

DISCUSSION PAPER SERIES

IZA DP No. 14096

**Identifying and Producing Effective
Teachers**

Seth Gershenson

FEBRUARY 2021

DISCUSSION PAPER SERIES

IZA DP No. 14096

Identifying and Producing Effective Teachers

Seth Gershenson

American University and IZA

FEBRUARY 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Identifying and Producing Effective Teachers

Teachers are among the most important school-provided determinants of student success. Effective teachers improve students' test scores as well as their attendance, behavior, and earnings as adults. However, students do not enjoy equal access to effective teachers. This article reviews some of the key challenges associated with teacher policy confronted by school leaders and education policymakers, and how the tools of applied economics can help address those challenges. The first challenge is that identifying effective teachers is difficult. Economists use value-added models to estimate teacher effectiveness, which works well in certain circumstances, but should be just one piece of a multi-measure strategy for identifying effective teachers. We also discuss how different policies, incentives, school characteristics, and professional-development interventions can increase teacher effectiveness; this is important, as schools face the daunting challenge of hiring effective teachers, helping teachers to improve, and removing ineffective teachers from the classroom. Finally, we discuss the supply and mobility of teachers, including the consequences of teacher absenteeism, the distribution of initial teaching placements, and the characteristics and preferences of those who enter the profession.

JEL Classification: I2

Keywords: teacher supply, teacher effectiveness, incentive pay, value-added models

Corresponding author:

Seth Gershenson
American University
4400 Massachusetts Ave NW
Washington, DC 20016
USA

E-mail: gershens@american.edu

1. Introduction

Human capital refers to the set of skills that enable one to create value. Economists have long recognized the importance of human capital, though formal study of the topic was popularized relatively recently by the publication of Gary Becker's eponymous book (1980). Formal schooling is an obvious, important, and easily measurable form of human capital that increases one's earning potential, health, and civic engagement (Lochner 2011). Coupled with the fact that communities as a whole benefit from a better educated populace in the form of increased productivity (Moretti 2004) and reduced crime (Lochner and Moretti 2004), it should come as no surprise that individuals and governments alike voluntarily invest in formal schooling.

The importance of formal education from both an individual and societal perspective, as well as the hope that all children are provided equal opportunities to succeed, has long motivated social scientists to identify the malleable determinants of student success. Just as a firm's or country's output can be modeled as a function of inputs (e.g., labor, raw materials), so too can a student's achievement be modeled as a function of school inputs (e.g., class size, teacher quality). Early empirical analyses of so-called *education production functions* are Hanushek (1979) and Murnane (1975).

In the 1990s, a "credibility revolution" took place in empirical economics that led researchers to be more critical of whether observed correlations between inputs (treatments) and outputs were the result of true causal relationships, or were driven by unobserved confounding factors (Angrist and Pischke 2010). Because few, if any, educational inputs are randomly assigned, this newfound scrutiny of empirical results led Todd and Wolpin (2003) to thoughtfully assess the conditions under which various specifications of the education production function identify causal effects of educational inputs on student outcomes.

One of the main insights of Todd and Wolpin, which continues to influence educational research today, is that we can proxy for unobserved historical inputs with a measure of lagged achievement (i.e., the previous year's test score). Controlling for past inputs facilitates apples-to-apples comparisons that might identify the causal effect of current inputs, such as teachers, on academic growth. This gives rise to so-called *value-added models*, or VAMs, which are now a workhorse tool for identifying the impact of school-provided educational inputs, particularly teachers, on student outcomes (Harris 2011; Koedel et al. 2015). VAMs take their name from the fact that they measure the growth in output (typically test scores) from one year to the next and then attribute that growth to specific current educational inputs.

VAM-based analyses of the education production function have identified two general, robust, oft-replicated facts that motivate this article's focus on teaching and teacher education. First, teachers are the most important school-provided educational input. Second, teachers vary widely in their effectiveness, both within- and between schools. These two findings are discussed in greater detail in Section 2, along with an overview of the estimation and specification of VAMs and how VAM-based estimates of teacher effectiveness compare to other measures of teacher effectiveness.

Given the importance of teachers, identifying, training, hiring, retaining, and cultivating effective teachers are first-order concerns for school leaders and educational policy makers. Section 3 reviews what we know about the observable qualifications that do (and do not) predict effective teaching, which could be targeted in teacher recruitment and compensation policies. Predicting who will be an effective teacher is fairly difficult, which complicates teacher hiring and assignment policy. Similarly, Section 4 reviews the evidence on pre-service teacher training and certification programs. Here, prospective teachers invest in and develop their human capital

in hopes of becoming (more) effective educators. There is scant evidence on which types of formal teacher training programs are most effective, though this is an active area of research.

Section 5 provides an overview of teacher labor supply. This includes an analysis of teachers' daily absences, as well as a broader discussion of who enters the teaching profession, the types of schools and classrooms novice teachers enter, and why. Policy plays a big role here too, as incentives and professional supports can be leveraged to influence teacher effort and persistence, both in their current school and in the profession at large. These are important issues, as effective teachers are unevenly distributed across schools and students.

This article is meant to serve as an introductory reading on teacher policy in an advanced undergraduate or graduate course in education policy or the economics of education. Importantly, it is not meant to be a comprehensive handbook or literature review, nor is it meant to fully describe the state-of-the-art methods for measuring teacher effects or conducting program evaluations; relevant review and methods citations are provided throughout. Lovenheim and Turner (2017) provide an accessible textbook introduction to the economics of education, several sections and chapters of which align with the issues relating to teachers discussed here. Murnane and Willet (2010) provide an accessible introduction to the quantitative methods used for causal inference in modern educational research. Similarly, this article is meant to provide the reader with a broad overview of how education economists study teacher effectiveness, teacher policy, and teacher labor markets; specific policy prescriptions or analyses are beyond its scope.

2. Measuring Teacher Effectiveness

Identifying effective teachers is difficult for two reasons. First, many factors relevant to student success, such as students' home environments, are outside of teachers' control. Second, student

and teacher assignments to classrooms are not random. For example, if students predisposed to strong academic performance are systematically assigned to Ms. Apple, we might incorrectly interpret the strong academic performance of her students as evidence of her teaching prowess, when her students would have done well in any classroom. This is the problem of *selection bias* that hinders the estimation of causal effects in observational data; without seeing the *counterfactual*, or what would have happened to Ms. Apple’s students had they had another teacher, we cannot naïvely attribute their success to Ms. Apple (Angrist and Pischke 2010).

It is important that we obtain valid measures of teacher effectiveness because we do not want to penalize or reward teachers for factors outside of their control. Luckily, the value-added models (VAMs) discussed in the introduction adjust for the non-random ways in which students are assigned to classrooms, as most non-random classroom assignments are made based on students’ performance in the prior year (Dieterle et al. 2015). Thus, simply by adjusting for academic achievement in the previous year, VAMs generally deliver valid estimates of teacher effectiveness (Chetty et al. 2014a; Guarino et al. 2015; Kane and Staiger 2008).

Formally, we model the achievement (A) of primary school student i in the self-contained grade- g classroom of teacher j in school s and year t as a linear regression model of the form:

$$A_{igjst} = \alpha A_{i,t-1} + \beta X_{it} + \gamma C_{-i,jt} + \delta_g + \theta_j + \pi_{st} + \varepsilon, \quad (1)$$

where X is a set of pre-determined student characteristics such as socioeconomic status, C is a set of classroom characteristics such as class size and the average prior academic performance of student i ’s peers, δ is a grade-level indicator, θ is the time-invariant teacher effect, π is a school-by-year indicator, and ε is an idiosyncratic error term. Equation (1) is known as a VAM because it relates current inputs, namely the teacher effect θ , to achievement *growth* because it is holding prior achievement constant. While a full discussion of the estimation of equation (1) is beyond

the scope of this article, the usual Ordinary Least Squares (OLS) estimator that treats the teacher effects as fixed effects (teacher indicators) is typically the most valid and robust approach.¹ For further discussion see Harris (2011) for an accessible introduction or Koedel et al. (2015) for a technical overview of best practices for estimation and specification of VAMs.

To identify the effectiveness of an individual teacher, point estimates of each θ_j could be used to rank teachers or to identify which teachers are above or below some threshold in the effectiveness distribution. To document the general importance of teachers in the education production function, analysts typically report the impact of a one standard deviation (SD) increase in teacher effectiveness on student achievement, or other aspects of the distribution of teacher effects such as the gain in achievement attributable to replacing an average teacher with a “good” teacher.

The SD of the teacher effects should be computed via a two-step process (Kane and Staiger 2008), as the SD of the estimated teacher fixed effects will include both sampling error and classroom-specific shocks unrelated to teacher effectiveness. First, take the classroom-level (teacher-year) average of the OLS residuals from estimates of equation (1) that leave the teacher effects in the model’s error term. Second, compute the covariance between each classroom’s average residual and that from a randomly chosen classroom taught by the same teacher in a different year. To avoid the effects of outliers, researchers typically repeat step two several times and then report the median of these multiple estimated SDs.

Value-added analyses of teacher effectiveness conducted in a variety of school contexts typically find that a one-SD increase in teacher effectiveness amounts to a test score gain of 10-20 percent of a test-score SD (Jackson et al. 2014). These effects tend to be larger in math than

reading, perhaps because children are more apt to learn and practice reading at home. Still, these are practically large effects that are at least as large, if not larger, than the impact of other school-provided inputs and amount to about 2-4 months of learning.² Table 1 summarizes some notable VAM-based estimates of teacher effectiveness from different contexts.

[Table 1 here]

The evidence is compelling, then, that effective teachers can significantly improve students' performance on standardized end-of-year math and reading tests. But this is not the sole metric on which teachers should be assessed, for at least three reasons. First, in an age of consequential accountability policies, we might worry that teachers will "teach to the test" at the expense of other non-tested subjects and skills. Second, there is more to good teaching than improving test scores, and some teachers might excel along some dimensions but not others. VAM scores alone will not inform teachers about where or how to improve. Finally, VAMs can only be estimated for teachers who have taught for multiple years in tested grades and subjects; this excludes many early-grade teachers and many secondary teachers.

These concerns have motivated many researchers to estimate variants of equation (1) that replace the achievement measures (A) with other educational outcomes such as absences, non-cognitive or socio-emotional skills, student beliefs and attitudes towards education, and course grades. We summarize some of these studies in Table 2, which generally find that teachers significantly affect their students' attendance habits, non-cognitive skills, and attitudes and beliefs; however, teachers who boost student test scores do not necessarily improve their non-academic outcomes, and vice versa, which reinforces the idea that effective teaching is multidimensional. A related line of research uses detailed administrative data linked across

multiple sources to show that teachers who boost students' test scores also significantly increase their students' lifetime earnings and likelihood of attending college (Chetty et al. 2014b). These findings reduce concerns that VAM-based measures of teacher effectiveness are merely picking up the transitory effects of "teaching to the test" and show how they could be applied to teachers in non-tested grades and subjects, since outcomes like student absences are available for all students.

[Table 2 here]

We do not want to rely on VAM-based measures that are biased (or perceived as biased by important stakeholders) in the sense that they capture factors outside of teachers' control. It is thus important to use *multiple measures* of teacher effectiveness. *Classroom observations* are another useful measure of teacher effectiveness that can be used alongside VAM-based measures or in cases where VAM-based measures are unavailable. Indeed, this is how most teachers are evaluated. The idea here is to have trained, neutral observers visit classrooms and use a rubric to rate teachers' classroom performance (Whitehurst et al. 2014). There are three general advantages of classroom observations. First, they can be conducted on all teachers in all grades and subjects. Second, they are perceived as more fair and transparent measures than their VAM analogs. Third, and perhaps most importantly, observations can provide real-time feedback to teachers on how they can improve. Observations are not a panacea, however, as they are costly and they too can be influenced by factors such as classroom composition, which are outside the teacher's control (Campbell and Ronfeldt 2018; Cohen and Goldhaber 2016).

Performance reviews and subjective evaluations of teachers by principals and students provide another tool for measuring teacher effectiveness. There is a long history of relying on these necessarily subjective measures and, once again, we might worry about implicit or explicit

biases in the principal's and students' ratings. While no measure is perfect, it is interesting and reassuring that principal/student ratings, observation scores, and VAM-based measures of teacher effectiveness tend to be at least modestly correlated (Grossman et al. 2014; Harris and Sass 2014; Sandilos 2019; Rockoff and Speroni 2011). For example, one study of a mid-sized school district in the Western US finds that principals can identify their most and least effective teachers (according to value-added scores) but struggle to differentiate between teachers in the middle of the effectiveness distribution (Jacob and Lefgren 2008). Similarly, principals in Chicago Public Schools were more likely to fire less effective teachers, as measured by VAM (Jacob 2011). Together, multiple measures do a pretty good job of identifying effective teachers, particularly in the tails of the distribution, and in identifying areas in which teachers can seek improvement.

3. Predicting Teacher Effectiveness

Given that teachers vary widely in effectiveness, an obvious and policy relevant question is whether we can predict who will be successful in the classroom a priori. This knowledge would help schools and districts know who to hire and help policymakers and universities recruit young people into the profession. Alas, it is generally difficult to predict who will be an effective teacher prior to observing them in the classroom (Staiger and Rockoff 2010).³ Specifically, the credentials related to teacher pay and certification requirements such as academic course work, knowledge tests, and graduate degrees have little, if any, association with teacher effectiveness.⁴

Two characteristics that are readily observable and do predict teaching effectiveness are teaching experience and teacher race. As is true in many occupations, teachers improve over

time. Improvements are greatest in the first few years, though on average teachers improve throughout their careers (Wiswall 2013). Experience is therefore a characteristic that should be rewarded in the pay scale and considered in hiring decisions.

Teacher race is another important, yet often overlooked, measure of teacher effectiveness. The reason is that same-race teachers produce notable gains in Black and Latino students' achievement, attendance, behavior, and educational attainment (Gershenson et al. 2021; Redding 2019). However, because the teaching force is disproportionately white, many students of color never get to enjoy the benefits of having a same-race teacher. This has led to many calls for a more diverse teaching force (Gershenson et al. 2021; Goldhaber et al. 2019). A summary of the evidence on same-race teacher effects is given in Table 3. Importantly, these effects have been documented for many outcomes, including college enrollment and high school graduation, and at nearly every level of schooling including preschool, primary and secondary school, university, and even law school. They also affect teacher behavior and teacher perceptions of students. See Gershenson et al. (2021) for a thorough discussion of the likely reasons for these race-match effects and how teacher policy can increase both the diversity of the teaching force and the frequency of same-race student-teacher exposures.

[Table 3 here]

We close this section with a discussion of economists' thinking on teacher certification and licensure.⁵ Nearly all US states and OECD (Organization for Economic Co-operation and Development) countries require that teachers hold some form of teaching certification or license. The intuitive rationale for occupational licensure is to ensure that individuals entering a profession have the necessary skills to perform the job safely and effectively (Kleiner 2000;

Nunn 2016). However, there are potential costs as well: rigid licensure policies might prevent effective teachers from entering the profession (Angrist and Guryan 2008; Lovenheim and Turner 2017, 306), disproportionately exclude individuals from underrepresented backgrounds from entering the profession (Goldhaber and Hansen 2010; Taylor et al. 2017), and exacerbate teacher shortages by preventing inter-state moves (Dee and Goldhaber 2017).

The typical US state teaching license requires a university degree, some coursework in education, and perhaps passing a test such as the Praxis. Thirty states also provide some type of incentive for teachers to acquire National Board Certification, which is typically on top of their state licensure requirements. Given the theoretical ambiguities about the impact of such policies, it is an empirical question as to whether the benefits outweigh the costs.

In terms of effectiveness, most of the literature finds modest effects of certification on student achievement (Goldhaber 2011). However, two nuances of these findings merit further discussion. First, in their study of Washington State, Cowan and Goldhaber (2016) find that the effectiveness premium for certified teachers applies only to those who passed the certification test on their first attempt; certified teachers who failed their initial attempt are neither more nor less effective than their non-certified colleagues. Moreover, the authors find a continuous relationship between teachers' scores on the certification test and their effectiveness in the classroom, but no discontinuity or "boost" in effectiveness upon crossing the certification threshold. This suggests that hiring and salary decisions should be nuanced and not simply view certification as a binary characteristic.

Second, in North Carolina, Black teachers who failed their first Praxis attempt but were otherwise qualified to teach could teach on a provisional license and performed just as well in the classroom as white teachers who passed their first attempt (Goldhaber and Hansen 2010).

Because Black students benefit from having same-race teachers and Black teachers are more likely to teach in schools with more Black students, these types of certification requirements may unintentionally reduce teacher effectiveness in schools that serve high shares of Black students and contribute to the lack of diversity in the teaching profession. More generally, taking the certification tests and required courses are costly in terms of both time and money, so mid-career professionals, individuals from low-income or historically under-represented backgrounds, and those who are unsure about their interest in teaching may be screened out by licensure policies.

Concerns about the unintended consequences of stringent teacher certification policies have led some jurisdictions to provide alternative routes into the classroom. One example of this is that some states allow *charter schools*, which are independently operated but publicly funded, to hire some noncertified teachers.⁶ For example, North Carolina only requires that 50 percent of a charter school's teachers be certified, which might help explain why the teaching force in the state's charter schools is more diverse than that in traditional public schools (Gershenson 2020).

Another solution is to allow *alternatively certified teachers* to enter the classroom. Alternatively certified teachers are either individuals allowed to teach while completing the full certification requirements or, perhaps in response to a teacher shortage in a specific geographic or subject area, simply met an alternative, laxer set of criteria. Alternative certification can be controversial both because fully certified teachers (and teacher unions) may resent individuals taking a perceived "short cut" into the profession and parents may worry about the commitment or quality of such teachers.

We have already seen that on average, certified teachers are only modestly more effective than their non- (or alternatively) certified counterparts. However, not all alternative certification

programs are created equal and it is important to know which are the most effective. One large and high-profile alternative certification program, Teach for America (TFA), is a bellwether that receives more than its fair share of both praise and criticism. TFA is a nonprofit organization that aims to recruit high-performing students from elite colleges and universities to commit to teaching for a minimum of two years in a school in a disadvantaged community. In lieu of any formal coursework, TFA candidates receive a 5-week bootcamp prior to entering the classroom, and weekly support throughout the year.

An initial concern with TFA was that the program was placing mostly white, well-off college seniors into schools attended mostly by students of color from low-income households; the pool of TFA teachers has grown more diverse over time, though there is still room for improvement (Del Ciello 2016). Another concern is that TFA teachers might leave teaching after fulfilling their initial two-year commitment (Donaldson and Johnson 2011). It is true that TFA teachers leave earlier than traditionally certified teachers; however, this is more than made up for by the fact that in both primary (Glazerman et al. 2006) and secondary (Xu et al. 2011) schools TFA teachers are significantly more effective than their traditionally-certified counterparts.

4. Producing Effective Teachers

Because teachers play a vital role in the development of students' human capital, yet vary widely in effectiveness, we now consider how both formal and informal training might increase teachers' human capital—and their effectiveness in the classroom. Just like students' learning is influenced by the quality of their teachers and peers, an analogous set of factors might influence teacher quality. This discussion distinguishes between training received prior to entering the classroom (pre-service) and that received while teaching (in-service).

In Section 3 we saw that advanced degrees, certification, and college course-taking predict teacher effectiveness modestly, if at all. A related question is whether some teacher preparation programs (TPPs) are more effective than others. We might expect some variation, since primary and secondary schools vary in their effectiveness and TPPs vary in their focus, resources, and instructional faculties. However, the nascent literature finds little difference across TPPs in their effectiveness (Goldhaber 2018; Koedel et al. 2015; von Hippel et al. 2016). In fact, these studies find more variation in teacher effectiveness *within* programs than between.

Student teaching is a central part of any TPP, where pre-service teachers work with a mentor teacher, helping in the classroom, designing lesson plans, and occasionally teaching a lesson. Here, there are notable within-TPP differences in their graduates' eventual effectiveness in the classroom by the effectiveness (as measured by VAM) of their assigned mentor (Goldhaber et al. 2020a). This is intuitive and consistent with the importance of individual teachers discussed in Section 2. There is great scope for improving the effectiveness of novice teachers by making more strategic assignments of student teachers to mentor teachers (Goldhaber et al. 2020b).

But mentoring and structured professional development (PD) opportunities need not stop once teachers enter the classroom. While traditional in-service PD, which pulls teachers out of the classroom for a full or half day during which they attend seminars, workshops, or lectures is generally ineffective (e.g., Jacob and Lefgren 2004), a more personalized form of in-service PD known as teacher coaching has shown great potential (Kraft et al. 2018; Kraft and Blazar 2018). Unlike traditional in-service PD that delivers generic content to a room full of teachers, coaching is one on one and delivers feedback and advice that is tailored to teachers' unique classrooms, skill sets, and challenges. Coaches observe teachers in the classroom and provide personal

feedback, which means that having quality coaches is key to the success of these interventions, as is having enough coaches to keep their workload manageable (Blazar and Kraft 2015).

A less enticing option is remediation targeted to low-performing teachers. Evidence from Chile finds that teachers who receive remedial training via classes experience a short-lived boost in effectiveness, but their effectiveness decreases while taking the classes and any gains fade out in a few years (Lombardi 2019). Moreover, being identified for remediation lowered teachers' confidence and pride, suggesting a stigma of the program. The main reason for the program's ineffectiveness, though, is likely its lack of personal feedback and communication to teachers.

Teachers also receive informal coaching and mentoring from their supervisors (principals) and their peers and there is mounting evidence that this influences teacher effectiveness. While on average all teachers improve with experience, Kraft and Papay (2014) show that these improvements are greatest in schools that have generally supportive environments. One aspect of a supportive environment is good colleagues. Jackson and Bruegmann (2009) show that teachers improve when they have more effective colleagues (as measured by VAM). These peer effects are strongest for novice teachers and persist over time, which suggests that teachers learn new skills from their colleagues.

Another way that teachers learn and improve is via information and feedback. Consistent with the evidence from teacher-coaching, Taylor and Tyler (2012) show that in Cincinnati, teachers' effectiveness (as measured by VAM) significantly increases after a classroom observation, with the largest gains accruing to teachers who were initially less effective. Of course, classroom observations might also increase teachers' effort by providing *accountability*, i.e., by reminding them that someone is watching.

Indeed, a fundamental assumption in economics is that people respond to incentives, so teachers might respond to both “carrots” (promised rewards) and “sticks” (threatened punishments). In their analysis of Washington, DC’s IMPACT teacher evaluation program, Dee and Wyckoff (2015) find evidence that teachers responded to both types of incentives.⁷ IMPACT threatened low-performing teachers with dismissal and offered high-performing teachers a cash bonus up to about \$25,000 (30 percent of salary) per year. Many low-performing teachers at risk of termination voluntarily left the district, while those who stayed improved. The cash incentive further increased the effectiveness of already effective teachers. Similarly, Biasi (2018) finds that when some districts in Wisconsin changed from seniority-based to flexible pay systems, teacher quality in those districts increased due to changes in both teacher effort and the composition of the teaching force.

However, the broader literature on the impact of financial incentives (i.e., *merit pay*) on teacher effectiveness is decidedly mixed. Several studies find null, modest, heterogeneous, or noisy effects (e.g., Brehm et al. 2017; Hill and Jones 2020; Goodman and Turner 2014; Speroni et al. 2020; Springer et al. 2011). There are a few potential reasons for this. First, it could be that some incentives are too small to create much change or that the effect is only realized among a small share of the teacher workforce. Second, and similarly, it could be that for merit pay to be effective it must be paired with other interventions or educational supports, as suggested by experimental evidence from Tanzania (Mbiti et al. 2019).

Finally, the most intriguing potential explanation comes from *behavioral economics*, which applies insights from psychology to economic models, to better understand how people make decisions.⁸ Specifically, *loss aversion* occurs when individuals care more about a loss of

\$M than a gain of \$M. If teachers, like most people, are loss averse, then how the merit pay policy is framed and implemented could matter greatly. Fryer et al. (2012) test this hypothesis by conducting an experiment in Illinois in which treated teachers are given financial bonuses at the start of the year, with the understanding that those bonuses must be returned at the end of the year if their students do not make sufficient learning gains, while control-group teachers partake in a “typical” merit pay program in which they receive the bonus at the end of the year if their students make sufficient gains. Importantly, the bonus was the same size for both groups, yet the authors find large learning gains (about 20 percent of a SD) in the treated “loss averse” group, but small, statistically insignificant effects of incentive pay in the control group. This result provides an important insight into how the power of incentives, and behavioral economics, might be used to increase teacher performance.

5. Teacher Labor Supply

Teacher labor supply refers to teachers’ willingness to work for a given wage, benefits, and working conditions. Labor economists view individuals’ employment decisions through a supply-and-demand framework, and teachers are no different (Lovenheim and Turner 2017, Chapter 12). Like anyone else, teachers have preferences about where and when they would like to work and there are varying levels of demand for their services across locales; this section reviews what we know about teachers’ preferences and their implications for student achievement. We discuss various dimensions of teachers’ labor supply such as daily labor supply (absences), initial placements and mobility across schools, and entry into the profession.

Existing teachers face a daily labor supply decision, as they always have the option of being absent. Anywhere from 5-10 percent of teachers in the US are absent on a given day (Roza

2007) and teacher absences are even more common in less developed countries (Chaudhury et al. 2006; Banerjee and Duflo 2006). Teacher absences are costly in at least four ways (Gershenson 2012). First, they are financially costly in the sense that US school districts pay over 1 billion dollars each year for substitute teachers to cover those absences. Second, there is an in-kind cost to non-absent teachers who in one way or another have to help cover their colleague's absence. Third, teacher absences are detrimental to student achievement, which we discuss in greater detail below, and finally, if teacher absences are not evenly distributed across schools, then the costs associated with teacher absences may exacerbate socioeconomic or racial disparities in educational outcomes.

There is now a robust literature that identifies the causal effect of teacher absences on student learning in numerous contexts, including NYC (Herrmann and Rockoff 2012), North Carolina (Clotfelter et al. 2009; Gershenson 2016 b), India (Duflo et al. 2012), Zambia (Das et al. 2007), and a large urban district in the Northeastern US (Miller et al. 2008). The effects are large: over the course of a school year, replacing a regular teacher with a substitute teacher is equivalent to replacing an average teacher with one at the 10th or 20th percentile of the effectiveness distribution, or about 10 percent of a test-score SD (Herrmann and Rockoff 2012). Moreover, the harmful effects of teacher absences likely exacerbate socio-demographic achievement gaps, as teacher absences are more common and more likely to go unfilled by substitute teachers in low-performing schools and schools serving low-income communities (Clotfelter et al. 2009; Miller et al. 2008; Gershenson 2012; Liu et al. 2020).⁹

Absences are harmful because in addition to generally disrupting the classroom, on average, substitute teachers are less effective than regular teachers. This raises the question of

whether teacher absences can be reduced. Some are unavoidable, of course, due to illness and family commitments, but some are discretionary. We know this because policies that hold teachers accountable for student achievement tend to reduce teacher absences. For example, Jacob (2013) shows that in Chicago Public Schools probationary teachers' absence rates significantly decreased once principals were given the ability to dismiss probationary teachers. Similarly, Gershenson (2016) shows that in North Carolina, teacher absences fell by about 10 percent in schools subject to sanctions under the No Child Left Behind Act (NCLB).

Financial incentives can reduce teacher absenteeism as well. An experiment in India finds that tying teacher salaries to regular attendance—and monitoring attendance with digital cameras—significantly reduced teacher absence rates and increased students' test scores (Duflo et al. 2012). In North Carolina, Ahn (2013) finds that relatively small bonuses of \$750 or \$1,500 tied to student achievement gains caused teachers to increase effort (reduce absences) when their receipt of the bonus was uncertain. Together, these studies show that while harmful, teacher absences are responsive to an array of policies and interventions that can both save money and improve student achievement.

The next dimension of teacher labor supply to consider is *where* teachers work. Two main facts drive this discussion. The first is known as the *draw of home*: teachers prefer to work in schools and districts similar to, and nearby (if not the same), those they attended as children (Boyd et al. 2005; Reiningger 2012). The second is that conditional on becoming a teacher, schools' working conditions are more important than salaries in determining where teachers want to work (Bacolod 2007 a). For example, an influential study of North Carolina's public-school teachers found that on average, all teachers, but particularly white teachers (who compose more than 80 percent of the teaching force), tend to leave lower-performing schools and schools

serving underrepresented minority student populations for higher-achieving schools and schools serving mostly white student populations (Hanushek et al. 2004).

Together, these two facts contribute to the inequitable distribution of teacher effectiveness across schools (Goldhaber et al. 2015; Lankford et al. 2002). So-called *hard to staff* schools, those with chronically low performance and high shares of economically disadvantaged students, have a harder time both recruiting and retaining teachers. The result is higher rates of turnover that lead to more novice teachers, which we have seen predicts teacher ineffectiveness, and more generally creates disruptions that harm student achievement in the schools that can least afford additional challenges (Hanushek et al. 2018; Ronfeldt et al. 2013; Sorensen and Ladd 2020). This unequal exposure to effective teachers and to teacher turnover undoubtedly contributes to persistent socioeconomic and racial disparities in students' educational outcomes.

Finally, we consider teacher labor supply on the extensive margin: who enters the profession, and why. Given the importance of teachers, there are many surveys of teachers that allow researchers to describe the characteristics of new and current teachers. The demographic composition of the teacher force has been fairly stable over the past several decades: it is disproportionately female (about 70-80 percent) and white (about 85 percent). One reason that women are overrepresented is the flexibility, and the ability to exit and return at the same rank and salary level, which appeals to women whose careers are interrupted by childbearing (Farkas et al. 2000; Flyer and Rosen 1997). Indeed, Grissom and Reiningger (2012) show that a nontrivial number of female teachers leave teaching for a few years and then return when their children are older. Redding and Baker (2019) show that racial gaps in teacher entry are primarily due to racial gaps in choosing education majors among college students, though racial disparities in college enrollment and completion contribute as well. Survey and qualitative interview data also suggest

that altruism and a desire to help others motivate individuals to become teachers (Farkas et al. 2000; Watt and Richardson 2012).

But how do teachers compare to nonteachers? Representative samples of the full population, not just of teachers, can shed light on this question. Salaries, the perceived job security, and comfortable retirement benefits certainly cause people to consider teaching (Bacolod 2007a). So too does parents' profession, as the children of teachers are significantly more likely to enter teaching themselves (Jacinto and Gershenson 2020). This likely contributes to the relatively static racial composition of the teaching force.

Finally, given that teaching is historically dominated by women, and for a time was one of the only professions available to women, it is important to consider how the opening of the broader labor market to women has influenced who enters teaching. A clever study by Bacolod (2007b) shows that between 1960 and 1990 the "quality" of new female teachers, as measured by test scores and the selectivity of their undergraduate institutions, fell. Meanwhile, the quality of women entering other professions such as law, medicine, and business increased. Interpreted through the lens of the Roy (1951) model of selection this is an intuitive outcome, as the opening of these higher-paying professions to women caused high-skill women, who in previous generations would have become teachers, to instead pursue more lucrative careers. However, more recently the quality of incoming teachers seems to be increasing among cohorts entering the profession in the 1990s and 2000s due to new efforts to recruit talented individuals (Master et al. 2018; Nguyen and Redding 2018), though this is occurring while increased state-level accountability pressure has reduced both the supply and quality of new teachers (Kraft et al. 2020). Once again, we see that teachers' labor supply is malleable.

6. Conclusion

This article introduces some of the myriad issues regarding teacher labor markets that are studied using the insights and tools of applied economics. Some of the topics discussed here are politically contentious and some of the literature bases yield mixed results. It is important to recognize that the impact of any policy or intervention is context dependent, implementation of those policies and interventions matters, and there may be challenges to delivering an effective intervention at scale. Due to data availability and the author's biases, most studies discussed here are from the US context, and among those studies from North Carolina, Florida, and Texas are overrepresented because these states were early in granting researchers access to high-quality administrative data.

Nonetheless, many of the key insights discussed here have been replicated in enough contexts and proved robust enough to inform our thinking on education policy moving forward. First and foremost, teacher policy is central to education policy, as teachers are the most important school-provided input. Good teachers make for good schools. Good teachers make the teachers around them better and improve students' learning, behavior, and ultimately their success in the labor market. However, identifying effective teachers is difficult and requires multiple measures of teacher effectiveness.

It is also clear that teachers improve over the course of their career and the best way to facilitate this is to provide generally supportive environments along with personalized feedback and instruction. However, traditional markers of teacher quality such as graduate degrees and certification are less predictive of teacher effectiveness and might even hinder efforts to recruit a racially diverse and representative teaching force. Teacher diversity is teacher quality, as there are large immediate and long-run benefits of being exposed to a same-race teacher, but in the

US, too many students of color never have that opportunity due to the unrepresentative, disproportionately white teaching force.

Finally, teachers, like anyone else, respond to incentives. However, those incentives must be well designed to avoid unintended consequences. They must also be substantial enough to alter teacher behavior, otherwise they are wasted resources. Insights from behavioral economics show how we can do more with less in this regard. The policies, practices, and interventions described in this article are critical to ensuring that we provide *all* students with access to effective instruction, which is a necessary but not sufficient condition to ensuring that all children have equal opportunities to develop the human capital necessary to flourish in the modern economy.

Notes

1. An important caveat is that in a secondary school setting in which students have multiple teachers per day, the model and estimation are necessarily more complicated (Jackson 2014; Mansfield 2015).
2. Based on annual (nine month) gains of 0.4-1.0 test-score SD in the primary grades, which are larger in the earlier grades (Baird and Pane 2019).
3. Empirically, education economists try to isolate the causal effect of observed teacher characteristics by replacing teacher fixed effects in a value-added model like that specified in equation (1) with a vector of the observed teacher characteristics.
4. Interestingly, a detailed survey of new teachers in New York City found that while individual survey items were not predictive of teaching effectiveness, they did load on to two factors that did, which might be interpreted as measuring the teacher's cognitive and non-cognitive skills (Rockoff et al. 2011).
5. See Goldhaber (2011) for a through discussion of teacher licensure around the world.
6. See Epple et al. (2016) for a thorough review of the broader literature on charter schools.
7. The authors estimate the effects of IMPACT using a regression discontinuity design (Murnane and Willett 2010, chapter 9), a powerful tool for causal inference, which intuitively compares teachers whose VAM scores were just above and just below an arbitrary threshold that determined eligibility for the incentives.
8. See Thaler (2016) and Kahneman (2011) for accessible introductions to behavioral economics.
9. See Gershenson (2012) and Liu et al. (2020) for a detailed discussion of substitute teacher labor markets.

References

- Aaronson, Daniel, Lisa Barrow, and William Sander (2007), "Teachers and student achievement in the Chicago public high schools," *Journal of Labor Economics*, 25 (1): 95-135.
- Ahn, Tom (2013), "The missing link: Estimating the impact of incentives on teacher effort and instructional effectiveness using teacher accountability legislation data," *Journal of Human Capital*, 7 (3): 230-273.
- Angrist, Joshua D., and Jonathan Guryan (2008), "Does teacher testing raise teacher quality? Evidence from state certification requirements," *Economics of Education Review*, 27 (5): 483-503.
- Angrist, Joshua D. and Jörn-Steffen Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics," *Journal of Economic Perspectives*, 24 (2): 3–30, doi: 10.1257/jep.24.2.3.
- Bacolod, Marigee (2007a), "Who teaches and where they choose to teach: College graduates of the 1990s," *Educational Evaluation and Policy Analysis*, 29 (3): 155-168.
- Bacolod, Marigee P (2007b), "Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality," *The Review of Economics and Statistics*, 89 (4): 737-751.
- Baird, Matthew D., and John F. Pane (2019), "Translating standardized effects of education programs into more interpretable metrics," *Educational Researcher*, 48 (4): 217-228.
- Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, (2010), "Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278," *Economic Policy Institute*.
- Banerjee, Abhijit, and Esther Duflo (2006), "Addressing absence," *Journal of Economic perspectives*, 20 (1): 117-132.
- Bau, Natalie, and Jishnu Das (2020), "Teacher value added in a low-income country," *American Economic Journal: Economic Policy*, 12 (1): 62-96.
- Becker, Gary S (1980), *Human capital*, Vol. 3, Chicago: University of Chicago press.
- Biasi, Barbara (2018), *The labor market for teachers under different pay schemes*, No. w24813, National Bureau of Economic Research.
- Blazar, David, and Matthew A. Kraft (2015), "Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment," *Educational evaluation and policy analysis*, 37 (4): 542-566.

- Blazar, David, and Matthew A. Kraft (2017), "Teacher and teaching effects on students' attitudes and behaviors," *Educational evaluation and policy analysis*, 39(1), 146-170.
- Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff (2005), "The draw of home: How teachers' preferences for proximity disadvantage urban schools," *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 24 (1): 113-132.
- Brehm, Margaret, Scott A. Imberman, and Michael F. Lovenheim (2017), "Achievement effects of individual performance incentives in a teacher merit pay tournament," *Labour Economics*, 44: 133-150.
- Campbell, Shanyce L., and Matthew Ronfeldt (2018), "Observational evaluation of teachers: Measuring more than we bargained for?," *American Educational Research Journal*, 55 (6): 1233-1267.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers (2006). "Missing in action: teacher and health worker absence in developing countries," *Journal of Economic perspectives*, 20 (1): 91-116.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014 a), "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 104 (9):2593–2632. doi:10.1257/aer.104.9.2593
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014 b), "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104 (9): 2633-79.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2007), *How and why do teacher credentials matter for student achievement? Working Paper No. w12828*, National Bureau of Economic Research.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2009), "Are teacher absences worth worrying about in the U.S.?" *Education Finance and Policy*, 4(2):115–149.
- Cohen, Julie, and Dan Goldhaber. "Building a more complete understanding of teacher evaluation using classroom observations," *Educational Researcher*, 45 (6): 378-387.
- Cowan, James, and Dan Goldhaber (2016), "National board certification and teacher effectiveness: Evidence from Washington State," *Journal of Research on Educational Effectiveness*, 9 (3): 233-258.
- Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan (2007), "Teacher shocks and student learning: Evidence from Zambia," *Journal of Human Resources*, 42(4):820–862.

- Dee, Thomas S. (2005), "A teacher like me: Does race, ethnicity, or gender matter?," *American Economic Review*, 95 (2): 158-165.
- Dee, Thomas S., and Dan Goldhaber (2017), "Understanding and addressing teacher shortages in the United States," *The Hamilton Project*, Brookings Institution.
- Dee, Thomas S., and James Wyckoff (2015), "Incentives, selection, and teacher performance: Evidence from IMPACT," *Journal of Policy Analysis and Management*, 34 (2): 267-297.
- Del Ciello, Deanna (2016), "Teach For America: Growing More Diverse," *Education Week*, Available online: <https://www.edweek.org/teaching-learning/teach-for-america-growing-more-diverse>, (accessed December 26, 2020).
- Delhommer, Scott (2019), "High School Role Models and Minority College Achievement". Working paper. University of Texas, Austin.
- Dieterle, Steven, Cassandra M. Guarino, Mark D. Reckase, and Jeffrey M. Wooldridge (2015), "How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added," *Journal of Policy Analysis and Management*, 34 (1): 32-58.
- Donaldson, Morgaen L., and Susan Moore Johnson (2011), "Teach For America Teachers: How Long Do They Teach? Why Do They Leave?," *Phi Delta Kappan*, 93 (2): 47–51, doi: 10.1177/0031721711109300211.
- Duflo, Esther, Rema Hanna, and Stephen P. Ryan (2012), "Incentives work: Getting teachers to come to school," *American Economic Review*, 102(4):1241-78.
- Egalite, Anna J., Brian Kisida, and Marcus A. Winters (2015), "Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement," *Economics of Education Review* 45: 44–52, doi: 10.1016/j.econedurev.2015.01.007
- Epple, Dennis, Richard Romano, and Ron Zimmer (2016). "Charter schools: A survey of research on their characteristics and effectiveness," *Handbook of the Economics of Education*, 5:139-208.
- Farkas, Steve, Jean Johnson, and Tony Foleno (2000), *A Sense of Calling: Who Teaches and Why. A Report from Public Agenda*, New York, NY.
- Flyer, Fredrick, and Sherwin Rosen, (1997), "The new economics of teachers and education," *Journal of Labor Economics*, 15 (1, Pt. 2): S104-S139. <https://doi.org/10.1086/209858>
- Fryer Jr, Roland G., Steven D. Levitt, John List, and Sally Sadoff (2012), "Enhancing the efficacy of teacher incentives through loss aversion: A field experiment (No. w18237)," National Bureau of Economic Research.

- Gershenson, Seth (2012), "How do substitute teachers substitute? An empirical study of substitute-teacher labor supply," *Economics of Education Review*, 31 (4): 410-430.
- Gershenson, Seth (2016 a), "Linking teacher quality, student attendance, and student achievement," *Education Finance and Policy*, 11 (2): 125-149.
- Gershenson, Seth (2016 b), "Performance standards and employee effort: Evidence from teacher absences," *Journal of Policy Analysis and Management*, 35 (3): 615-638.
- Gershenson, Seth (2019), "Student-Teacher Race Match in Charter and Traditional Public Schools," *Thomas B. Fordham Institute*.
- Gershenson, Seth, Constance A. Lindsay, and Michael Hansen (2021). *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom*, Cambridge: Harvard Education Press.
- Gershenson, Seth, Cassandra M.D. Hart, Joshua Hyman, Constance Lindsay, and Nicholas W. Papageorge (2018), "The long-run impacts of same-race teachers (Working Paper No. 25254)," *National Bureau of Economic Research*, <https://doi.org/10.3386/w25254>.
- Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge (2016), "Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations," *Economics of Education Review*, 52: 209-24, doi: 10.1016/j.econedurev.2016.03.002.
- Glazerman, Steven, Daniel Mayer, and Paul Decker (2006), "Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes," *Journal of Policy Analysis and Management*, 25 (1): 75-96.
- Goldhaber, Dan (2011), "Licensure: Exploring the value of this gateway to the teacher workforce," *Handbook of the Economics of Education*, 3:315-339.
- Goldhaber, Dan (2019), "Evidence-based teacher preparation: Policy context and what we know," *Journal of Teacher Education*, 70 (2): 90-101.
- Goldhaber, Dan, John Krieg, and Roddy Theobald (2020 a), "Effective like me? Does having a more productive mentor improve the productivity of mentees?," *Labour Economics*, 63: 101792.
- Goldhaber, Dan, John Krieg, Natsumi Naito, and Roddy Theobald (2020 b), "Making the most of student teaching: The importance of mentors and scope for change," *Education Finance and Policy*, 15 (3): 581-591.
- Goldhaber, Dan, and Michael Hansen (2010), "Race, gender, and teacher testing: How informative a tool is teacher licensure testing?," *American Educational Research Journal*, 47 (1): 218-251.

- Goldhaber, Dan D., Pete Goldschmidt, and Fannie Tseng (2013), "Teacher value-added at the high-school level: Different models, different answers?," *Educational Evaluation and Policy Analysis*, 35 (2): 220-236.
- Goldhaber, Dan, Lesley Lavery, and Roddy Theobald (2015), "Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students," *Educational researcher*, 44 (5): 293-307.
- Goldhaber, Dan, Roddy Theobald, and Christopher Tien (2019), "Why we need a diverse teacher workforce," *Phi Delta Kappan*, 100 (5): 25-30.
- Goodman, Sarena F., and Lesley J. Turner (2013), "The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program," *Journal of Labor Economics*, 31 (2): 409-420.
- Grissom, Jason A., and Christopher Redding (2016), "Discretion and Disproportionality: Explaining the Underrepresentation of High-Achieving Students of Color in Gifted Programs," *AERA Open*, doi: 10.1177/2332858415622175.
- Grissom, Jason A., and Michelle Reininger (2012), "Who comes back? A longitudinal analysis of the reentry behavior of exiting teachers," *Education Finance and Policy*, 7(4): 425–454. https://doi.org/10.1162/EDFP_a_00075
- Grossman, Pam, Julie Cohen, Matthew Ronfeldt, and Lindsay Brown (2014), "The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment," *Educational Researcher*, 43 (6): 293-303.
- Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge (2015), "Can value added measures of teacher performance be trusted?," *Education Finance and Policy*, 10 (1) :117–156.doi:10.1162/EDFP_a_00153
- Hanushek, Eric A. (1979), "Conceptual and empirical issues in the estimation of educational production functions," *Journal of human Resources*: 351-388.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin (2004), "Why public schools lose teachers," *Journal of Human Resources*, 39(2): 326–354.
- Hanushek, Eric A., Steven G. Rivkin, and Jeffrey C. Schiman (2016), "Dynamic effects of teacher turnover on the quality of instruction," *Economics of Education Review*, 55: 132-148.
- Harris, Douglas N. (2011), *Value-Added Measures in Education: What Every Educator Needs to Know*, Cambridge: Harvard Education Press.

- Harris, Douglas N., and Tim R. Sass (2014), "Skills, productivity and the evaluation of teacher performance." *Economics of Education Review*, 40: 183-204.
- Hart, Cassandra M. D. (2020), "An Honors Teacher Like Me: Effects of Access to Same-Race Teachers on Black Students' Advanced-Track Enrollment and Performance," *Educational Evaluation and Policy Analysis*, 42 (2): 163–87.
- Herrmann, Mariesa A., and Jonah E. Rockoff (2012), "Worker absence and productivity: Evidence from teaching," *Journal of Labor Economics*, 30: 749–782.
- Hill, Andrew J., and Daniel B. Jones (2020), "The Impacts of Performance Pay on Teacher Effectiveness and Retention Does Teacher Gender Matter?," *Journal of Human Resources*, 55(1): 349-385.
- Holt, Stephen B., and Seth Gershenson (2019), "The impact of demographic representation on absences and suspensions," *Policy Studies Journal*, 47 (4): 1069-1099.
- Jacinto, Alberto, and Seth Gershenson (2020), "The Intergenerational Transmission of Teaching," *American Educational Research Journal*.
- Jackson, C. Kirabo (2014), "Teacher quality at the high school level: The importance of accounting for tracks," *Journal of Labor Economics*, 32(4): 645-684.
- Jackson, C. Kirabo (2018), "What do test scores miss? The importance of teacher effects on non-test score outcomes." *Journal of Political Economy*, 126 (5): 2072-2107.
- Jackson, C. Kirabo, and Elias Bruegmann (2009), "Teaching students and teaching each other: The importance of peer learning for teachers," *American Economic Journal: Applied Economics* 1(4): 85-108.
- Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger (2014), "Teacher effects and teacher-related policies," *Annu. Rev. Econ.* 6(1): 801-825.
- Jacob, Brian A. (2013), "The effect of employment protection on teacher effort," *Journal of Labor Economics*, 31 (4): 727-761.
- Jacob, Brian A., and Lars Lefgren (2004), "The impact of teacher training on student achievement quasi-experimental evidence from school reform efforts in Chicago," *Journal of Human Resources*, 39 (1): 50-79.
- Jacob, Brian A., and Lars Lefgren (2008), "Can principals identify effective teachers? Evidence on subjective performance evaluation in education," *Journal of labor Economics*, 26 (1): 101-136.

- Jacob, Brian A. (2011), "Do principals fire the worst teachers?," *Educational Evaluation and Policy Analysis*, 33 (4): 403-434.
- Joshi, Ela, Sy Doan, and Matthew G. Springer (2018), "Student-Teacher Race Congruence: New Evidence and Insight from Tennessee," *AERA Open*, 4 (4): 1–25, doi: 10.1177/2332858418817528.
- Kahneman, Daniel (2011). *Thinking, fast and slow*. Macmillan.
- Kane, Thomas J., and Douglas O. Staiger (2008), "Estimating teacher impacts on student achievement: An experimental evaluation," *NBER Working Paper No. 14607*.
- Kleiner, Morris M. (2000), "Occupational licensing," *Journal of Economic Perspectives*, 14 (4): 189-202.
- Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff (2015), "Value-added modeling: A review," *Economics of Education Review*, 47: 180-195.
- Koedel, Cory, Eric Parsons, Michael Podgursky, and Mark Ehlert (2015), "Teacher preparation programs and teacher quality: Are there real differences across programs?," *Education Finance and Policy*, 10 (4): 508-534.
- Kraft, Matthew A. (2019), "Teacher effects on complex cognitive skills and social-emotional competencies," *Journal of Human Resources*, 54(1):1-36.
- Kraft, Matthew A., and David Blazar (2018), "Taking teacher coaching to scale: Can personalized training become standard practice?," *Education Next*, 18 (4): 68-75.
- Kraft, Matthew A., and John P. Papay (2014), "Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience," *Educational evaluation and policy analysis*, 36 (4): 476-500.
- Kraft, Matthew A., David Blazar, and Dylan Hogan (2018), "The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence," *Review of educational research*, 88 (4): 547-588.
- Kraft Matthew A., Eric J. Brunner, Shaun M. Dougherty, and David J. Schwegman (2020), "Teacher Accountability Reforms and the Supply and Quality of New Teachers," *Journal of Public Economics*, 188: 104212.
- Lankford, Hamilton, Susanna Loeb, and James and Wyckoff (2002), "Teacher sorting and the plight of urban schools: A descriptive analysis," *Educational evaluation and policy analysis*, 24(1): 37-62.

- Lindsay, Constance A., and Cassandra M. D. Hart (2017), "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina," *Educational Evaluation and Policy Analysis*, 39 (3): 485–510, doi: 10.3102/0162373717693109.
- Liu, Jing, and Susanna Loeb (2021), "Engaging teachers: Measuring the impact of teachers on student attendance in secondary school," *Journal of Human Resources*, 56 (2): 343–379.
- Liu, Jing, Susanna Loeb, and Yingand Shi (2020), "More Than Shortages: The Unequal Distribution of Substitute Teaching," *Education Finance and Policy*: 1-48.
- Lochner, Lance (2011), "Nonproduction benefits of education: Crime, health, and good citizenship." *Handbook of the Economics of Education*, 4: 183-282.
- Lochner, Lance, and Enrico Moretti (2004), "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports," *American economic review*, 94 (1): 155-189.
- Lombardi, María (2019), "Is the remedy worse than the disease? The impact of teacher remediation on teacher and student performance in Chile," *Economics of Education Review*, 73: 101928.
- Lovenheim, Michael, and Sarah E. Turner (2017), *Economics of education*. Macmillan Higher Education.
- Mansfield, Richard K (2015), "Teacher quality and student inequality," *Journal of Labor Economics*, 33 (3): 751-788.
- Master, Benjamin, Min Sun, and Susanna Loeb (2018), "Teacher workforce developments: Recent changes in academic competitiveness and job satisfaction of new teachers," *Education Finance and Policy*, 13(3): 310-332.
- Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani (2019), "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania," *The Quarterly Journal of Economics*, 134 (3): 1627-1673.
- Miller, Raegen T., Richard J. Murnane, and John B. Willett (2008), "Do teacher absences impact student achievement? Longitudinal evidence from one urban school district," *Educational Evaluation and Policy Analysis*, 30 (2): 181–200.
- Moretti, Enrico (2004), "Human capital externalities in cities," *Handbook of regional and urban economics*, 4: 2243-2291.
- Murnane, Richard J. (1975), "The impact of school resources on the learning of inner-city children."
- Murnane, Richard J., and John B. Willett (2010), *Methods matter: Improving causal inference in*

educational and social science research. Oxford University Press.

- Nguyen, Tuan D., and Christopher Redding (2018), "Changes in the demographics, qualifications, and turnover of American STEM teachers, 1988–2012." *AERA Open* 4 (3): 2332858418802790.
- Nunn, Ryan (2016), "Occupational Licensing and American Workers," *The Hamilton Project*.
- Redding, Christopher (2019), "A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes," *Review of educational research*, 89 (4): 499-535.
- Redding, Christopher, and Dominique J. Baker (2019), "Understanding racial/ethnic diversity gaps among early career teachers," *AERA Open*, 5 (2): 1–17.
<https://doi.org/10.1177/2332858419848440>
- Reininger, Michelle (2012), "Hometown disadvantage? It depends on where you're from: Teachers' location preferences and the implications for staffing schools," *Educational Evaluation and Policy Analysis*, 34 (2): 127-145.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005), "Teachers, schools, and Academic Achievement," *Econometrica*, 73 (2): 417–458. doi:10.1111/j.1468-0262.2005.00584.x
- Rockoff, Jonah E. (2004), "The impact of individual teachers on student achievement: Evidence from panel data," *American Economic Review*, 94(2): 247–252.
doi:10.1257/0002828041302244
- Rockoff, Jonah E., and Cecilia Sperroni (2011), "Subjective and objective evaluations of teacher effectiveness: Evidence from New York City," *Labour Economics*, 18 (5): 687-696.
- Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger (2011), "Can you recognize an effective teacher when you recruit one?," *Education finance and Policy*, 6 (1): 43-74.
- Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff (2013), "How teacher turnover harms student achievement," *American educational research journal*, 50 (1): 4-36.
- Rothstein, Jesse (2010), "Teacher quality in educational production: Tracking, decay, and Student Achievement," *Quarterly Journal of Economics*, 125 (1): 175–214.
doi:10.1162/qjec.2010.125.1.175
- Roy, Andrew Donald. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3 (2): 135-146.

- Roza, Marguerite (2007), *Frozen Assets*, Washington, DC: Education Sector Reports.
- Sandilos, Lia E., Wesley A. Sims, Kate E. Norwalk, and Linda A. Reddy (2019), "Converging on quality: Examining multiple measures of teaching effectiveness," *Journal of school psychology*, 74: 10-28.
- Slater, Helen, Neil M. Davies, and Simon Burgess (2012), "Do teachers matter? Measuring the variation in teacher effectiveness in England," *Oxford Bulletin of Economics and Statistics*, 74 (5): 629-645.
- Sorensen, Lucy C., and Helen F. Ladd (2020), "The hidden costs of teacher turnover," *AERA Open*, 6 (1): 2332858420905812.
- Speroni, Cecilia, Alison Wellington, Paul Burkander, Hanley Chiang, Mariesa Herrmann, and Kristin Hallgren (2020), "Do Educator Performance Incentives Help Students? Evidence from the Teacher Incentive Fund National Evaluation," *Journal of Labor Economics*, 38 (3).
- Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J. R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher (2011), "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)," *Society for Research on Educational Effectiveness*.
- Staiger, Douglas O., and Jonah E. Rockoff (2010), "Searching for effective teachers with imperfect information," *Journal of Economic perspectives*, 24 (3): 97-118.
- Taylor, Eric S., and John H. Tyler (2012), "The effect of evaluation on teacher performance," *American Economic Review*, 102 (7): 3628-51.
- Taylor, Erika D., Stacey Pelika, and Andy Coons (2017), "To What Extent Are Ethnic Minority Teacher Candidates Affected by High-Stakes Assessments? NEA Research Brief NBI No. 16," *National Education Association*.
- Thaler, Richard H. (2016), "Behavioral economics: Past, present, and future," *American Economic Review*, 106 (7): 1577-1600.
- Todd, Petra E., and Kenneth I. Wolpin (2003), "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113 (485): F3-F33.
- Tran, Long, and Seth Gershenson (2021), "Experimental Estimates of the Student Attendance Production Function," *Educational Evaluation and Policy Analysis*. doi: 10.3102/0162373720984463
- Vinopal, Katie (2018), "Understanding Individual and Organizational Level Representation: The Case of Parental Involvement in Schools," *Journal of Public Administration Research and Theory*, 28 (1): 1–15, doi: 10.1093/jopart/mux036.

- von Hippel, Paul T., Laura Bellows, Cynthia Osborne, Jane Arnold Lincove, and Nick Mills (2016), "Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different?," *Economics of Education Review*, 53: 31-45.
- Watt, Helen MG, and Paul W. Richardson (2012), "An introduction to teaching motivations in different countries: Comparisons using the FIT-Choice scale," *Asia-Pacific Journal of Teacher Education*, 40 (3): 185-197.
- Whitehurst, Grover J., Matthew M. Chingos, and Katharine M. Lindquist (2014), "Evaluating teachers with classroom observations," *Brown Center on Education Policy: Brookings Institute*.
- Wiswall, Matthew (2013), "The dynamics of teacher quality," *Journal of Public Economics*, 100 (2013): 61-78.
- Xu, Zeyu, Jane Hannaway, and Colin Taylor (2011), "Making a difference? The effects of Teach for America in high school," *Journal of Policy Analysis and Management*, 30 (3): 447-469.

Table 1. Value-added Estimates of a 1-SD Increase in Teacher Effectiveness on Test Scores

Math SD (1)	Reading SD (2)	Locale (3)	Level (4)	Study (5)
0.11	0.10	New Jersey	Primary	Rockoff (2004)
0.11	0.10	Texas	Primary	Rivkin et al. (2005)
0.15	0.11	North Carolina	Primary	Rothstein (2010)
0.16	0.12	NYC	Primary	Chetty et al. (2014 a)
0.19	0.11 ^a	Pakistan	Primary	Bau and Das (2020)
0.13	-	Chicago	Secondary	Aaronson et al. (2007)
0.07	0.02	North Carolina	Secondary	Jackson (2014)
0.13	0.06	US	Secondary	Goldhaber et al. (2013)
0.17 ^b	-	England	Secondary	Slater et al. (2012)

a Urdu, the national language of Pakistan.

b Averaged across math, science, and reading.

Abbreviations: standard deviations (SD), New York City (NYC).

Table 2. Value-added Estimates of Teacher Effects on Non-academic Outcomes

Outcome (1)	Effect (2)	Locale (3)	Level (4)	Study (5)
Absences	-0.07	North Carolina	Primary	Gershenson (2016 a)
Absences	-50%	Large urban district	Secondary	Liu and Loeb (2020)
Grit / Effort	0.14	6 large MET (US) districts	Primary	Kraft (2019)
“Non-cognitive skill” index ^a	0.08	North Carolina	Secondary	Jackson (2018)
Self-reported: self-efficacy	0.14	4 large US districts	Primary	Blazar and Kraft (2017)
happiness	0.31			
behavior	0.15			
College at 20	37%	NYC	Primary	Chetty et al. (2014 b)
Earnings at 28	1.3%			
Teenage birth	4.6%			

a Index is a weighted average of grades, on-time grade progression, absences, and suspensions. Effect refers to effect of a 1-standard deviation (SD) increase in teacher effectiveness unless otherwise noted. MET refers to the Measures of Effective Teaching Study.

Table 3. Effects of Race Match on Student and Teacher Outcomes

Outcome (1)	Effect (2)	Locale (3)	Level (4)	Study (5)
Math score SD	0.08	TN	Primary	Joshi et al. (2018)
Math score SD	0.03	NC	Primary	Clotfelter et al. (2007)
Math score SD	0.03	FL	Primary	Egalite et al. (2015)
HS graduation	7%	TN STAR	Primary	Gershenson et al. (2018)
College enrollment	13%			
HS graduation	3%	TX	Secondary	Delhommer (2019)
College enrollment	3%			
Chronic Absence	10%	NC	Primary	Holt and Gershenson (2019)
Chronic Absence	26%	TN STAR	Primary	Tran and Gershenson (2021)
Suspension	5%	NC	Primary	Lindsay and Hart (2017)
Parental involvement	8%	US (ECLS)	Primary	Vinopal (2018)
Take advanced course	7%	NC	Secondary	Hart (2020)
Teacher has high expectations	30%	US (ELS)	Secondary	Gershenson et al. (2016)
Teacher has negative view of student behavior	30 to 50%	US (NELS)	Primary	Dee (2005)
Recommended to gifted and talented program	200%	US (ECLS)	Primary	Grissom and Redding (2016)

Notes: Negative signs are omitted for negative outcomes: same-race teachers *reduce* the likelihood of “bad” outcomes such as chronic absence or suspensions. STAR refers to the Tennessee STAR class-size reduction experiment. ECLS refers to the Early Childhood Longitudinal Study-Kindergarten Cohort. ELS refers to the 2001 Educational Longitudinal Study. NELS refers to the 1988 National Educational Longitudinal Study. ECLS, ELS, and NELS are all conducted by the US Department of Education.