations. Moreover, this similarity suggests that the school-FE estimates are not biased by, say, teachers who have high baseline expectations also being more effective teachers. We prefer the school-FE estimates in panel A as the baseline estimates due to their increased efficiency: the teacher-dyad FE estimates' standard errors are about twice as large as those for the baseline school-FE model. Finally, over-identification tests fail to reject that the 2SLS estimates that rely on different sets of IVs are equivalent. Indeed, looking across columns within rows of Appendix Table S10, we see that the 2SLS estimates tend to be of similar sign, size, and significance, irrespective of the instruments used in 2SLS estimation.

Never-the-less, both sets of instruments are potentially problematic. The first set of IVs is invalid if teacher perceptions capture unobserved student traits that enter the education production function, as opposed to transitory, between-classroom variation in behavior. Meanwhile, the (Kling, 2006)-style instruments are problematic if their optimism relates to student-varying teacher-level factors (such as student-specific effectiveness) which is not captured by teacher fixed effects that (by construction) remain constant across students. Thus, we view the 2SLS estimates as a robustness test. They provide evidence that the main results are robust to using different sources of variation to generate teacher disagreements in order to identify the impact of teacher expectations on student outcomes. The similarity between the OLS and variously-specified 2SLS estimates suggests that these various approaches are triangulating a real, causal effect of teachers' expectations on long-run student outcomes.

4 A Joint Model of Expectations and Outcomes

In this section, we develop and estimate a joint model of teacher expectations and student outcomes. The model formalizes the idea that teacher disagreements can be used to identify causal estimates of the impact of teacher expectations on student outcomes. The model posits an unobserved latent factor θ_i that uniquely determines the objective probability, absent teacher expectations, that students complete a college degree. Teacher expectations are treated as measurements of this latent factor. Teacher bias is treated like forecast error, defined as the difference between expectations and what a student would achieve absent bias.

The model serves two key purposes. First, it provides a different approach to estimate the impact of teacher expectations on student outcomes, one that explicitly incorporates the idea that the same set of factors — summarized by θ_i and some of which are unobserved by the econometrician — jointly determine teacher expectations and student college degree completion. Second, by recovering θ_i, we can compute teacher bias or forecast error, defined as the difference between teacher expectations and θ_i. We can thus use the estimated model to examine the distribution of biases for different teacher-student pairs. In particular, we examine bias for different teacher and student race pairs. For example, white teachers have lower average expectations than do black teachers for the same black student. Recovering bias allows us to assess whether in such cases black teachers are too optimistic or white teachers are too pessimistic (or both).

4.1 Theoretical Model

Let y_i be the outcome variable of interest, T_{ji}, $j \in \{E, M\}$, be the variables measuring teacher j's expectations about student i's outcome. Let the true model of educational attainment be

$$y_i = c + \theta_i + b_{Ei}\gamma_E + b_{Mi}\gamma_M + \epsilon_{Yi}, \tag{4}$$

where $b_{ji} = b_{ji}(T_{ji}, \theta_i)$ represents teacher j's bias for student i and is a function of teacher j's expectation and the latent factor, and ϵ_{Yi} is a mean-zero educational achievement shock. The parameters of interest are the coefficients γ_j, that map these biases to outcomes. Similar to Cunha et al. (2010), we assign an economic interpretation to θ. This is not a student fixed effect, nor should it be interpreted as a measure of student ability or skill. Rather, it is a latent variable that captures heterogeneity in the objective probability that a student observed in the 10th grade will eventually graduate college.[29] That is, $c + \theta_i$ gives the expected probability that a student i will graduate from college in the absence of teacher biases ($b_{Ei} = b_{Mi} = 0$). The same latent variable will be used in the production function of teacher expectations to capture how teachers observe many of the factors that determine this objective probability. Including biases in teachers' expectations in the education production function is an innovation of the current study that formally allows for self-fulfilling prophecies.

We initially assume that teacher expectation production functions are defined as follows:

$$T_{Ei} = c_E + \phi_E \theta_i + \epsilon_{Ei} \tag{5}$$

$$T_{Mi} = c_M + \phi_M \theta_i + \epsilon_{Mi}, \tag{6}$$

Using the production function (equation (4)) along with the teacher expectations equations, we define bias as the difference between teacher expectations and the objective college completion probability, which we define as expected y_i. Here, the expectation is conditional on teachers assuming no impact of their bias (or that their bias is equal to zero), which we can relax in a way we discuss below. Formally, bias is defined as

$$\begin{aligned} b_{ji} &= T_{ji} - E[Y_i|\theta_i, b_{Ei} = 0, b_{Mi} = 0] \\ &= T_{ji} - c - \theta_i \\ &= (c_j - c) + (\phi_j - 1)\theta_i + \epsilon_{ji} \end{aligned} \tag{7}$$

To generate bias, teacher expectations can thus deviate from the objective college completion probability in three ways. First, the mean of expectations could be systematically different, captured by the difference between c_j and c. Second, teachers may have different beliefs about the role of the latent factor, captured by ϕ_j. Notice, if $\phi_j > 0$, the magnitude of bias rises with θ_i. If $\phi_j < 0$, it falls for students with a higher objective likelihood of college completion. Third, there is random forecast error, which we assume is independent of the disturbance term in education production function.[30]

We highlight two features of the baseline, linear model. First, the model as written implicitly assumes that the impact of expectations is the same as the impact of bias. To see this, substitute in the definition of bias in equation (7) into equation (4) and rearrange

[29] Our interpretation of θ_i as a factor that maps directly into the singular probability that an individual completes a four-year college degree means that it is sensible to be modeled as a singleton. If, as in Cunha et al. (2010), θ_i represented the skill(s) that facilitate college completion, it would make more sense to treat it as a multidimensional vector.

[30] The simple linear model expressed above is identified using standard arguments from the measurement error literature (see Kotlarski (1967)).

terms. The education production function can be rewritten as:

$$y_i = (1 - \gamma_E - \gamma_M)c + (1 - \gamma_E - \gamma_M)\theta_i + T_{E_i}\gamma_E + T_{M_i}\gamma_M + \epsilon_{Y_i}. \tag{8}$$

That is, while replacing teacher biases with teacher expectations would change the interpretation of some of the parameters of the model, the impact of teacher expectations and and of teacher bias are both governed by γ.

Second, the way in which we define bias assumes that teachers, when forming expectations, do not know that their own expectations can directly affect the education production. We can reformulate the model in order to relax this assumption. Specifically, teachers know that the impact of their own expectations on outcomes is equal to γ_j and form their expectations accordingly. We continue to assume that teachers view their own expectations as being unbiased. Teacher expectations become a recursive function of the student's θ and of teachers' expectations, and the reported expectation is a fixed point of that recursive formulation. To simplify exposition, we assume $c = c_E = c_M = 0$. Formally,

$$\begin{aligned} Y_i &= \theta_i + b_{Ei}\gamma_E + b_{Mi}\gamma_M + e_i^Y \\ T_{Ei} &= \phi_E\theta_i + T_{Ei}\gamma_E + \epsilon_{Ei} \\ &= \alpha_E\phi_E\theta_i + \alpha_E\epsilon_{Ei} \\ T_{Mi} &= \phi_M\theta_i + T_{Mi}\gamma_M + \epsilon_{Mi} \\ &= \alpha_M\phi_M\theta_i + \alpha_M\epsilon_{Mi} \end{aligned} \tag{9}$$

where $\alpha_j = \frac{1}{1-\gamma_j}$. We obtain the third and fourth expressions by solving for the T_{ji}. Notice, the estimated γ parameters in the education production function have the same interpretation as in the baseline model. However, the γ also enter the teacher expectation functions since teachers take into account how their expectations affect outcomes. Notice, given the definition of α, teacher expectations become arbitrarily large (i.e., there is no fixed point) if γ_j approaches 1. This point is non-trivial as it implies that as long as the impact of teacher expectations is not too large, even if teachers are aware of the impact of their expectations, expectations (and bias) do not spiral towards infinity, i.e., teachers cannot generate arbitrarily high outcomes solely through high expectations.[31]

[31] There are a number of other possible related extensions. For example, teachers can be aware of the impact of their own expectations and the impact of the other teacher's expectations and form their expectations accordingly. We can also assume that teachers view whether other teachers are biased. In the model outlined above using the recursive formulation, bias as defined in equation (7) is $b_{ji} = \frac{\phi_j + \gamma_j - 1}{1-\gamma_j} + \frac{\epsilon_{ji}}{1-\gamma_j}$. In all cases, parameters attached to teacher expectations remain the same, as does the size restriction on the γ needed for there to be a fixed point, though the α's become more complicated formulas.

4.2 Empirical Implementation

Instead of estimating the linear model directly, we impose a probit functional form on the outcome and teacher expectations to address the binary nature of expectations and outcome variables. We also return to the case in which teachers are not aware of the impact of bias on student outcomes.[32] As before, the outcome is college completion, a binary variable denoted y_i, which takes the value 1 if student i graduates from a 4-year college and 0 otherwise. The probability that $y_i = 1$ is given by:

$$Pr(y_i = 1) = \Phi(c + \theta_i + G_i\beta + b_{Ei}\gamma_E + b_{Mi}\gamma_M), \tag{10}$$

where Φ is the standard normal cdf, and 9th grade GPA (G) is added as an additional control following our findings in Section 3. According to equation (10), college completion is a function of a constant c and a latent factor θ_i, where we assume that

$$\theta_i \sim N(0, \sigma_\theta^2). \tag{11}$$

Together, c, θ_i, and 9th-grade GPA (G) determine the objective probability that student i, absent teacher bias (b), will attain a four-year college degree, where G is included because of evidence that lower 9th-grade GPA predicts a higher likelihood of teacher disagreements.

The model also allows for differences by student race in the education production function parameters along with differences by student and teacher race in the production of teacher expectations. This allows racial mismatch between teachers and students to affect whether and to what degree teachers are biased. This feature of the model is motivated by earlier research using the same data set showing that black and white teachers, when evaluating the same black student, have different expectations about the student's educational attainment (Gershenson et al., 2016). This means that blacks and whites may exhibit different distributions of completing a college degree, which would be captured by race-specific differences in c (mean) and σ_θ. Again, given our interpretation of θ, these differences are not purely ability differences, but also reflect variation in the inputs received by students that could affect long-run educational outcomes, such as early childhood investments and school quality. Teacher biases, expressions for which are derived below, are given by b_{ji}, where $j \in \{E, M\}$ indexes the teacher and the γ parameters map biases to outcomes.

[32] That is because the fixed point algorithm described above does not lead to analytic expressions in the probit formulation, which would complicate estimation. The binary nature of the outcome and expectations variables seems to be of first-order importance, which we thus address in our main results. Reassuringly, treating the model as linear, in which case the γ do not change if teachers know that biases can affect outcomes, yields similar results to those assuming a probit formulation.

We jointly estimate teacher-expectation and student-outcome equations as functions of θ_i and G_i. Teacher expectations, denoted T_{ji} for teachers $j \in \{E, M\}$, are given by:

$$Pr(T_{ji} = 1) = \Phi(c_j + \phi_j \theta_i + G_i \beta_j + D_{ji} \times [c_{j,D} + \phi_{j,D} \theta_i + G_i \beta_{j,D}]). \tag{12}$$

The indicator D_{ji} takes the value of one if student i faces an other-race subject-j teacher, and zero otherwise. This captures how teacher-student racial mismatch can change how teachers form expectations for a given student with a singular objective probability of college completion. In other words, racial mismatch between teachers and students can affect whether and to what degree teachers are biased. We define bias by combining equations (10) and (12):

$$b_{ji} \equiv T_{ji} - \Phi(c + \theta_i + G_i \beta) \tag{13}$$

so that bias is simply the difference between what a teacher reports (T_{ij}) and the objective probability that the student would complete a college degree given θ_i and G_i. This definition of bias implies that b_{ji} is continuous, increases 1:1 with T_{ji}, and is $\in (-1, 1)$.[33] According to equation (13), teacher bias arises when a teacher's expectations diverge from information that is common to both of them, including ninth-grade GPA and the latent factor θ_i. The model captures several potential sources of bias in teacher expectations. Based on the patterns observed in section 2.1, we allow teachers to be wrong on average, meaning c_j and $(c_j + c_{j,D})$ can deviate from c. Teachers may also be wrong about how θ_i maps into outcomes, which occurs if $\phi_j \neq 1$. For example, the reduced form finding that teachers seem to over-estimate low and high educational attainment outcomes despite our controlling for a host of observables could mean that $\phi_j > 1$.[34] Teachers may also be biased in how they map observable grades G_i to outcomes, in which case $\beta_j \neq \beta$. Finally, teachers may be wrong for idiosyncratic reasons, which is captured by the standard-normal disturbances that lead to the probit functional form in equation (12).

[33] In Appendix C, we estimate several alternative models. In Appendix C.1, we estimate models using alternative definitions of bias along with a model where, rather than include bias in the outcome equation, we allow the outcome to be a direct function of teacher expectations. In Appendix C.2, we estimate a model that permits correlation in bias for two teachers evaluating the same student. In Appendix C.3, we estimate a model that relies on parameter restrictions for identification and does not require distributional assumptions or additional data (beyond teacher expectations and student outcomes) and yields estimates of γ as analytic expressions that are functions of moments from the data. In Appendix C.4, we estimate a model similar to the one in Appendix C.3, but with additional data.

[34] The term ϕ_j may capture how teachers have biased beliefs about how a given θ_i affects outcomes. It may also capture that teachers correctly map ability to outcomes, but mis-estimate θ_i. We cannot separately identify these effects. Similarly, the term β_j may represent that teachers are biased in the mapping or in their observation of G_i. Again, we are unable to separately identify these mechanisms. For ease of interpretation, we will assume that teachers observe G_i and θ_i, but incorrectly map these to outcomes when forming expectations.

4.3 Identification and Estimation of the Joint Model

There are two points to discuss regarding identification of the econometric model defined by equations (10)-(13). The first is whether the estimated γ are accurately interpreted as causal. Analogous to what is required for identification of the reduced form models estimated in Section 3.4, the argument is that teacher bias is exogenous in the production function of student outcomes. The second is that we need sufficient data to estimate the distribution of the latent factor, which is not a trivial condition. We discuss each in point in turn.

For the γ_j to be given causal interpretations in equation (10), the biases (denoted b) must be exogenous, conditional on θ, c, and G. The b, in turn, mechanically explain why teacher expectations might diverge for a given student. Intuitively, this means that information that teachers use to form expectations, but which is not used by both teachers, does not directly affect college-going. Information about student i that is commonly used by both teachers to form expectations is captured in θ_i, c, and ninth-grade GPA. In other words, identification relies on a similar argument to our main results: that disagreements are conditionally random. This was argued in Section 3.4.

As written, the model described in equations (10)-(13) in Section 4.2 is not econometrically identified in that there are not enough measurements to identify all model parameters. There are two reasons. First, we cannot identify latent factors with discrete outcomes absent further data. Second, the two expectation equations, which are used as imperfect measurements of student abililty, are also included as regressors (via the b) in the outcome equation. One way to achieve identification is to place additional restrictions on parameters as in Heckman et al. (2006). In Appendix C.3, we show that if we restrict $\gamma_E = \gamma_M \equiv \gamma$ and $\phi_E = \phi_M \equiv \phi$, and and replace the probit functional forms with linear models with years of schooling as the outcome variable, we obtain an identified system of equations .

Parameter restrictions are a useful alternative when there are not obvious exclusion restrictions on additional data, i.e., variables that only enter either the expectations or the outcome equations, but not both. Typically, it is difficult to defend such exclusions. Fortunately, two exams (a math and a reading test) were administered to all ELS-2002 students. Results from these exams were not revealed to students or teachers. Therefore, the exams can be used as additional (mis)measurements of student ability, but do not enter into the student outcome equation once we have conditioned on θ_i. In other words, scores on these exams should only be associated with educational attainment because they reflect factors that would likewise affect college completion, but not because teachers observe them.[35]

[35] To assess the validity of using these additional measurements, in one of the alternative models in Appendix C.4, we estimate a version of the linear years-of-schooling model with the parametric assumptions

We also control for 9th grade GPA in the outcome equation, allow grades to affect teacher expectations, and use grades to identify θ_i. This is useful for a couple of reasons. First, we might be concerned that math and reading test scores do not contain the full set of skills that teachers observe, in which case there would be bias in the impact of teacher forecast error on y. Several papers (e.g., Cunha et al., 2012) argue that test scores might not measure non-cognitive skills, such as motivation or grit, but that grades would. Moreover, we do not want to see grades as independent of θ, which requires that we model their relationship with θ. Finally, we want to illustrate how teacher bias can be due to a misreading of the mapping of skills to outcomes, where some skills are observed by the econometrician and some are not.

Formally, we add three measurement equations:

$$S_{ji} = c_{Sj} + \phi_{Sj}\theta_i + e_{Sji}, j \in \{E, M\} \tag{14}$$

$$G_i = c_G + \phi_G\theta_i + e_{Gi} \tag{15}$$

where S_j is the test score in subject j. In the equations, e_{SEi} and e_{SMi} follow normal distributions with $N(0, \sigma_{S,j})$ for $j \in \{E, M\}$, independent across i and j. Further, e_{Gi} are assumed to be independent of e_{SEi} and e_{SMi} and to follow a truncated normal distribution with mean 0 and standard deviation σ_G, where the upper and lower cutoff values are equal to the draw of e_{Gi} that equate GPA to 4.0 and 0, respectively. Appendix D shows formally that the addition of these three measurement equations identifies the system of equations (10)-(15).

The econometric model is described in equations (10)-(15). We collect the parameters to be estimated into a vector denoted Ξ:

$$\Xi = \left\langle c, \sigma_\theta, \beta, \{\gamma_j, c_j, \phi_j, \beta_j, c_{j,D}, \phi_{j,D}, \beta_{j,D}, c_{S,j}, \phi_{S,j}\}_{j\in\{E,M\}}, c_G, \phi_G, \sigma_G \right\rangle. \tag{16}$$

We estimate Ξ using simulated maximum likelihood (Hajivassiliou and Ruud, 1994). In the inner loop of the estimation algorithm, we compute the likelihood for a particular set of candidate parameters, which are indexed by (g) and denoted $\Xi^{(g)}$. To calculate the log likelihood for a given set of candidate parameters $\Xi^{(g)}$, we first draw the latent factor K times for each individual i. We denote each draw $\theta_{ik}^{(g)}$.[36] For each $\theta_{ik}^{(g)}$, we use distributional assumptions

on error terms used in our main model along with the additional data needed for identification of the main model. We show that estimates of γ in that model are nearly identical to estimates of the restricted non-parametric model described above. Similarity of results gives us some confidence in our approach of using these additional measurements to achieve identification of our main model.

[36]Prior to estimating, we draw a block matrix of size $N \times K$ from a standard normal distribution once and denote it Ψ, where N is the number of individuals in the sample and K is the number of simulation

on the error terms, additional candidate parameters, and data to calculate the likelihood contribution for teacher expectations ($P_{T_\tau}(T_{i,j}|\theta_{ik}^{(g)})$, $j \in \{E, M\}$). Next, for each draw, we calculate bias using equation (13). Then, we calculate the likelihood contribution for college completion, denoting the probability $P_y(y_i|\theta_{ik}^{(g)})$. Similarly, we compute the likelihood contributions for the test scores and for ninth-grade GPA, denoting these densities $f_E(S_{E,i}|\theta_{ik}^{(g)})$, $f_M(S_{M,i}|\theta_{ik}^{(g)})$, and $f_G(G_i|\theta_{ik}^{(g)})$, respectively. Using these components, we calculate the value of the likelihood for each draw of the latent factor $\theta_{ik}^{(g)}$ as:

$$\begin{aligned} L_{ik}^{(g)} &= P_y(y_i|\theta_{ik}^{(g)}) \times \textstyle\prod_{\tau \in \{E,M\}} P_{T_\tau}(T_{\tau i}|\theta_{ik}^{(g)}) \\ &\times \; f_E(S_{Ei}|\theta_{ik}^{(g)}) \times f_M(S_{Mi}|\theta_{ik}^{(g)}) \times f_G(G_i|\theta_{ik}^{(g)}). \end{aligned} \tag{17}$$

After constructing $L_{ik}^{(g)}$ for each individual i and draw k, we then average $L_{ik}^{(g)}$ over the K draws for each individual. Finally, we take the log and then sum over all N individuals to obtain the log-likelihood: i.e., we compute:

$$l^{(g)} = \sum_{i=1}^{N} \log \left(\frac{1}{K} \sum_{k=1}^{K} L_{ik}^{(g)} \right). \tag{18}$$

In the outer loop, we repeat the inner loop for different sets of candidate parameters until the log likelihood function is maximized. We use quasi-Newton methods to choose candidate parameters.[37]

4.4 Estimates

Tables 9 and 10 report parameter estimates of the education and teacher expectation production functions defined by equations (10) and (12), respectively.[38] Column (1) of Table 9 reports parameter estimates for white students, and the estimated γ suggest that teacher expectations have positive, statistically significant effects on the probability that white students complete a 4-year degree.[39] The estimated β is positive and statistically significant,

draws, set to 1,000. At each draw, $\theta_{ik}^{(g)}$ denotes the value of the latent factor for individual i and draw k. It is element (i, k) in Ψ multiplied by $\sigma_\theta^{(g)}$. This helps to avoid the so-called "chattering" effect, which can lead to different values of the likelihood function given the same parameters due to differences in random draws at each parameter set.

[37] We also repeat the estimation algorithm for different sets of starting values to help ensure that we have not found a local maximum.

[38] Table S11 in Appendix B reports the "nuisance parameter" estimates from measurement equation (15).

[39] We refer to effects of bias and teacher expectations interchangeably since there is a 1:1 relationship between these constructs, by definition, in equation (12). We also report parameter estimates where we use actual teacher expectations instead of bias in Appendix Table S13. As previously explained, γ estimates are statistically indistinguishable.

indicating that students with higher 9th-grade GPAs are significantly more likely to earn a four-year college degree than their counterparts with lower GPAs. This result is intuitive and provides a useful check of the model, since GPA is a known proxy for academic ability that predicts college completion (Bound and Turner, 2011). The magnitudes of these probit coefficients cannot be directly interpreted, so the bottom panel of Table 9 reports the APE of teachers' expectations, the main independent variables of interest, on the likelihood of earning a four-year degree.[40]

The APEs indicate that for white students, on average, the impact of either teacher changing from not expecting to expecting a college degree is about a 20 percentage point increase in the likelihood of the student completing a four-year degree. These estimates are remarkably similar to the average partial effects shown in Table 5 of 0.13 and 0.14 for math teachers and ELA teachers, respectively. The similarity between these two approaches lends additional credence to the interpretation of these estimates as causal effects of teacher expectations on students' long-run educational attainment. These effects, moreover, translate into statistically significant elasticities of college completion with respect to biases of about 0.12 to 0.13. Column (2) of Table 9 reports parameter estimates for black students, and the estimated γ once again suggest that teacher expectations have positive effects on educational attainment. However, only the ELA teacher's expectation is statistically significant at traditional confidence levels, and this coefficient is larger for ELA versus for math teachers, again consistent with results using OLS regressions found in Table 5.[41] The estimated β is once again positive and statistically significant, though smaller in magnitude than that for white students.

The variance and mean of θ, via the probit function, determine the objective probability (absent teacher bias and along with GPA) that a student will complete college. Consistent with realized educational outcomes, a comparison of columns (1) and (2) shows that the distribution of θ for black students is centered to the left of that for white students, and exhibits greater variance. This means that upon reaching the tenth grade, black students are already disadvantaged relative to their white counterparts in terms of college completion probability. Again, this does not reflect their ability, but instead captures racial disparities in the multitude of investments over the lifecycle, including factors such as school quality, neighborhood effects, and early childhood environments and resources. Our model is designed to separate the objective probability (which teachers use to form their expectations) from the impact of teacher expectations via self-fulfilling prophecies.

[40]Standard errors for the APE are computed via the delta method. The APE are evaluated at the mean value of θ, which is zero by construction.

[41]It is worth noting that for black student the math and ELA γ are not significantly different from one another.

Table 10 reports the parameter estimates of the teacher expectation production functions. The first two columns report the parameter estimates for white students' ELA and math teachers, respectively. The production of teacher expectations for white students is broadly similar across subjects: the other-race teacher indicators are both statistically insignificant, as are their corresponding APE, which is consistent with the lack of a racial-mismatch effect on teachers' expectations for white students. Also, intuitively, teachers' expectations are increasing in both θ and 9th-grade GPA. The results for black students, reported in columns (3) and (4), are broadly similar.

However, there is one notable difference: for black students, there are significant negative effects of student-teacher racial mismatch on teachers' expectations. This is consistent with estimates reported in Gershenson et al. (2016). Specifically, pooled estimates of student-FE LPMs in Gershenson et al. (2016) find that racial mismatch reduces the probability that teachers expect a black student will complete a college degree by 0.09. However, when allowing the effect to vary by subject, the authors find that the racial-mismatch effect is about twice as large for math teachers (0.15) as for ELA teachers (0.07). This pattern, and the effect sizes, are remarkably similar to those reported in columns (3) and (4) of Table 10. That the measurement error model estimated here produces similar evidence regarding the impact of student-teacher racial mismatch on teachers' educational expectations for black students, despite using a demonstrably different econometric approach and estimation procedure, cross-validates the measurement error model and lends additional support to the causal interpretation of the estimated impact of teacher expectations on educational attainment.

4.5 The Distribution of Bias by Race

Thus far, the model confirms previous results suggesting that student-teacher racial mismatch reduces teachers' educational expectations for black students. However, the results in Table 10 do not speak directly to long-debated questions about whether, to what extent, and in what direction teacher expectations are biased. The model developed in section 4.2, and specifically equation (13), provide answers to these questions.

Figure 4 plots kernel density estimates of the distributions of the biases in teachers' expectations separately by student race, subject, and student-teacher race congruence.[42] Panel A shows the distributions of ELA teachers' biases. For both same- and other-race ELA teachers of both white and black students, the average bias is positive. In other words,

[42] Another way to illustrate these differences is using contour plots, which are presented in Figure S1 in Appendix B. These plots (heat maps) depict higher concentrations as brighter colors.

teachers are overly optimistic on average, which is consistent with patterns observed in the raw ELS data documented in Table 1 and in Figure 2. Also, for both same- and other-race ELA teachers the average amount of bias is similar for both white and black students. However, the average positive bias (over-optimism) is slightly larger for black students when evaluated by a black teacher. This is consistent with evidence of smaller effects of student-teacher racial mismatch on ELA teachers' expectations for black students. The similarity in means is somewhat misleading, however, as it obfuscates more pronounced differences across the distribution. Specifically, there is more mass at zero bias for blacks than for whites, as many teachers accurately predict that black students will not complete college, and this is true for both same- and other-race teachers. There is similarly more mass at one (the upper bound of bias) for blacks than whites, which is due to both same- and other-race teachers being more likely to expect black students to complete college, even when the objective probability of them doing so is nil. White students, meanwhile, are more likely than blacks to receive positive bias in the range of about 0.1 to 0.7, which means that both same- and other-race teachers are more likely to give white students the "benefit of the doubt" and expect a four-year degree when their objective probability of completing college is in the 30-90% range.

Panel B of Figure 4 similarly plots the distributions of math teachers' biases. Many of the qualitative patterns observed in Panel A for ELA teachers are present here: biases are positive on average for all students, blacks are more likely than whites to receive zero bias, and on average, black students receive more positive bias (over-optimism) than white students when evaluated by black teachers, while the opposite is true for white teachers' expectations. However, differences in the bias distributions of same- and other-race math teachers are significantly more pronounced than the corresponding differences for ELA teachers. This is to be expected, given the result in Table 10 that the effect of racial mismatch on expectations is significantly larger for math teachers than for ELA teachers. Indeed, these mean differences are driven by a notable increase in the frequency of objectively correct (zero-bias) expectations and a flattening of the right tail of the bias distribution for other-race teachers' expectations for black students. This raises a nuanced, but important point: other-race math teachers' expectations for black students may be more accurate (less biased) than those made by black math teachers. However, this accuracy has the potential to propagate racial gaps in educational attainment, since we have argued that high expectations, even overly optimistic ones, have a positive impact on college completion. The results in Figure 4 indicate that on average, all teachers are too optimistic about students' college-completion potential, but the degree of overoptimism is greater for black students assessed by black teachers relative to white teachers.

4.6 Bias and Racial Attainment Gaps

We have demonstrated racial differences in the production of bias along with the impact of expectations (including biased ones) on outcomes. However, we have yet to investigate how these two mechanisms interact to contribute to the racial gap in college completion. We begin to do so here, by noting that the model distinguishes between three types of racial differences that can influence racial gaps in educational attainment:

1. Initial conditions, including ninth-grade GPA and the latent factor θ_i, which combine to identify the objective likelihood of college completion (net of the impact of bias) at the time tenth-grade teachers form expectations.

2. The mapping between initial conditions and teacher expectations governed by the parameters in equation (12); i.e., racial disparities in the teacher expectations faced by students with the same θ_i and G_i.

3. The production function of student outcomes governed by parameters in equation (10).

Figure 5 illustrates how each of these factors contributes to racial disparities.[43] The figure plots the CDF of the probability that black and white students will obtain a four-year college degree, assuming that all students have white teachers.

In the upper-right panel of Figure 5, we simulate the black-white college completion gap under the counterfactual in which blacks are assigned the same initial conditions as whites, i.e., the same distribution of θ_i and of G_i. Not surprisingly, this closes much of the attainment gap, as many of the differences in the distribution of educational attainment arise from factors occurring prior to the tenth grade. Still, even with the same initial conditions, black students do not face the same distribution of college completion as white students. This means that some of the gap can be explained by how initial conditions map to expectations along with racial differences in how expectations produce outcomes.

One interesting feature of the upper right panel of Figure 5 is that black students with initial conditions suggesting a low probability of college completion might do better than their white counterparts if assigned the same initial conditions. The reason is that some black students with lower initial conditions may face higher positive bias. This can be seen in Figure 4, where black students are more likely to face optimistic teachers. Nonetheless, towards the upper end of the distribution, whites outperform blacks despite having the

[43] For each counterfactual simulation, this is done by drawing e_{Gi}, e_{Ei}, e_{Mi}, and θ 100,000 times using the distributional assumptions outlined in subsection 4.3 given our parameter estimates and simulating GPA, as well as ELA and math teacher expectations using equations (12) and (15). The probability that black and white students will obtain a four-year college degree is then calculated using equation (10).

same ninth-grade GPA and the same objective probability of completing college. Again, since θ_i does not represent innate ability, these results suggest that two students enter the tenth grade having the same objective probabilities (net of bias) of finishing college might experience different outcomes. This discrepancy is due to racial differences in the production and impact of biases, which thus exacerbates existing gaps.

To illustrate this point, the lower-left panel of Figure 5 shows what happens if black and white students not only have the same initial conditions, but also the same mapping from initial conditions to teacher expectations. This has a relatively small additional impact on the gap, which can be seen in the lower right panel, where both counterfactuals are simulated. Notice, for individuals with relatively low or relatively high objective probabilities of college completion, the impact of the production of teacher bias is nearly zero. In fact, some black students in the lower tails are harmed if they face the same production of bias as white students. This is because black students with low θ_i tend to face higher expectations from white teachers. For black students in the middle of the distribution, however, facing the same mapping from initial conditions to teacher expectations as whites is helpful in promoting college completion. This finding is consistent with the distributions of bias plotted in Figure 4, which indicate that white students who begin with objective probabilities of college completion that are neither very high nor very low are more likely to be given the "benefit of the doubt" than are black students. Given that expectations matter, this can raise the attainment gap through self-fulfilling prophecies.

The lower right panel of Figure 5 also illustrates that the remainder of the gap is closed when blacks counterfactually face the same education production function as whites (governed by the parameters in equation (10)). Part of the production function difference is due to differences in γ, particularly differences in math teachers' γs across races. Another difference is in β, which may reflect disparities in school quality.[44] In general, Figure 5 demonstrates that most of the attainment gap between blacks and whites arises from factors that occur prior to our observing them in the tenth grade, which is not surprising and underscores the importance of interventions in early-childhood and primary-school education. Still, initial conditions do not account for the entire gap, which is concerning since it means that teacher expectations widen the gap. This is due to racial differences in the impact of bias on outcomes, but also due to differences in the production of bias.

To explore these patterns a bit further, we plot teachers' expectations for black students as a function of θ_i while making different assumptions about how their expectations are formed. The top panels of Figure 6 show how white ELA and math teachers' expectations

[44] Indeed, if black students face white students' γ, but different β, a small gap remains.

change when we impose that, for a given θ, black students face the same expectation normally given to a white student. It is immediately apparent that not all black students are helped by such change in expectation formation. Indeed, at low levels of θ white teachers have higher expectations for black students than for white students. However, at high levels of θ, black students benefit from this change. This is consistent with the distributions of bias presented in Figure 4.

The bottom panels of Figure 6 show how the expectations of ELA and math teachers for black students change when the expectation is formed by a black — rather than white — teacher. Among ELA teachers there is a muted increase in expectations at all levels of θ of having more black teachers. For math, the effect is much larger due to the larger impact of racial mismatch on math teachers' expectations. Importantly, white students are not hurt by the addition of other-race teachers, as shown in the corresponding Figure S2 in Appendix B for white students. If black students faced more black teachers, they would face higher expectations. Moreover, if white teachers formed expectations for black students the same way they did for white students with the same θ, black students with moderate to higher θ would be exposed to more positive bias. Coupled with our key finding that teacher expectations matter for student outcomes, these results provide some preliminary evidence on policy implications. In particular, policies shifting the expectations production function faced by black students (through changes in white teachers' expectation or the racial composition of teachers) could raise black students' achievement and thus contribute to closing attainment gaps.

5 Conclusion

We provide a framework to estimate the causal impact of teacher expectations on student outcomes. Our approach jointly estimates education and teacher expectation production functions using data from a nationally representative longitudinal survey of U.S. high school students. Our identification strategy leverages teacher disagreements, an idea that we formalize using insights from the measurement-error literature. Our analysis suggests that teacher expectations are not just accurate forecasts of student outcomes, but that biases also influence outcomes by becoming self-fulfilling prophecies. Moreover, we find that the production of white teachers' expectations places black students at a disadvantage. For a given objective probability of college completion, white teachers are less optimistic about black students. Policies that would put black students on the same footing as white students in terms of how teacher expectations are formed could narrow attainment gaps, though only slightly. Not surprisingly, large gaps in the objective probability of college completion exist by the time students reach the tenth grade and these would need to be addressed with earlier interventions.

Our results suggest several areas for future research. One extension would consider variation by gender, family income, or other factors in how teacher expectations diverge from objective probabilities. Another set of extensions would use our framework for different populations or contexts. One possibility would be to examine the role of biased beliefs for students in earlier grades to assess whether they have stronger, or longer-lasting, effects. More generally, our approach could be used to assess the role of expectations in driving behavior in other contexts where individuals are tasked with making economic decisions under uncertainty. However, doing so using our framework would require data on multiple reports of expectations of a given outcome. Thus, the approach we develop here along with our results suggest that collecting multiple reports of subjective expectations could be used to identify causal effects of expectations on outcomes.

References

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Ashenfelter, Orley C and Alan B Krueger, "Estimates of the Economic Returns to Schooling from a New Sample of Twins," *American Economic Review*, 1994, *84* (5), 1157–73.

Bailey, Martha J. and Susan M. Dynarski, "Inequality in Postsecondary Education," in G. Duncan and R. Murnane, eds., *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*, New York, NY: Russell Sage Foundation, 2011, pp. pp. 171–132.

Becker, G.S., *Human Capital: A Theoretical Analysis with Special Reference to Education*, Columbia University Press, 1964.

Boser, Ulrich, Megan Wilhelm, and Robert Hanna, "The Power of the Pygmalion Effect: Teachers Expectations Strongly Predict College Completion," *The Center for American Progress*, 2014, pp. 1–7.

Bound, John and Gary Solon, "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling," *Economics of Education Review*, 1999, *18* (2), 169–182.

— **and Sarah Turner**, "Dropouts and Diplomas: The Divergence in Collegiate Outcomes," in E. Hanushek, S. Machin, and L. Woessmann, eds., *Handbook of the Economics of Education*, Vol. 4, New York, NY: Russell Sage Foundation, 2011, pp. 573–613.

Brophy, Jere E., "Research on the Self-Fulfilling Prophecy and Teacher Expectations," *Journal of Educational Psychology*, 1983, *75* (5), 631.

Burgess, Simon and Ellen Greaves, "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities," *Journal of Labor Economics*, 2013, *31* (3), 535–576.

Cameron, Stephen V. and James J. Heckman, "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 2001, *109* (3), 455–499.

Card, David, "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 1999, *3*, 1801–1863.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *The American Economic Review*, 2014, *104* (9), 2593–2632.

— **, — , and —** , "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104* (9), 2633–79.

Chiang, Chun-Fang and Brian Knight, "Media Bias and Influence: Evidence from Newspaper Endorsements," *The Review of Economic Studies*, 2011, pp. 795–820.

Cunha, Flávio, Irma Elo, and Jennifer Culhane, "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation," NBER Working Paper 2013.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Dee, Thomas S., "Are There Civic Returns to Education?," *Journal of Public Economics*, 2004a, *88* (9), 1697–1720.

— , "Teachers, Race, and Student Achievement in a Randomized Experiment," *Review of Economics and Statistics*, 2004b, *86* (1), 195–210.

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

Delavande, Adeline and Hans-Peter Kohler, "HIV/AIDS-related Expectations and Risky Sexual Behaviour in Malawi," *Review of Economic Studies*, 2016, *83* (1), 118–164.

DellaVigna, S. and E. Kaplan, "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

der Klaauw, Wilbert Van and Kenneth I Wolpin, "Social Security and the Retirement and Savings Behavior of Low-Income Households," *Journal of Econometrics*, 2008, *145* (1), 21–42.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach, "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion," *Journal of Policy Analysis and Management*, 2013, *32* (4), 692–717.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos, "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom," *The American Economic Review*, 2014, *104* (8), 2567–2591.

Ferguson, Ronald F., "Teachers' Perceptions and Expectations and the Black-White Test Score Gap," *Urban Education*, 2003, *38* (4), 460–507.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps, "Leaving Boys Behind: Gender Disparities in High Academic Achievement," *Journal of Human Resources*, 2015, *50* (3), 549–579.

Fryer, Roland G., "Racial Inequality in the 21st Century: The Declining Significance of Discrimination," NBER Working Paper 2010.

Gentzkow, Matthew and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas W. Papageorge, "The Long-Run Impacts of Same-Race Teachers," NBER Working Paper No. 25254, 2018.

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

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," 2015.

Goldberger, Arthur S., "Structural Equation Methods in the Social Sciences," *Econometrica*, 1972, *40* (6), 979–1001.

Gregory, Anne and Francis Huang, "It Takes a Village: The Effects of 10th Grade College-Going Expectations of Students, Parents, and Teachers Four Years Later," *American Journal of Community Psychology*, 2013, *52* (1-2), 41–55.

Grossman, Michael, "Education and Nonmarket Outcomes," *Handbook of the Economics of Education*, 2006, *1*, 577–633.

Hajivassiliou, Vassilis A. and Paul A. Ruud, "Classical Estimation Methods for LDV Models Using Simulation," *Handbook of Econometrics*, 1994, *4*, 2383–2441.

Hanushek, Eric A. and Steven G. Rivkin, "Generalizations about Using Value-Added Measures of Teacher Quality," *American Economic Review*, 2010, *100* (2), 267–271.

_, John F. Kain, and Steven G. Rivkin, "Why Public Schools Lose Teachers," *Journal of Human Resources*, 2004, *39* (2), 326–354.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411.

Hu, Yingyao and Susanne M. Schennach, "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 2008, *76* (1), 195–216.

Hudomiet, Peter, Gábor Kézdi, and Robert J Willis, "Stock Market Crash and Expectations of American Households," *Journal of Applied Econometrics*, 2011, *26* (3), 393–415.

Hurd, Michael D, "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, 2009, *1*, 543.

Jackson, C. Kirabo, "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation," *Journal of Labor Economics*, 2009, *27* (2), 213–256.

Jacob, Brian A., "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education," *Economics of Education Review*, 2002, *21* (6), 589–598.

Jacob, Brian A and Tamara Wilder, "Educational Expectations and Attainment," NBER Working Paper 2010.

Jones, Daniel B and Andres Hill, "Self-Fulfilling Prophecies in the Classroom," Mimeo, University of South Carolina 2018.

Jöreskog, Karl G. and Arthur S. Goldberger, "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 1975, *70* (351a), 631–639.

Jussim, L. and J.S. Eccles, "Teacher Expectations: II. Construction and Reflection of Student Achievement.," *Journal of Personality and Social Psychology*, 1992, *63* (6), 947–961.

Jussim, Lee and Kent D. Harber, "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies," *Personality and Social Psychology Review*, 2005, *9* (2), 131–155.

Kane, Thomas J and Douglas O Staiger, "Estimating teacher impacts on student achievement: An experimental evaluation," Technical Report, NBER Working Paper No. 14607 2008.

Kling, Jeffrey R, "Incarceration Length, Employment, and Earnings," *American Economic Review*, 2006, *96* (3), 863–876.

Kotlarski, Ignacy, "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 1967, *20* (1), 69–76.

Lareau, Annette, *Unequal Childhoods: Class, Race, and Family Life*, Univ of California Press, 2011.

_ **and Elliot B. Weininger**, "Class and the Transition to Adulthood," in G. Lareau and E. B. Weininger, eds., *Social Class: How Does It Work*, New York, NY: Russell Sage Foundation, 2008, pp. 118–151.

Lavy, Victor and Edith Sand, "On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases," NBER Working Paper No. 20909 2015.

Lochner, Lance and Enrico Moretti, "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 2004, *94* (1), 155–189.

Loury, Glenn C., *The Anatomy of Racial Inequality*, Harvard University Press, 2009.

Lundberg, Shelly, Robert A. Pollak, and Jenna Stearns, "Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing," *The Journal of Economic Perspectives*, 2016, *30* (2), 79–101.

Machin, Stephen, Olivier Marie, and Sunčica Vujić, "The Crime Reducing Effect of Education," *Economic Journal*, 2011, *121* (552), 463–484.

Manski, C.F., "Measuring Expectations," *Econometrica*, 2004, *72* (5), 1329–1376.

Manski, Charles F, "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?," in "Studies of Supply and Demand in Higher Education," University of Chicago Press, 1993, pp. 43–60.

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

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos, "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 2004, *88* (9), 1667–1695.

Rist, Ray, "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education," *Harvard Educational Review*, 1970, *40* (3), 411–451.

Rosenthal, Robert and Lenore Jacobson, "Pygmalion in the Classroom," *The Urban Review*, 1968, *3* (1), 16–20.

Savage, L.J., *The Foundations of Statistics*, Wiley, 1954.

Staiger, Douglas O. and Jonah E. Rockoff, "Searching for effective teachers with imperfect information," *Journal of Economic Perspectives*, 2010, *24* (3), 97–118.

Terrier, Camille, "Giving a Little Help to Girls? Evidence on Grade Discrimination and its Effect on Students' Achievement," Centre for Economic Performance Discussion Paper No. 1341 2015.

Wiswall, Matthew and Basit Zafar, "Determinants of College Major Choice: Identification Using an Information Experiment," *Review of Economic Studies*, 2015, *82* (2), 791–824.

Figures and Tables

Table 1: Analytic Sample Means — Students

Sample (Students) :	All (1)	White (2)	Black (3)	Male (4)	Female (5)
Educational Attainment					
Completed College or more	0.45	0.49	0.29	0.43	0.47
Completed < HS Diploma	0.01	0.01	0.02	0.01	0.01
Education Completed, Years	14.67	14.83	14.08	14.51	14.81
	(2.06)	(2.06)	(1.84)	(2.05)	(2.07)
Teacher Expectations					
College or More, English	0.64	0.67	0.48	0.60	0.67
Expect < HS, English	0.01	0.01	0.03	0.02	0.01
ELA Teacher Expected Years	15.65	15.78	14.86	15.48	15.80
	(2.23)	(2.14)	(2.21)	(2.29)	(2.16)
College or More, Math	0.63	0.66	0.44	0.61	0.65
Expect < HS, Math	0.01	0.01	0.03	0.01	0.01
Math Teacher Expected Years	15.51	15.65	14.66	15.43	15.59
	(2.09)	(1.99)	(2.07)	(2.16)	(2.03)
Teacher Expectations Disagree	0.21	0.20	0.25	0.21	0.21
Math Teacher has Higher Expectation	0.10	0.10	0.11	0.10	0.10
Academic Background					
Reading Assessment	52.82	54.67	46.71	52.39	53.21
	(9.83)	(9.26)	(8.99)	(10.20)	(9.47)
Math Assessment	53.01	54.71	45.77	54.00	52.12
	(9.67)	(8.78)	(8.88)	(10.13)	(9.15)
9th grade GPA	2.92	3.02	2.44	2.82	3.01
	(0.78)	(0.73)	(0.76)	(0.78)	(0.77)
Demographics and Socioeconomic Status					
Household Income < 20K	0.11	0.06	0.26	0.09	0.13
Household Income > 100K	0.18	0.21	0.08	0.19	0.17
Mother has ≤ HS diploma	0.34	0.29	0.39	0.32	0.35
Mother has a Bachelor's or More	0.31	0.34	0.23	0.33	0.29
Teacher					
ELA Teacher Non-White	0.10	0.05	0.26	0.10	0.10
Math Teacher Non-White	0.11	0.06	0.21	0.11	0.11
ELA Teacher Black	0.04	0.02	0.20	0.04	0.04
Math Teacher Black	0.04	0.02	0.16	0.03	0.04
Observations	6060	3970	610	2870	3190

Notes: This table presents means of variables where students are the unit of analysis. Standard deviations for non-binary variables are reported in parentheses. HS denotes high school. 9th-grade GPAs are on a 4.0 scale. Math and reading assessment scores are on a 0-100 scale. All sample sizes are rounded to the nearest 10 in accordance with NCES regulations for restricted data.

Table 2: Analytic Sample Means — Teachers

Sample (Teachers) :	All Teachers (1)	Math Teachers (2)	English Teachers (3)	White Teachers (4)	Black Teachers (5)	Male Teachers (6)	Female Teachers (7)
Teacher Characteristics							
Non-White	0.11	0.11	0.10	0.00	1.00	0.10	0.11
Math Teacher	0.50	1.00	0.00	0.50	0.47	0.62	0.43
Male	0.35	0.44	0.27	0.36***	0.26	1.00	0.00
Years of Experience	14.89	15.35	14.44	15.17	15.01	15.56	14.53
	(10.76)	(10.74)	(10.77)	(10.80)	(11.28)	(11.61)	(10.25)
$\leq$ Three years of experience	0.16	0.14	0.19	0.15***	0.21	0.16	0.16
No teaching certificate	0.17	0.15	0.18	0.16***	0.21	0.21	0.14
Major in subject taught	0.48	0.47	0.49	0.49	0.48	0.42	0.51
Has graduate degree	0.47	0.48	0.47	0.49	0.45	0.51	0.46
Student Demographics							
American Indian	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Asian	0.08	0.08	0.08	0.07***	0.05	0.09	0.07
Black	0.10	0.10	0.10	0.09***	0.47	0.09	0.11
Hispanic	0.12	0.12	0.12	0.10	0.12	0.13	0.11
Multiple Race	0.04	0.04	0.04	0.04	0.04	0.05	0.04
Male	0.47	0.47	0.47	0.47	0.44	0.51	0.45
Observations	12130	6060	6060	10830	470	4300	7820

Notes: This table presents means of variables where teachers are the unit of analysis. Standard deviations for non-binary variables are reported in parentheses. Asteriks in column (4) denotes significance from t-test of mean difference between White and Black teachers at the 1% significance level. All sample sizes are rounded to the nearest 10 in accordance with NCES regulations for restricted data.

Table 3: Teacher Expectations Production Function

	ELA Teacher Exp.						Math Teacher Exp.			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HH income 20K - 35K	0.07**			0.02	0.02	0.10***			0.05**	0.05**
	(0.03)			(0.02)	(0.02)	(0.03)			(0.02)	(0.02)
HH income 35K - 75K	0.19***			0.09***	0.09***	0.19***			0.08***	0.08***
	(0.02)			(0.02)	(0.02)	(0.02)			(0.02)	(0.02)
HH income 75K - 100K	0.25***			0.11***	0.11***	0.27***			0.13***	0.13***
	(0.03)			(0.02)	(0.02)	(0.03)			(0.02)	(0.02)
HH income > 100K	0.29***			0.12***	0.12***	0.30***			0.13***	0.13***
	(0.03)			(0.02)	(0.02)	(0.03)			(0.02)	(0.02)
Student is American Indian		-0.21**		-0.04	-0.04		-0.25**		-0.08	-0.08
		(0.10)		(0.07)	(0.07)		(0.10)		(0.10)	(0.10)
Student is Asian		0.11***		0.05**	0.05**		0.12***		0.05**	0.05**
		(0.03)		(0.02)	(0.02)		(0.03)		(0.02)	(0.02)
Student is Black		-0.17***		-0.01	-0.00		-0.21***		-0.05**	-0.05**
		(0.03)		(0.02)	(0.02)		(0.03)		(0.02)	(0.02)
Student is Hispanic		-0.17***		-0.04*	-0.04*		-0.16***		-0.03	-0.03
		(0.03)		(0.02)	(0.02)		(0.02)		(0.02)	(0.02)
Student is Multiple Race		-0.05		0.00	-0.00		-0.07**		-0.02	-0.02
		(0.03)		(0.03)	(0.03)		(0.03)		(0.03)	(0.03)
GPA for all 9th grade courses			0.35***	0.33***	0.33***			0.35***	0.34***	0.34***
			(0.01)	(0.01)	(0.01)			(0.01)	(0.01)	(0.01)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Characteristics	No	No	No	No	Yes	No	No	No	No	Yes
Observations	6060	6060	6060	6060	6060	6060	6060	6060	6060	6060
R^2	0.28	0.27	0.49	0.50	0.50	0.28	0.28	0.50	0.50	0.51
Adjusted R^2	0.19	0.18	0.43	0.44	0.44	0.20	0.19	0.44	0.44	0.44

Notes: $*p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$. The dependent variable is a binary indicator equal to one if the teacher expects the student to complete a four-year college degree or more, and zero otherwise. Parentheses contain standard errors that are robust to clustering at the school level. Estimates are from OLS regressions of equation (1). Student socioeconomic status (SES) controls, teacher controls, and 9th grade GPA are included in all specifications. Student SES controls include indicators for household income and mother's educational attainment as well as indicators for student race, sex, and if a language other than English is spoken at home. Teacher controls include teacher race and gender dummies, years of experience, and whether or not the teacher majored in the subject he or she teachers. School FE refers to school fixed effects.

Table 4: Transition Matrices of Disagreements in Teacher Expectations

Math Teacher Expectation	English Teacher Expectation: HS or Less	Some College	Bachelor's or More	Total
	All Students (N = 6060)			
HS or Less	7.12	4.79	1.62	13.54
Some College	4.32	9.90	9.27	23.49
Bachelor's or More	1.62	8.36	52.99	62.97
Total	13.06	23.06	63.88	100.00
	White Students (N = 3970)			
HS or Less	5.47	4.16	1.21	10.82
Some College	4.18	9.56	9.09	22.83
Bachelor's or More	1.33	8.31	56.71	66.35
Total	10.97	22.02	67.00	100.00
	Black Students (N = 610)			
HS or Less	13.18	8.07	2.63	23.88
Some College	7.08	12.52	12.03	31.63
Bachelor's or More	2.63	8.07	33.77	44.48
Total	22.90	28.67	48.43	100.00

Notes: HS denotes high school. Each entry reports the percentage of observations that fall in the particular math teacher expectation-ELA teacher expectation category.

Table 5: OLS Estimates of Effect of Expectations on Educational Attainment

	All Students									White	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ELA Teacher Exp.	0.48***		0.31***	0.30***	0.24***	0.15***	0.14***	0.17***	0.16***	0.14***	0.17**
	(0.01)		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.08)
Math Teacher Exp.		0.48***	0.31***	0.31***	0.25***	0.16***	0.13***	0.15***	0.13***	0.14***	0.11
		(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.07)
Teacher Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student SES	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9th Grade GPA	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
School FE	No	No	No	No	No	No	Yes	No	Yes	Yes	Yes
Teacher Dyad FE	No	No	No	No	No	No	No	Yes	No	No	No
Observations	6060	6060	6060	6060	6060	6060	6060	3600	3600	3970	610
R^2	0.22	0.22	0.28	0.30	0.34	0.37	0.45	0.59	0.46	0.48	0.65
Adjusted R^2	0.22	0.22	0.28	0.29	0.34	0.37	0.38	0.18	0.37	0.39	0.31

Notes: $^*p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. The dependent variable is a binary indicator equal to one if the student completed a four-year college degree or more, and zero otherwise. Parentheses contain standard errors that are robust to clustering at the school level. Estimates are from OLS regressions of equation (2). Student socioeconomic status (SES) controls include indicators for household income and mother's educational attainment as well as indicators for student race, sex, and if a language other than English is spoken at home. Teacher controls include teacher race and gender dummies, years of experience, and whether or not the teacher majored in the subject he or she teaches. School FE refers to school fixed effects and Teacher Dyad FE refers to Math-ELA teacher pair fixed effects. Estimates in Column (9) are from a school fixed effects model, but estimated on the subsample of students for whom teacher dyad fixed effects are identified.

Table 6: Average Partial Effects of Teacher Expectation on Educational Outcomes

	All Students						White	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome: Never enroll in college								
ELA Teacher Exp.	-0.18***		-0.13***	-0.12***	-0.09***	-0.06***	-0.06***	-0.02
	(0.01)		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Math Teacher Exp.		-0.17***	-0.10***	-0.10***	-0.08***	-0.05***	-0.05***	-0.04
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Outcome: Enroll but not complete college								
ELA Teacher Exp.	-0.25***		-0.16***	-0.15***	-0.12***	-0.06***	-0.07***	-0.11**
	(0.01)		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
Math Teacher Exp.		-0.26***	-0.18***	-0.17***	-0.14***	-0.08***	-0.09***	-0.01
		(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)
Outcome: Complete College								
ELA Teacher Exp.	0.44***		0.28***	0.27***	0.21***	0.13***	0.12***	0.13***
	(0.01)		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Math Teacher Exp.		0.43***	0.28***	0.27***	0.22***	0.13***	0.14***	0.05
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Teacher Controls	No	No	No	Yes	Yes	Yes	Yes	Yes
Student SES	No	No	No	No	Yes	Yes	Yes	Yes
9th Grade GPA	No	No	No	No	No	Yes	Yes	Yes
Observations	6060	6060	6060	6060	6060	6060	3970	610

Notes: *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is a categorical variable equal to 0 if student has never enrolled in college, 1 if the student has enrolled, but not completed, college, and 2 if the student has completed a 4-year college degree. Parentheses contain standard errors that are robust to clustering at the school level. We report average partial effects computed from estimates from a multinomial logit regression. Student socioeconomic status (SES) controls include indicators for household income and mother's educational attainment as well as indicators for student race, sex, and if a language other than English is spoken at home. Teacher controls include teacher race and gender dummies, years of experience, and whether or not the teacher majored in the subject he or she teaches.

Table 7: Mechanisms: How Teacher Expectations Affect Outcomes

	GPA		Hrs/wk, Tot. Hwk		Expectations	
	(1)	(2)	(3)	(4)	(5)	(6)
ELA Teacher Exp.	0.43***	0.16***	0.37*	0.32	0.12***	0.10***
	(0.02)	(0.02)	(0.18)	(0.18)	(0.02)	(0.02)
Math Teacher Exp.	0.38***	0.11***	0.59**	0.51**	0.09***	0.08***
	(0.02)	(0.02)	(0.19)	(0.19)	(0.02)	(0.02)
Lagged Control	No	Yes	No	Yes	No	Yes
Mean of Dep. Var	3.04	3.04	6.40	6.40	0.82	0.82
Mean of Lagged Var	.	2.98	.	10.53	.	0.86
Observations	5580	5580	5070	5070	5330	5330

Notes: *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable in columns (1)-(2) is 12th grade GPA on a 4.0 scale. The dependent variable in columns (3)-(4) is usual hours per week spent on homework. The dependent variable in columns (5)-(6) is an indicator variable equal to 1 if the student expects to complete college. All dependent variables are measured at the first follow-up survey, when most of the respondents are in the 12th grade. Estimates are from OLS regressions. Lagged control indicates that the lagged dependent variable measured during the initial survey is included as a regressor. Student socioeconomic status (SES) controls and teacher controls are included in all specifications. Student SES controls include indicators for household income and mother's educational attainment as well as indicators for student race, sex, and if a language other than English is spoken at home. Teacher controls include teacher race and gender dummies, years of experience, and whether or not the teacher majored in the subject he or she teachers. Except for column (1), 9th grade GPA is used as an additional control.

Table 8: Testing the Exogeneity of Teacher Bias

Regression No.	Variable	Estimates	Standard Error
1	9th-grade GPA	-0.0723***	(0.0097)
2	$\lvert S_E - S_M \rvert$	0.0023	(0.0014)
3	$\lvert S_E - S_M \rvert$	-0.0005	(0.0031)
	$\lvert S_E - S_M \rvert^2$	0.0002	(0.0002)
4	S Ever Bullied	0.0011	(0.0148)
5	S Got in Fight	-0.0015	(0.0207)
6	S Participated in Science Fair	-0.0163	(0.0193)
7	S Finds Class Interesting	-0.0041	(0.0132)
8	S Ever in College Prep	0.0159	(0.0147)
9	P Thinks S Has Disability	-0.0392	(0.0245)
10	Passive (ELA)	0.0053	(0.0209)
11	Never attentive (ELA)	-0.1887***	(0.0563)
12	Rarely attentive (ELA)	-0.1054***	(0.0329)
13	Sometimes attentive (ELA)	0.0339*	(0.0197)
14	Mostly attentive (ELA)	0.0526***	(0.0127)
15	Strongly agree reading is fun	-0.0290*	(0.0171)
16	Agree reading is fun	0.0007	(0.0133)
17	Disagree reading is fun	0.0073	(0.0138)
18	Hours spent on ELA Homework in school	0.0010	(0.0026)
19	Hours spent on ELA Homework out of school	-0.0001	(0.0022)
20	Total hours on ELA Homework	0.0002	(0.0015)
21	Passive (Math)	-0.0305	(0.0192)
22	Never attentive (Math)	-0.0407	(0.0771)
23	Rarely attentive (Math)	-0.1261***	(0.0314)
24	Sometimes attentive (Math)	0.0062	(0.0209)
25	Mostly attentive (Math)	0.0635***	(0.0129)
26	Strongly agree math is fun	-0.0019	(0.0222)
27	Agree math is fun	-0.0075	(0.0138)
28	Disagree math is fun	0.0053	(0.0130)
29	Hours spent on Math Homework in school	0.0009	(0.0023)
30	Hours spent on Math Homework out of school	0.0004	(0.0021)
31	Total hours on Math Homework	0.0005	(0.0013)

Notes: *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each row reports the coefficient(s) of interest from a unique regression. The number of observations for each regressions and summary statistics of the variables examined here are reported in Table S8 in Appendix B. S refers to the student and P refers to the parent. S_E and S_M are ELA and math standardized test scores, respectively. All regressions control for student race, sex, 9th-grade GPA, household income and mother's educational attainment, indicators for single parent household, if a language other than English is spoken at home, and school fixed effects. In regression 3, the quadratic terms are jointly insignificant (F-stat = 1.63, p-value= 0.20.)

Table 9: Education Production Function Estimates

	Whites	Blacks
γ_E	0.52***	0.50***
	(0.06)	(0.16)
γ_M	0.55***	0.23
	(0.06)	(0.16)
β	0.50***	0.27**
	(0.05)	(0.11)
c	-0.46***	-0.83***
	(0.05)	(0.14)
σ_θ	0.51***	0.80***
	(0.05)	(0.14)
APE		
b_E	0.18***	0.14***
	(0.02)	(0.05)
b_M	0.20***	0.07
	(0.02)	(0.04)
Elasticities		
b_E	0.12***	0.18***
	(0.02)	(0.06)
b_M	0.13***	0.08
	(0.02)	(0.05)
N	3970	610

Notes: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$. Parameter estimates of equation (10) are reported. The dependent variable is a binary indicator equal to one if the student completed a four-year college degree or more, and zero otherwise. Standard errors are computed by constructing the Hessian of the likelihood function using outer product measure. To compute the outer product measure, we calculate two-sided numerical derivatives of the likelihood function for each estimated parameter. In each direction, the derivative is calculated by perturbing each parameter and then computing the likelihood. Standard errors for the average partial effects (APE) and elasticities are calculated using the delta method. Parameters related to ELA teacher expectations are marked with a subscript E, and parameters related to Math teachers are denoted with a subscript M.

Table 10: Teacher Expectation Production Function Estimates

	Whites		Blacks	
	ELA	Math	ELA	Math
	(1)	(2)	(3)	(4)
c	0.58***	0.56***	0.47**	0.53***
	(0.03)	(0.03)	(0.19)	(0.19)
c_D	-0.09	0.23	-0.26	-0.53***
	(0.13)	(0.15)	(0.21)	(0.2)
ϕ	1.47***	1.68***	0.94***	1.38**
	(0.18)	(0.2)	(0.32)	(0.55)
ϕ_D	-0.45	0.00	-0.21	-0.52
	(0.45)	(0.39)	(0.32)	(0.51)
β	0.55***	0.5***	0.44**	0.14
	(0.04)	(0.04)	(0.18)	(0.21)
β_D	0.23	0.16	0.05	0.31
	(0.19)	(0.14)	(0.2)	(0.23)
APE				
D	-0.03	0.06	-0.10*	-0.27***
	(0.04)	(0.04)	(0.06)	(0.07)
N	3970		610	

Notes: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$. Parameter estimates of equation (12) are reported. Standard errors are computed by constructing the Hessian of the likelihood function using outer product measure. To compute the outer product measure, we calculate two-sided numerical derivatives of the likelihood function for each estimated parameter. In each direction, the derivative is calculated by perturbing each parameter and then computing the likelihood. Standard errors for the average partial effects (APE) are calculated using the delta method.

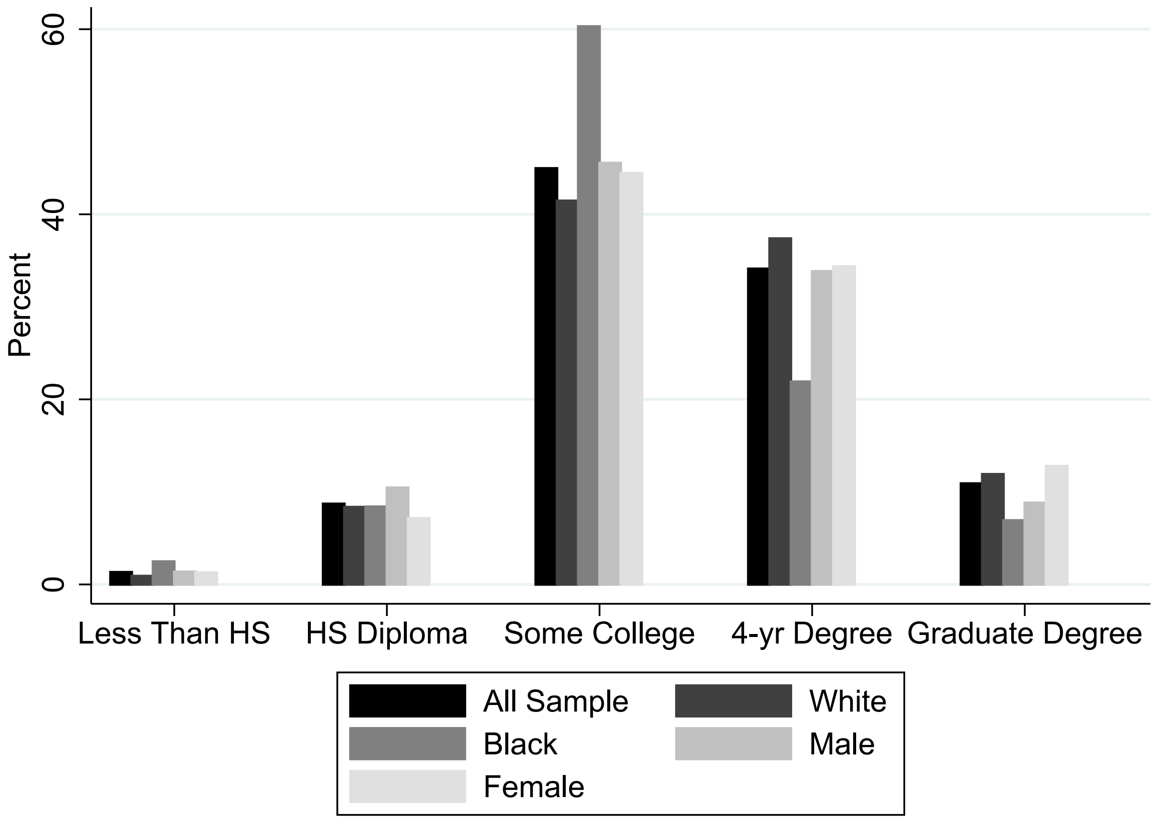
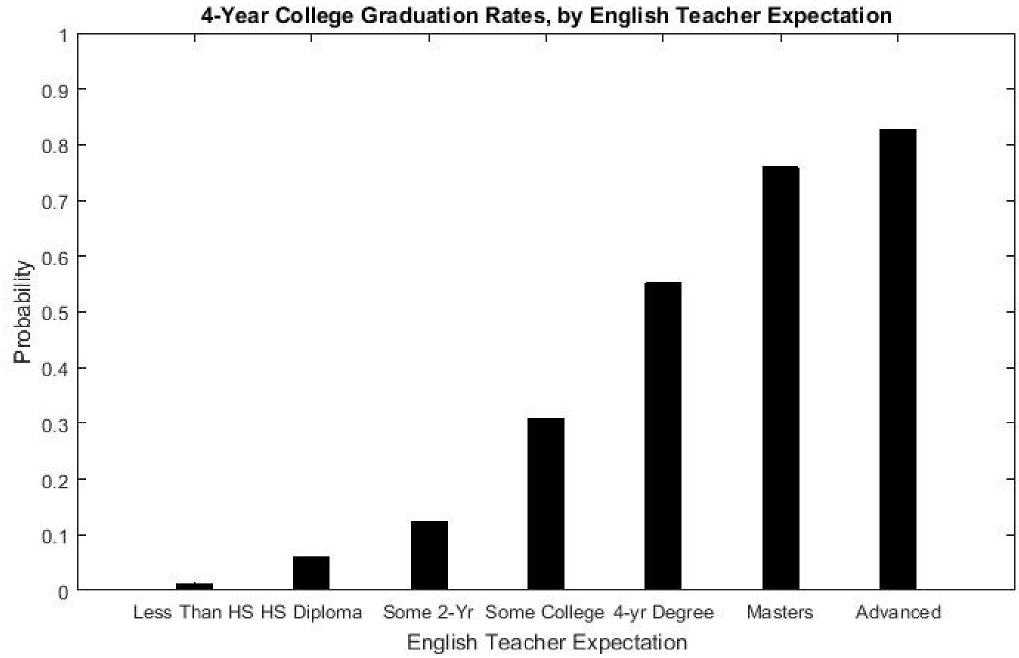
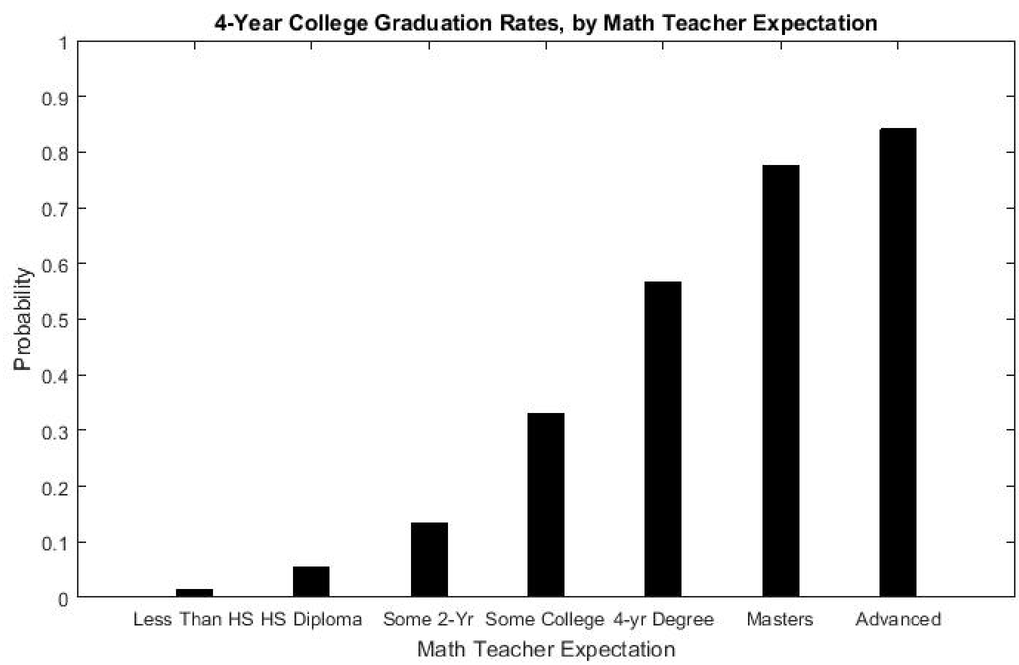


Figure 1: Educational Attainment, by Subgroup. This figure is a histogram of the percentage of the subsample of students who fall in the given educational attainment category. HS is high school. Graduate degree includes masters, Ph.D. and professional degrees.

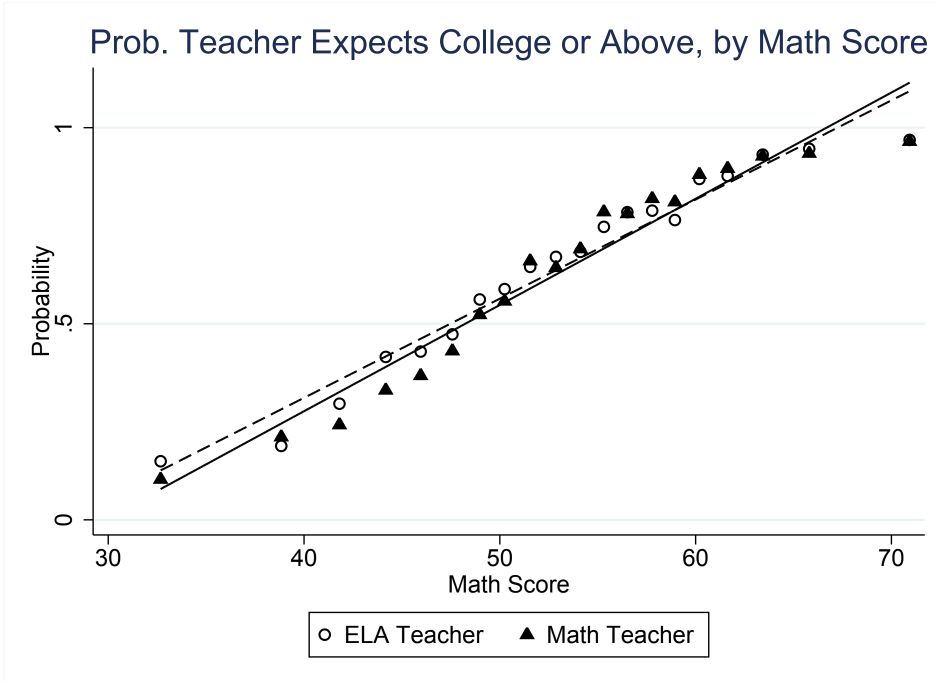


(a)

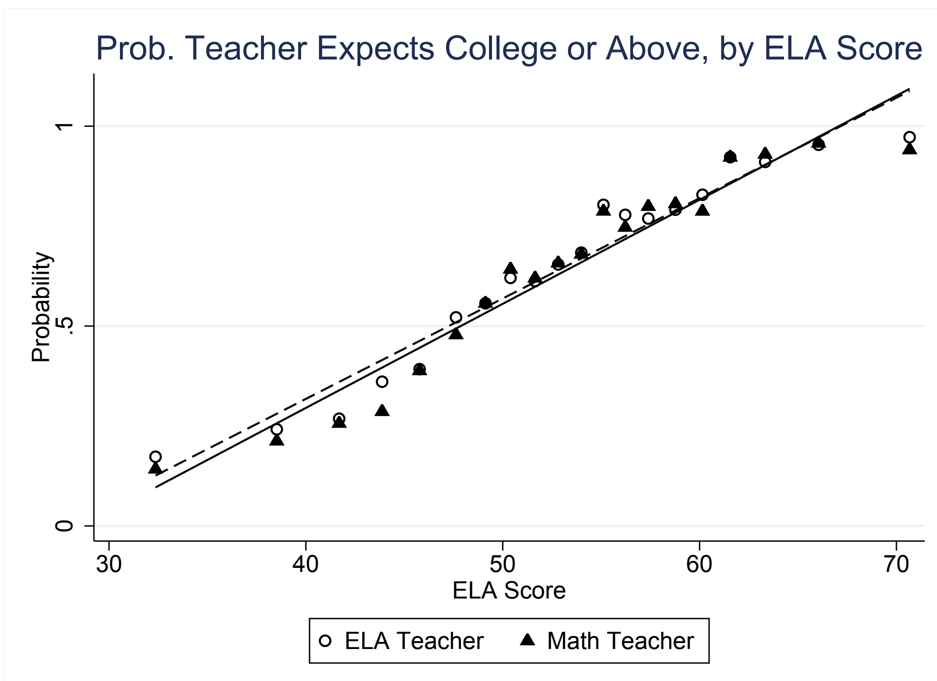


(b)

Figure 2: Teacher Expectations and Student Outcomes. Panel 2(a) shows the percentage of students who complete a four year college degree by ELA teacher expectations. Panel 2(b) plots respective percentages by math teacher expectations.

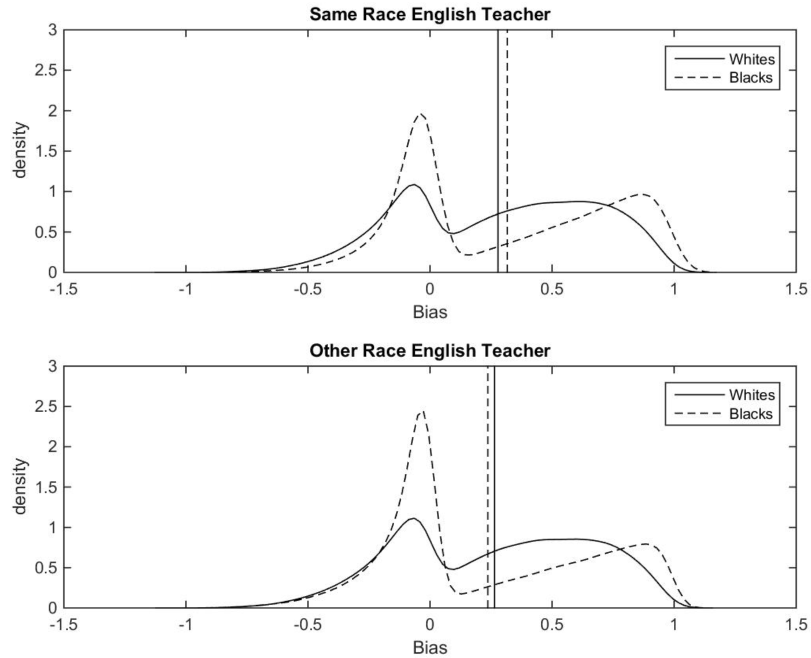


(a)

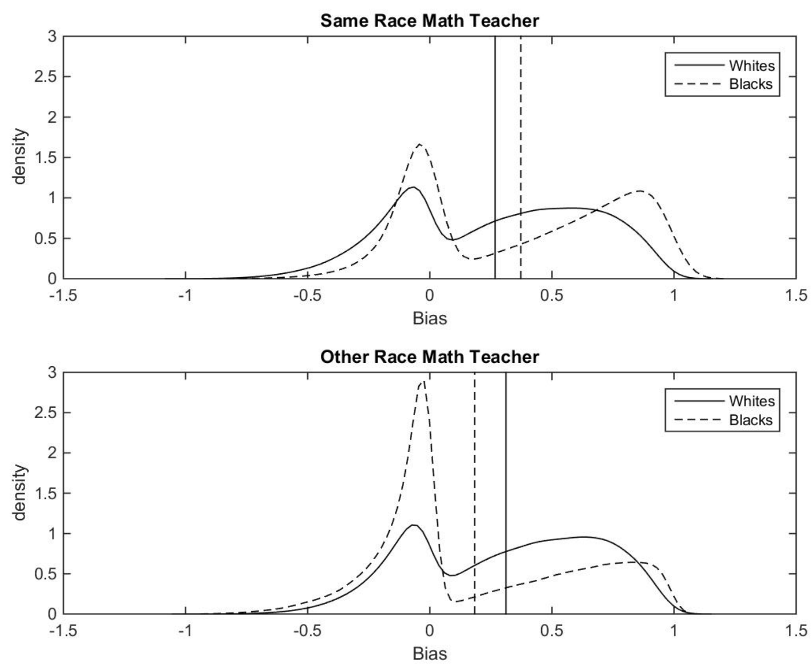


(b)

Figure 3: Math and ELA Scores and Teacher Expectations. This figure shows binned scatterplots, along with fitted regression lines for ELA (dashed lines) and math teacher expectations (solid lines), by math and ELA score, respectively.



(a)



(b)

Figure 4: Distribution of Bias by Student Race. These figures show probability distribution functions of teacher bias for different teacher and student race pairs. Vertical lines represent mean bias. Panel 4(a) shows the distribution of bias for white and black students with same and other race ELA teachers. Panel 4(b) shows the analogous distributions of math teacher bias. Bias is defined in equation (13).

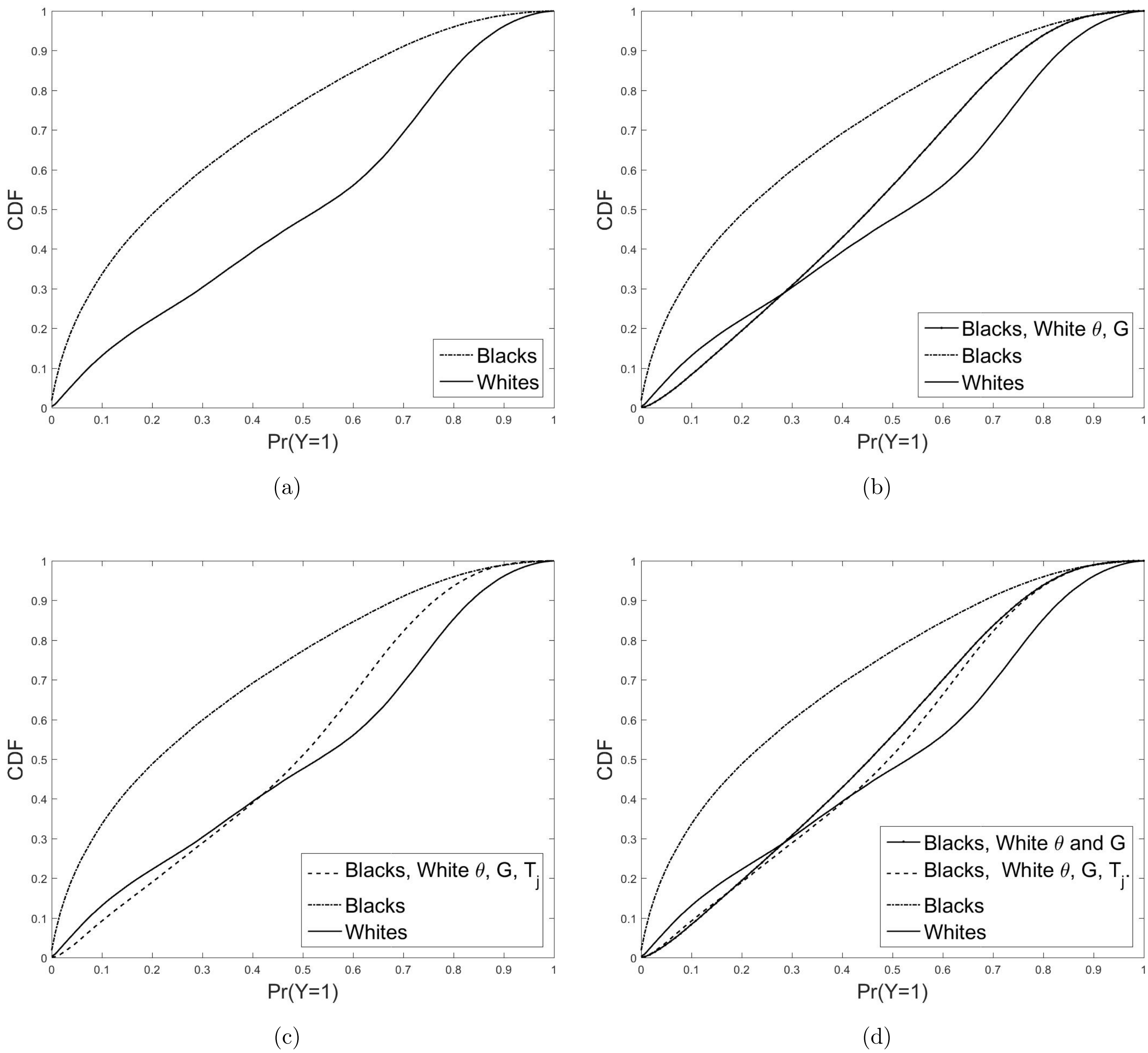


Figure 5: Cumulative Distribution Functions (CDFs) of College Completion Probability for Students of White Teachers. Black and white denote student race. θ is the latent factor that measures the objective probability of completing college (net of GPA and bias), G is 9th-grade GPA, T_j is the the expectation of the subject-j teacher, and Y is a binary indicator for college completion. Panel 5(a) plots the actual CDFs of $Pr(Y = 1)$ for black and white students who have white teachers. Panel 5(b) plots the distribution under the counterfactual in which black students have the same θ and G as white students. Panel 5(c) plots the distribution under the counterfactual in which black students face the same teacher expectation production function *and* the same θ and G as white students. Panel 5(d) combines the three previous plots.

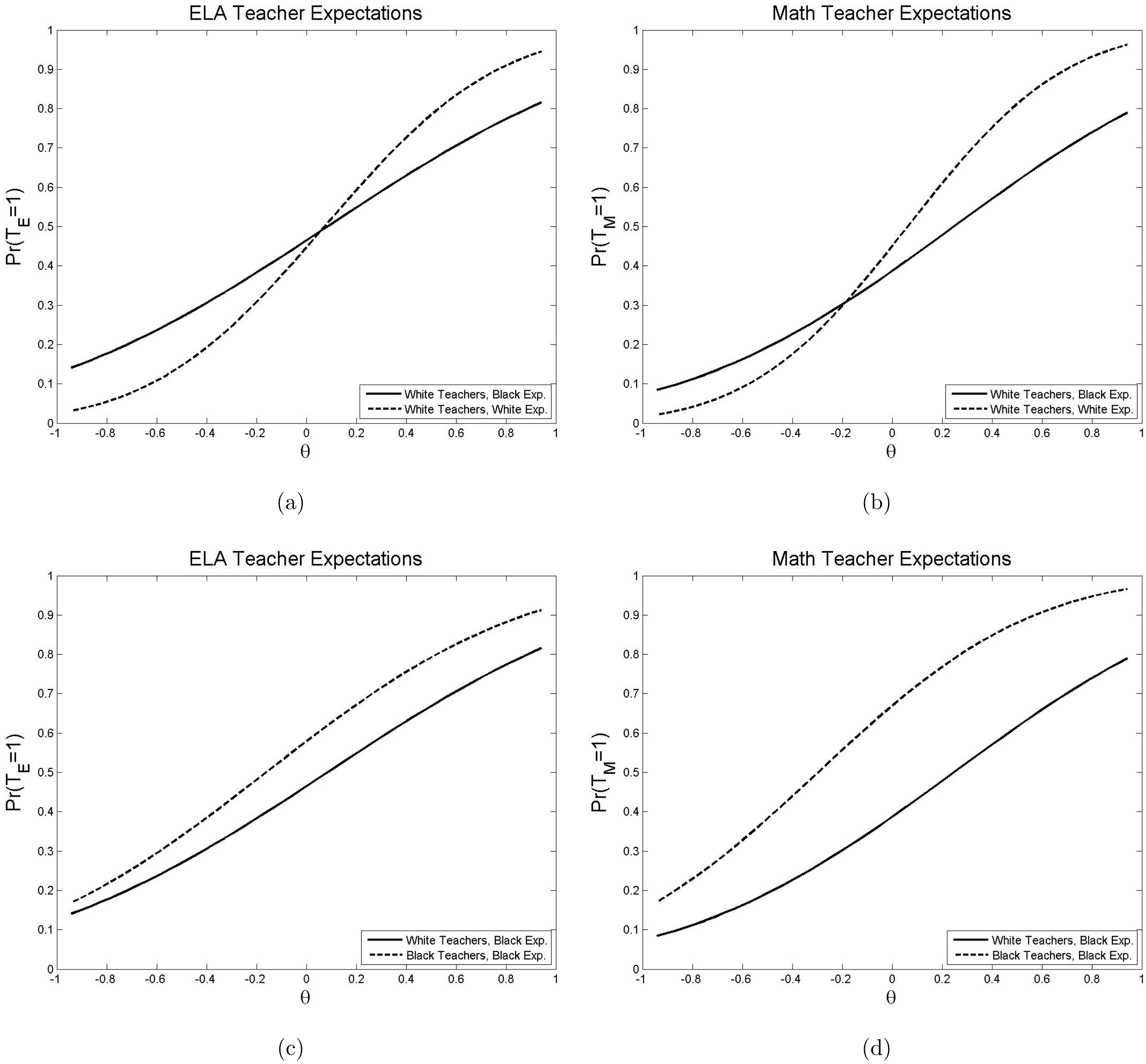


Figure 6: Teacher Expectations for Black Students. θ is the latent factor that measures the objective probability of completing college (net of GPA and bias) and T_j is the expectation of the subject-j teacher. Panel 6(a) shows how teacher expectations change when black students face the same expectation production function from white ELA teachers as white students. Panel 6(b) shows how the expectations change in the counterfactual scenario for math teachers. Panels 6(c) and 6(d), respectively, compare white and black ELA and math teachers' expectation for black students with given θ.