GERALD R. FORD SCHOOL OF PUBLIC POLICY
EDUCATION POLICY INITIATIVE
UNIVERSITY OF MICHIGAN

Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools

Brian Jacob
Jonah E. Rockoff
Eric S. Taylor
Benjamin Lindy
Rachel Rosen

Education Policy Initiative
Gerald R. Ford School of Public Policy
735 S. State Street
Ann Arbor, Michigan 48109

EPI Working Paper 2016 | March, 2016

TEACHER APPLICANT HIRING AND TEACHER PERFORMANCE: EVIDENCE FROM DC PUBLIC SCHOOLS

Brian Jacob
Jonah E. Rockoff
Eric S. Taylor
Benjamin Lindy
Rachel Rosen

Originally posted as NBER Working Paper No. 22054

We first thank the District of Columbia Public Schools, in particular Michael Gaskins, Anna Gregory, Brooke Miller, Jason Kamras, and Scott Thompson. Generous financial support was provided by the Smith Richardson Foundation. We received helpful comments and suggestions from seminar participants at Brown, Chicago, Clemson, Cornell, Delaware, Johns Hopkins, Kentucky, LSU, New York Fed, NYU, Paris School of Economics, Princeton, Stanford, UC Santa Barbara, APPAM, and AEFP. The authors of this publication were consultants to the District of Columbia Public Schools. The terms of this relationship and this publication have been reviewed and found to be in accordance with the DCPS policy on objectivity in research by the Office of Talent and Culture and by the Office of Instructional Practice District of Columbia Public Schools.

Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools
Brian Jacob, Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen
March 2016
JEL No. I2,J2,M51

ABSTRACT

Selecting more effective teachers among job applicants during the hiring process could be a highly cost-effective means of improving educational quality, but there is little research that links information gathered during the hiring process to subsequent teacher performance. We study the relationship among applicant characteristics, hiring outcomes, and teacher performance in the Washington DC Public Schools (DCPS). We take advantage of detailed data on a multi-stage application process, which includes written assessments, a personal interview, and sample lessons, as well as the annual evaluations of all DCPS teachers, based on multiple criteria. We identify a number of background characteristics (e.g., undergraduate GPA) as well as screening measures (e.g., applicant performance on a mock teaching lesson) that strongly predict teacher effectiveness. Interestingly, we find that these measures are only weakly, if at all, associated with the likelihood of being hired, suggesting considerable scope for improving teacher quality through the hiring process.

Brian Jacob
Gerald R. Ford School of Public Policy
University of Michigan
735 South State Street
Ann Arbor, MI 48109
and NBER
bajacob@umich.edu

Jonah E. Rockoff
Columbia University
Graduate School of Business
3022 Broadway #603
New York, NY 10027-6903
and NBER
jonah.rockoff@columbia.edu

Eric S. Taylor
Harvard Graduate School of Education
Gutman Library 469
6 Appian Way
Cambridge, MA 02138
eric_taylor@gse.harvard.edu

Benjamin Lindy
1110 Main St.
Cincinnati, OH 45202
benjamin.lindy@gmail.com

Rachel Rosen
MDRC
16 E. 34th St., 19th Floor
New York, NY 10016
rachel.rosen@mdrc.org

"The best means of improving a school system is to improve its teachers. One of the most effective means of improving the teacher corps is by wise selection."

Ervin Eugene Lewis, Superintendent of Schools, Flint, Michigan, 1925

The idea that teacher selection is important for improving education is not new, but it finds much support in recent research documenting variation in teachers' impacts on student short- and long-run outcomes (Chetty et al. 2014a, 2014b, Jackson 2013, 2014). While some have pointed out the benefits of removing teachers who perform poorly on the job (Gordon et al. 2006, Hanushek 2011, Goldhaber and Theobald, 2013, Adnot et al. 2016), improving selection among existing applicants may be even more *cost effective* because it avoids substantial costs: exposing students to ineffective teachers (Staiger and Rockoff, 2010) and compensating teachers for increased risk of job separation (Rothstein, 2015).[1]

Nevertheless, establishing rigorous methods to select individuals likely to become successful teachers has proven difficult. Limited progress has not been due to a lack of attention by academics, but rather to the use of small samples, the focus on a small set of teacher characteristics found in administrative data, and the lack of high quality performance measures on teachers.[2] One conclusion from this work is that selection using basic credentials such as certification and completion of graduate education is likely to yield few benefits. Economists, though latecomers to the issue of teacher quality, have consistently found that these credentials have little or no power to explain variation in performance across teachers (Rockoff 2004, Rivkin et al. 2005, Kane et al. 2008, Harris and Sass 2011).

[1] In addition, collection of performance data on teachers (e.g. standardized student testing, classroom observation, portfolios of student work) requires significant public resources and often entails difficult labor negotiations (e.g., Baker and Santora 2013) while schools and school districts have wide freedom to require applicants to submit information as part of the hiring process.
[2] Morsh and Wilder (1954) provide an extensive review of hundreds of studies conducted over the first half of the 20th century, beginning with Meriam (1906).

In this paper, we use uncommonly detailed data on teacher job applications, employment, and performance in the Washington, DC Public Schools (hereafter DCPS) to gain insights into how various measures might be used to improve teacher selection at the hiring stage. Our data and setting present several advantages for addressing this issue. First, DCPS implements a centralized multi-stage application process that provides us with a large sample of applicants for whom we have a range of characteristics, including scores on written assessments, personal interviews, and teaching auditions. Second, passing each stage of the application process was based on meeting a particular score threshold, which helps us overcome issues of selection into teaching into DCPS.[3] Third, DCPS conducts annual evaluations of all of its teachers under its "IMPACT" system, under which a wide variety of performance data is collected.[4] This allows us to evaluate teacher performance in all grades and subjects, rather than limiting our sample to teachers whose students take annual standardized tests.

We find that several background characteristics (e.g., undergraduate GPA) as well as screening measures (e.g., applicant interview scores) strongly predict teacher effectiveness. These results are quite robust to various methods to control for selection into the teaching work force in DCPS. Indeed, these measures are only weakly, if at all, associated with the likelihood of receiving a job offer or being hired.[5] Our results suggests that there exists considerable scope for improving teacher quality through the selection process. We find that teachers whose applications are in the top quartile of predicted effectiveness have on-the-job performance that is

[3] As we explain in further detail below, applicants who passed through the entire application process have a discontinuously higher probability of receiving a job offer and working in DCPS, but there is still considerable variation in applicant scores, conditional on passing. This allows us to estimate the relationship of these application scores to both hiring and teacher performance, while accounting for selection into the sample using the passing threshold.
[4] Dee and Wyckoff (2013) describe IMPACT's incentives in greater detail and demonstrate that the program affected teacher turnover and measured performance on various IMPACT components.
[5] We also find that few background characteristics, and none of the application screening measures, are significantly associated with the probability of retention within DCPS.

roughly 0.65 standard deviations higher on average than those whose applications are in the bottom quartile. To underscore the large magnitude of this finding, we compare it to the on-the-job improvement exhibited by DCPS teachers over their initial years on the job, as this is the point in teachers' careers when performance has been consistently shown to improve rapidly (e.g., Rice 2013, Ost 2014, Papay and Kraft 2015). Among the new teachers in our sample who remain in DCPS for three years, their average 3-year growth in performance is only 0.30 standard deviations, less than half of the difference in performance between top- and bottom-quartile applicants entering DCPS.

This paper contributes to the body of recent work on the importance of personnel policies within the economics of education as well as the wider literature in labor and personnel economics on employee selection. As noted in a review by Oyer and Schaeffer (2011), while economists have accomplished much to understand the importance of employee incentives, "the literature has been less successful at explaining how firms can find the right employees in the first place." Recent work has examined the role of personal referrals (Schmutte 2015, Burks et al. 2015), placement agencies (Stanton and Thomas 2013) and, most similar to our setting, objective testing technologies (Hoffman et al. 2015) in the hiring process. With greater access to detailed information within firms and institutions on hiring and performance, a great deal may be learned about the economics of employee selection.

The paper proceeds as follows. Section 2 provides an overview of current knowledge regarding teacher hiring practices and earlier research that informs our study. In Section 3 we describe teacher application, hiring, and evaluation processes in DCPS. We present empirical findings on selection into DCPS, job performance, and attrition in Section 4, and discuss the major conclusions from our findings in Section 5.

2. What Do We Know About Teacher Selection?

We devote this section to providing an overview of existing work on the issue of teacher selection, as much of it comes from fields other than economics. Economists have paid far greater attention to the labor supply of teachers than to labor demand. Many studies examine determinants of attrition and mobility between schools among employed teachers, mainly focusing on the impacts of compensation and student/school characteristics (e.g., Dolton and van der Klaauw, 1995, 1999; Hanushek et al. 2004; Scafidi et al. 2007; Clotfelter et al. 2008; Jackson 2009; Boyd et al. 2011).[6] An assumption (sometimes implicit) in much of this work is that employment outcomes stem from teachers' choices, not those of school or district administration. Below, we restrict our attention to papers focusing more narrowly on the questions of teacher labor demand and employee selection. Other work in economics presents evidence on how teacher characteristics relate to high-stakes evaluations by principals (Jacob and Walsh 2011), and how retention decisions by school principals respond to lowering the cost of dismissal (Jacob 2013) or the provision of new teacher performance data (Rockoff et al. 2012). These papers are related to our study, but deal with on-the-job personnel assessment, rather than hiring.

2.1 Economics Literature

There is a small literature in economics on the demand for teaching applicant characteristics and the effectiveness of teacher selection processes. Ballou (1996) finds that high achieving college graduates with teaching credentials are not more likely to receive a job offer,

[6] There are also studies focusing on how teacher effort, measured by absence, responds to incentives such as employment protection (Jacob 2013), paid absence policies (Jacobson 1989) or school accountability (Ahn 2013).

despite being no less likely to apply for teaching jobs or accept offers given to them. He concludes that "public school officials undervalue cognitive skills and subject matter knowledge when screening new applicants and that hiring decisions are suboptimal as a result." Recent work finds some supporting evidence for this conclusion. Hinrichs (2014) sent resumes with randomly generated characteristics to a representative sample of schools nationwide and measured total rates of response as well as responses with invitations for an interview. This has the distinct advantage of isolating labor demand, but is limited to basic resume credentials and to initial interest, rather than ultimate hiring decisions. He finds that applicants from more selective colleges were slightly more likely to receive a response, but those with higher undergraduate GPA were not. A large effect was found for an applicant's geographic location, with applicants from within the same state having much higher response rates.

Boyd et al. (2011) study the labor market for teachers transferring within the same school district, using data from New York City to discern whether transfer applicants with certain characteristics are more likely to be hired. While the use of applications data allows them to measure and control for some components of labor supply, they cannot observe declined job offers, so this aspect of labor supply could potentially affect their conclusions. Nevertheless, they find that, conditional on application, transfer applicants are more likely to be hired if they possess higher certification exam scores, attended more competitive colleges, had more years of experience, or if their previous students had higher achievement growth. This last result suggests that school principals are able to identify more effective teachers among transfer applicants.

Studying dismissals of probationary teachers in the Chicago Public Schools, Jacob (2011) finds that principals are more likely to release individuals with more absences and lower value-added scores, and less likely to release individuals with stronger educational qualifications measured by things such as the competitiveness of their undergraduate college and whether they had ever failed a teacher certification exam.

In contrast, earlier work by Kane and Staiger (2005) presents evidence from a natural experiment suggesting that schools are not effective at choosing the most effective applicants among new hires. Taking advantage of a state incentive for reduced class sizes, Los Angeles dramatically increased hiring of new elementary teachers in 1997, from around 1,300 per year to over 3,300.[7] Despite the hiring spike (and a clear increase in new hires that did not possess proper credentials), estimates of performance were no worse for teachers hired in 1997 than for cohorts hired in preceding years.

Some recent research has produced promising results regarding how indices of teacher characteristics predict teaching performance, although these have included measures collected in low-stakes research surveys (Rockoff et al. 2011) or administrative data unavailable to schools and school districts (Boyd et al. 2008).[8] Moreover, these studies focus on the characteristics of teachers who are already hired rather than data on the characteristics of a pool of teacher applicants, preventing them from addressing issues of selection bias.

In a study which is most closely related to our work, Goldhaber et al. (2014) examine data from Spokane, Washington, where teacher applications are scored subjectively based on education, qualifications and certifications, experience, letters of recommendation, and narrative

---

[7] Hiring remained high for additional years, and Los Angeles did not increase teacher salaries.

[8] Some suggestive evidence also comes from two alternative certification programs in New York City. Rockoff and Speroni (2011) find that math teachers hired through the Teaching Fellows program were slightly more effective in their first year of teaching if they had a high rating during program selection. Dobbie (2011) finds that an index of eight criteria used to select applicants into the Teach for America (TFA) program are positively related to effectiveness among teachers during their first years of teaching.

statements.[9] They find that teachers with higher rated applications have significantly higher impacts on student achievement (i.e., value-added) and higher retention rates, and they conclude that selection due to non-random hiring has little impact on their results.[10]

2.2 Hiring Preferences and Processes in Teaching

Outside of economics, there is a significant literature on teacher hiring. These studies focus mostly on either (a) the characteristics that principals look for in job applicants and (b) the processes used by schools to recruit, screen, and select teachers. Here we give a general sense of these literatures and their findings; for more detailed reviews see Pounder (1989), Rutledge et al. (2008), or Mason and Schroeder (2010).

Studies of the values principals place on teacher applicant characteristics are based on qualitative interviews and surveys, typically with small samples drawn from either one district or a limited geographic area (e.g., Place and Kowalski 1993, Abernathy et al. 2001, Harris et al. 2007, Cannata and Engel 2011). These analyses generally indicate that principals place greater weight on personal traits (e.g., "honesty", "good character", "ability to work with peers", "respect" or "compassion" for students) that may be more difficult to assess than credentials like academic achievement or years of prior teaching experience.

No nationally representative study on methods used for teacher hiring exists. However, a number of studies, spanning many years and various geographic areas, provide a fairly consistent picture. The two methods employed in these studies are either to ask school district

[9] Applications are scored first by central human resources staff and, if the score meets a cutoff, are scored again by school-level personnel; those with high scores in the second stage are brought in for a formal interview by the school principal and/or staff.

[10] They assess the scope of selection bias using instrumental variables methods. Applicant scores are strongly related to hiring probability in their setting, but are sometimes not calculated correctly by district staff who must add up scores on multiple criteria by hand. These arithmetic errors serve as instruments to address the selection problem.

administrators about their hiring practices or to survey teachers about their experiences being hired for their most recent job (Liu and Kardos 2002, Liu and Moore-Johnson 2007).

Applicants almost always submit written applications with information including a resume and proof of certification, as well as transcripts and recommendation letters. From there, a subset of applicants is invited for in-person interview. These surveys also indicate that teachers are usually interviewed more than once as they progress toward being hired.

Submission of writing samples, a portfolio of work, or delivering a sample lesson are all far less common than the in-person interview. None of the early studies we reviewed mentioned any written evaluations other than a cover letter and they all report that a small fraction (typically less than 15 percent) of districts observed applicants teaching a lesson prior to hire.[11] More recently, Strauss (1998) reports that roughly 25% of districts surveyed in Pennsylvania solicit writing samples and roughly one third request teacher exam scores (e.g., National Teacher Examination or Praxis Series). Balter and Duncombe (2005), surveying New York State school districts, report that 60% require a writing sample, 30% require a teaching portfolio, and two thirds require certification exam scores.[12] These surveys also report between 40 and 50 percent of districts using a sample classroom presentation, which may indicate a trend toward greater use. However, surveys of teachers (Liu and Kardos 2002, Liu and Moore-Johnson 2007) across several states typically find only about 15% giving a sample lesson prior to being hired.[13] One

[11] See Neely (1957), Diekrager (1969), Nalley (1971), Hansen (1976), and Luthy (1982).

[12] In addition to letters of recommendation, Pennsylvania districts reported placing high weight on college major and grade point average (but low weight on test scores, essays, or institution attended) when deciding whom to interview. In New York, recommendations and college major are also given high weight in screening prior to the interview, but low weights are given to grade point average, institution attended, and scores on certification exams (or other screening tests).

[13] The fraction of teachers who taught a sample lesson was 6.5 percent in California, 14 percent in Florida, 14.6 percent in Michigan, 19.6 percent in Massachusetts, and 23 percent in New Jersey.

caveat to this conclusion is that student teachers and teachers' aides who are hired to teach full-time, will likely have been observed while teaching, albeit outside of the formal hiring process.[14]

2.3 Validity of Interviews and Job-Task Observations in Employee Selection

A large literature in applied (or "industrial") psychology examines the power of interviews to extract reliable and accurate information about the future success of potential employees. Much of the early research in this field exposed low reliability and validity (see Schmitt 1976), but more recent work demonstrates that structured interviews (i.e., pre-selected questions with rubrics for coding answers) can predict outcome variables such as evaluations of employee performance by supervisors (see Arvey and Campion 1982, Hunter and Hunter 1984, Motowidlo et al. 1990, McDaniel et al. 1994).

A small set of studies focus on interviews for teachers specifically. However, this research typically examines how teacher characteristics (e.g., gender, age) and interview structure (e.g., a single interviewer vs. a panel) affect hiring decisions, and many of these studies use actors instead of actual teachers (e.g., Bolton 1969, Young and Pounder 1986, Young 2005). Few studies test whether interview decisions predict future success in teaching, but there is some evidence, albeit in small samples, of a positive relationship between teachers' interview ratings and supervisor ratings of job performance (Mickler and Solomon 1986) and student achievement gains (Webster 1988).

We know of no study that focuses on the predictive validity of sample lessons done as part of a hiring process.[15] However, the power of job simulations to predict productivity is a

[14] Liu and Moore-Johnson (2007) report that 20 percent of teachers worked in their current schools in some capacity before they were hired, and Strauss (1998) finds that about one third of school districts try to fill full-time teaching positions with current substitutes or part-time teachers.

well-researched issue in the field of industrial psychology (see Wernimont and Campbell 1968, Hunter and Hunter 1984), and there is a large literature on the relationship between student achievement and observed teacher behaviors—typically measured with trained observers using low-inference coding systems or "rubrics." This research has consistently found positive relationships between observed teacher behavior and student learning outcomes.[16] Nevertheless, while research indicates that effective teachers can be identified through observation, this evidence comes from full-time teachers in normal classroom settings and may not accurately reflect the evaluation of sample lessons presented during the hiring process.

3. Application, Hiring, and Performance Evaluation in DCPS

Traditionally, the hiring process for teachers in DCPS was largely decentralized, with school principals making independent decisions on whom to hire with few restrictions aside from licensing and certification. Principals could hire teachers from a variety of sources, including student teachers in their schools, through relationships with local teacher preparation programs, or through central office applications. In 2009, DCPS created TeachDC, a multi-stage, centralized application process which is the focus of our analysis and which we describe in greater detail in Section 3.1. TeachDC aims to streamline hiring by screening out less desirable applicants and giving principals a list of "recommended" candidates who successfully completed the process. We examine hiring and performance for TeachDC applicants from 2011 to 2013.

---

[15] Some indirect evidence is presented by Wede (1996), who analyzed data on subjective performance evaluations of teachers from a school district that incorporated a sample lesson as part of its hiring process. Several years later, average evaluations of those hired during this period were not statistically different than those hired in prior years.

[16] Recent work includes Holtzapple (2003), Schacter and Thum (2004), Milanowski (2004), Kimball et al. (2004), Gallagher (2004), Kane et al. (2011), and Kane et al. (2012). Brophy and Good (1984) review earlier research.

Information on “recommended” TeachDC applicants is made available to principals via an online database. Recommended applicants are listed in alphabetical order, with links to their resumes, and can be filtered by subject area to help principals find candidates. Principals can also navigate through the online database to find out further information on how the applicants scored in the Teach DC process. While we know that DCPS principals are provided with information on the online database during a regularly occurring and mandatory meeting of school administrators, the district does not track whether principals used the database, nor whether they proceeded beyond the list of candidates to view applicants’ scores in any of the hiring stages. As we show below, evidence suggests principals used the list of recommended candidates, but did not rely on the detailed application scores to select applicants.[17]

While completion of TeachDC can help applicants find a job, being on the recommended list is not required in order to be hired by a DCPS principal. TeachDC recommended candidates are quite likely to be hired (see Table 2, discussed below), but they comprise only about one quarter of new teachers hired from 2011 to 2013. An additional quarter of new hires applied to TeachDC but did not complete the process, either because they failed one of the selection stages or stopped voluntarily. Roughly half of news hires during these years did not apply through the TeachDC process.[18]

While we do not examine applicants outside of TeachDC, for comparison purposes we present summary statistics (see Table 1) on all teachers in DCPS, breaking them up by whether

[17] In personal correspondence, DCPS officials indicated their belief that few principals accessed information beyond examining teachers in the recommended pool for the subject in which they were interested in hiring.
[18] This other half of new hires is divided between two groups. First, about 10 percent of new hires applied through other rigorous selection processes like Teach For America and the DC Teaching Fellows, which focus on individuals who lack teaching certification but have outstanding prior achievements and demonstrated leadership in other professions or activities. Over the period 2011-2013, DCTF and TFA brought in, respectively, roughly 100 and 60 new DCPS teachers. These teachers participate in training programs in the summer prior to starting their teaching jobs and take courses to obtain teaching certification during their first few years of employment. Second, the remaining approximately 40 percent of new hires applied directly to local school principals.

the teacher is a new DCPS hire who applied to TeachDC from 2011-2013 (our primary analytic sample), a new hire who did not apply to TeachDC, or a veteran teacher (hired before 2011). Relative to veteran teachers, new hires are younger, more likely to teach in middle schools, and have lower performance evaluation scores. This is true regardless of whether they applied to TeachDC. There are few noticeable differences between new hires who applied to TeachDC and those which did not, although TeachDC applicants appear to have somewhat better performance evaluations in their first year (e.g., -0.4 vs. -0.6 standard deviations on our normalized measure, the details of which are provided in section 3.2).

3.1 The TeachDC Application Process

We focus on the TeachDC selection process as it occurred from 2011-2013. Each year, from roughly February through July, candidates submit applications to Teach DC. The online application system first collects background information such as applicants' education history, employment experience, and eligibility for licensure. Applicants who don't already hold a DC license and whose credentials make them ineligible to obtain one prior to the start of the school year are not allowed to proceed further, and we do not analyze these ineligible applications.[19]

Following collection of this preliminary information, district officials review applications in several stages; we discuss these stages in detail below, as they changed somewhat from year to year. In 2011, there were four stages of evaluation; two written evaluations (general essays and subject-specific assessments), an interview, and a teaching audition. In 2012 and 2013, the general essay was dropped, and applicants were assessed on the remaining three stages. Also, in

[19] To be licensed in DC, teachers must have a bachelor's degree, complete a teacher preparation program (traditional or alternative), and pass both the PRAXIS I and relevant PRAXIS II exams (or substitute exams). Teachers licensed in another state are also generally eligible for a DC license.

2011 many of the teaching auditions were done live in DCPS classrooms, but, due to logistical difficulties, audition videos were used for the remaining 2011 cases and in all of 2012 and 2013.

At the end of each stage, applicants who pass a specified performance threshold are allowed to proceed. Applicants who pass all stages (and a background check) are included in the recommended pool seen online by principals. On average, for those who made it through the process, it took roughly six weeks from the initial application to the pass/fail determination at the final stage.

Table 2 shows the number of applicants evaluated in each recruiting year and each stage, as well as whether or not they passed the stage and the fraction of applicants hired in each possible stage outcome. There were roughly 2,500 applicants per year, of which roughly 13 percent were hired into DCPS. Roughly 60-70% of applicants completed the subject-specific written assessment and 30-40% of applicants completed the interview. However, the number of applicants completing the audition rose significantly after 2011; as we explain in more detail below, Teach DC changed from live auditions in 2011 to video submissions in 2012 and 2013.

As mentioned above, applicants did not have to make it into the Teach DC recommended pool in order to be hired into DCPS.[20] In Table 2, we see that in both 2011 and 2012, the percentage of applicants hired among those not even evaluated in the initial stage was only slightly below average. However, among applicants who are evaluated in each stage, those who failed the evaluation are less likely to be hired than those who passed. Among those applicants who passed the final audition stage and made it into the recommended pool, the fraction hired

[20] Table 1 shows that 28% of the 80 candidates who failed the audition stage in 2011 were nonetheless hired by the district. The analogous figures in 2012 and 2013 are notably lower. Based on conversations with DCPS officials, we believe that some of these candidates were actually moved ahead to the recommended pool to increase the choices available to principals. In analyses we do not present (but which are available upon request), we confirm that all of our results are robust to excluding these 80 applicants or to treating them as recommended.

was 48 percent, 40 percent, and 52 percent in years 2011, 2012, and 2013, respectively. Thus, it seems clear that applicants who make it into the TeachDC recommended pool are far more likely to be hired, supporting the notion that principals use this list as a source for job candidates.

To give a better sense of how the TeachDC process worked in practice, we briefly summarize the key aspects of each stage during the three years on which we focus. In 2011, applicants first submitted online essays of 200-400 words which were scored by one of several district office reviewers for content and writing quality.[21] In addition to the essays used for selection at this stage, applicants were asked additional questions that were not used in the selection process and were not provided to principals that hired new teachers. Importantly, applicants were not told explicitly that these items were different than the essays or any other information that they submitted, so these data are likely indicative of responses that DCPS would receive if they were to be used in the selection process.[22]

Applicants answered 50 multiple-choice questions from the Haberman Star Teacher Pre-Screener (Haberman, 1993), a commercial teacher applicant screening instrument. Used by a number of large urban school districts throughout the U.S., the Haberman Pre-Screener is intended to provide school officials with guidance on how effective a particular candidate is likely to be in an urban classroom. Prior research has indicated a positive relationship between Haberman scores and teacher performance in the classroom (Rockoff et al. 2011).[23]

[21] One essay was on instructional strategies for low-performing students, and the other on the use of student achievement data. These essays were scored by on a 4 point scale (in 0.1 point increments), and a composite score was calculated using weights of 40% for the content of each essay and 20% for overall writing quality. As a general rule, applicants proceeded if they achieved a composite score of 2.0 or higher. In addition, DCPS officials selected a random 20% subset of applicants with scores below 2.0 to pass, although applicants with the minimum possible score (1.0 on both essays) were not eligible to be selected.
[22] Prior to this entire section, applicants were informed that some of the questions were part of a pilot program, but were not told which items were part of the pilot and which were not.
[23] This assessment was developed by interviewing teachers thought to be highly effective and designing questions to capture their attitudes and beliefs. The Haberman Foundation also produces an interview protocol and scoring rubric

In addition, applicants answered multiple-choice questions to measure the "Big Five" personality traits (Costa and McCrae, 1992) and Grit, defined as "the tendency to sustain interest in and effort toward very long-term goals"(Duckworth and Quinn, 2009). [24] While our intention was to examine measures of the Big Five and Grit, a factor analysis (see Appendix Table A1) reveals that applicants' answers are inconsistent with the instruments' designs. The only trait from these surveys that aligns well with a cohesive set of personality questions is Extroversion. All questions other than Extroversion line up along two factors corresponding to whether the question was normally scored (e.g., measuring conscientiousness, the item "Is a reliable worker") or reverse scored (e.g., measuring conscientiousness, the item "Tends to be disorganized"). We believe that this was due to the fact that the questions were asked as part of a job application rather than a low-stakes survey, and candidates may have "faked" their responses to appear more attractive to DCPS officials.[25] Hence, in the analysis below, we include three personality measures (Extroversion, "Positive Spin", and "Negative Spin"), and the Haberman test score in regressions of DCPS hiring and teacher performance.

---

which is intended to assist district officials in identifying individuals likely to be effective urban school teachers, although this protocol was not used in DCPS during the period of our study. The average score (out of 50) for 2011 TeachDC applicants was 34.2, with a standard deviation of 4.7, and similar to the average score of 31.9 (standard deviation 4.8) found by Rockoff et al. (2011) for a sample of recently hired NYC math teachers.

[24] Personality traits were measured using a shortened version of the Big Five Inventory (John, Donahue, and Kentle 1991) in which applicants express their degree of agreement with how a phrase (e.g., "I am talkative") describes them. The 16 items focused mostly on Extroversion (5 questions) and Conscientiousness (5 questions), two traits linked to job performance in earlier studies (Barrick and Mount, 1991; Rockoff et al., 2011), with less emphasis on Agreeableness (2 questions), Neuroticism (2 questions), or Openness to New Experience (2 questions). Grit was measured using a similar instrument developed by Duckworth and Quinn (2009) with eight items, such as "is not discouraged by setbacks" and "has difficulty maintaining focus on projects that take more than a few months." The definition of Grit is provided at: https://sites.sas.upenn.edu/duckworth , accessed on March 17, 2014.

[25] A comparison with responses of roughly 400 recently hired New York City math teachers on a low-stakes survey of the Big Five (Rockoff et al. 2008, Table 2) supports this notion. NYC teachers reported levels (on a 5 point scale) of 4.11, 4.04, and 3.85 for, respectively, Agreeableness, Conscientiousness, and Openness to New Experiences. Each had a standard deviation of about 0.5. In stark contrast, the 2011 TeachDC applicants' average reported Agreeableness, Conscientiousness, and Openness to New Experiences were 4.63, 4.67, and 4.66. For other evidence on self-report bias in this context, see Mueller-Hanson et al. (2003) and Donovan et al. (2014).

Applicants in all three years took a subject-specific written assessment to assess their pedagogical content knowledge (PCK) and knowledge of instructional practices. Applicants selected a subject area (e.g., art, math, Biology) and level (i.e., elementary, middle, or high school) to which they were applying, and then were asked to complete a subject- and level-specific task. Most applicants were asked to read a case-study in which a student demonstrates misunderstanding of the subject matter and to write a 300-400 word essay explaining the nature of the student's misconceptions and describing instructional strategies for resolving them. In 2011 and 2012, applicants for math teaching positions were required to complete the Knowledge of Mathematics for Teaching (KMT) test, a multiple choice test intended to measure understanding and skills distinctly valuable to teaching math (Hill et al. 2004).[26] Essay content and writing quality were scored by DCPS personnel and these scores (plus the KMT test score, when applicable) were averaged to determine whether the applicant passed to the next stage. The passing threshold varied somewhat across years and was altered within the year for certain subject areas in order to obtain enough qualified applicants.

Applicants who passed the subject-specific essay stage were invited for a 30 minute interview and to submit a 10 minute demonstration lesson. Interviews were conducted by the same DCPS personnel who scored the subject-specific essays, as well as several "Teacher Selection Ambassadors" (TSAs). TSAs were highly rated DCPS teachers who received training from DCPS staff in order to assist with the TeachDC selection process.[27]

---

[26] Elementary school applicants wrote an essay assessing content knowledge in English language arts in addition to taking the KMT test. Applicants for middle school math positions in these two years completed the KMT but did not have to complete an additional essay. In 2013, DCPS did not administer the KMT assessment, instead relying on essays alone to evaluate each candidate's content knowledge.

[27] Interviews could be done in person or over the phone, and applicants were asked to respond to a series of structured questions covering five areas: track record of success, response to challenges, contribution to work environment, ownership of high expectations, and continuous learning. For example, under "response to challenges," interviewees were asked, "tell me about the most significant behavior challenge that you've

The demonstration or “mini” lesson could be done in person or submitted by video. Applicants were allowed to choose the topic and had the option to provide lesson materials. DCPS officials scored applicant performance according to selected dimensions of the Teaching and Learning Framework (TLF), the same rubric used to measure classroom performance under the DCPS IMPACT teacher evaluation system, which we describe in more detail below.[28] Applicant performance on the mini-lesson and interview were combined to yield a final score, and applicants scoring above a specified threshold, which varied somewhat across years, were invited to proceed to the final stage. In 2013, DCPS did not require the mini-lesson and applicants were evaluated on the basis of the interview alone.

The final stage in the TeachDC process consisted of a teaching audition in which the applicant taught a complete lesson of approximately 30 minutes. All auditions in 2011 were conducted in DCPS classrooms but were videotaped for evaluation. In 2012, applicants were permitted to submit a videotaped teaching lesson in lieu of the “live” audition, while in 2013 auditions were based completely on video submissions. In each year, DCPS staff and TSAs evaluated the auditions using the same DCPS classroom observation protocol (i.e., the TLF rubric mentioned above), with each audition rated by one TSA.[29]

---

encountered with a student (or group),” with follow-up questions like “what did you do first to address the challenge,” “what was the result,” and “what ultimately happened.” Applicants’ responses were scored on a 4-point scale using a detailed rubric.

[28] Applicants receive a score of 1-4 in five areas: lead well-organized objective-driven lessons, explain content clearly, engage students in learning at all levels, check for student understanding, and maximize instructional time. The scoring rubric is quite detailed and the current version can be found at: http://dcps.dc.gov/page/impact-overview. To provide an example of how scores are anchored, some of the language describing a “4” in “maximize instructional time” includes “routines, procedures, and transitions are orderly, efficient, and systematic with minimal prompting from the teacher.” By contrast, a score of “1” is described by “routines or procedures are not evident or generally ineffective; the teacher heavily directs activities and transitions.”

[29] Applicants received scores from 1-4 on several different elements, with all element scores combined to yield a final score. In 2013, approximately 15% of interviews and 30% of the auditions were checked by a DCPS staff member as part of a “random audit” to assess the reliability of TSA ratings. The correlation between the average scores initially assigned and those after review was 0.87 for interviews, although 45% had at least one component score changed and 17% had the final recommendation overturned. Only 20% of reviewed auditions had any

3.2 Performance Evaluation (DCPS IMPACT)

Each DCPS teacher's performance evaluation for the previous school year is summarized in a single "IMPACT" score, which determines personnel decisions ranging from termination to significant salary increases. An IMPACT score is composed of several performance measures, which vary depending on the grade(s) and subject(s) to which the teacher is assigned. Our data include final IMPACT scores and all component scores (described below) for all district teachers in the school years 2011-12 through 2013-14.

The first component of the IMPACT score is based on measures of student learning. For teachers of math or reading in grades 4 through 10, this component includes an "individual value-added score" (IVA) based on the DC Comprehensive Assessment System (DC-CAS) standardized tests.[30] These teachers, known as "Group 1", represent about 15 percent of DCPS teachers. All teachers are evaluated with a "Teacher Assessed Student Achievement" score (TAS). At the start of the school year each teacher sets student learning goals based on non-DC-CAS assessments which are scored by the teacher, as well as weights if multiple assessments are used. The principal must approve the assessments, weights, and learning goals. At the end of the year, the principal validates the assessment scores and evaluates accomplishment of the learning goals using a rubric.[31] Additionally, in 2011-12 (and earlier years), 5 percent of all teachers' final IMPACT score is a measure of school value-added on DC-CAS tests.

---

component score changed, leading to roughly 10% of reviewed auditions having the final recommendation overturned.

[30] In the 2011-12 school year, the first in our data, "individual value-added" was only measured for grades 4 through 8.

[31] In the 2011-12 school year (and before) IVA was the only student learning component for Group 1 teachers even though these teachers do have TAS scores.

The second component of all teachers' evaluation is a classroom observation score. Each teacher is typically observed five times during the year, three times by a school principal and twice by a "master educator" (i.e., an experienced teacher who conducts observations full-time at many schools). Teachers' performance during classroom observations is scored using the district's own Teaching and Learning Framework (TLF) rubric.[32] Observers assign scores in several areas of practice that are averaged within observations, and then these composites are averaged across observations.[33]

The remaining two evaluation components are assessed by the school principal and assistant principals. Principals rate each teacher's "commitment to the school community" (CSC) using a rubric that covers partnerships with parents, collaboration with colleagues, and support for school-wide initiatives and high expectations. Last, the school principal can deduct points from a teacher's final IMPACT score on the basis of poor attendance, tardiness, disrespect of others, or failure to follow policies and procedures. This last component is known as "core professionalism" (CP).

Teachers' final IMPACT scores are a weighted average of the various component scores; Appendix Table A2 summarizes the weights, which changed between the school years 2011-12 and 2012-13. The final IMPACT score determines the teacher's impact rating category, based on pre-specified ranges. There are five possible ratings: ineffective, minimally effective, developing, effective, and highly effective. However, in 2011-12 there were only four categories; the "developing" category was introduced in 2012-13.

[32] The TLF rubric is modified somewhat for teachers in kindergarten and younger classrooms, and teachers who work with special education or English language learner students in non-traditional settings. During the period of our data, a separate rubric was used for teachers working with students with autism.

[33] Examples of areas of practice include "explains content clearly", "engages students at all learning levels", "provides students multiple ways to move toward mastery", "checks for student understanding", "maximizes instructional time and builds a supportive", and "learning-focused classroom."

Teachers in the “ineffective” category are immediately dismissed. Teachers are also dismissed if they fall in the “minimally effective” category for two consecutive years. At the other end of the distribution, teachers scoring in the “highly effective” category receive a one-time bonus of as much as $25,000. If a teacher is rated highly effective for two consecutive years, she receives a substantial permanent increase in salary; Dee and Wyckoff (2013) estimate this could be worth as much as a 29 percent increase in the present value of a teacher’s total earnings over a 15 year horizon.

3.3 Data and Descriptive Statistics

We use data on over 7,000 individuals who applied through Teach DC in the years 2011-2013 and who were eligible for a teaching license in DC.[34] We analyze subsequent hiring and performance data from the school years 2011-12 through 2013-14.[35] Thus, we have three cohorts of candidates and new hires, and can observe retention and performance for the 2011 applicants for up to three years.

Two limitations in our data are worth noting. The first is that we only have information on job offers for two years. Contrasting the outcomes “offer” and “hire” is useful for disentangling labor supply from labor demand. Among all the TeachDC applicants who received

[34] We drop 198 applicants who participated in a Fast Track application option in 2011. Our results are not sensitive to including these applicants.

[35] We focus on applicants who applied for teaching jobs, and among those applicants identify new hires who are working as a DCPS teacher (as opposed to working as a counselor or administrator or any other role). We define a DCPS teacher as someone who (i) held a teaching position, as recorded in district human resources data and identified by union status, (ii) at some point during the school year. This definition includes individuals who were hired, worked in the fall, but left midyear. It also includes individuals hired midyear. Part-year teachers sometimes do not have job performance data (IMPACT scores), but we nevertheless count them as new hires. Additionally, some DCPS employees who are not officially holding a teaching position do have teaching responsibilities, and are scored in the IMPACT teacher performance evaluation program. In addition to the definition above, we count anyone with IMPACT teaching scores as a DCPS teacher. There are only two such teachers among our applicants, and the results are not sensitive to excluding them.

a formal job offer in 2012, for example, approximately four out of five accepted that offer (82.4 percent). The acceptance rate was slightly higher, 89.3 percent, for applicants who reached the recommended pool. However, based on correspondence with DCPS officials, we believe some applicants received informal offers from the school principal before the principal recorded the offer formally in the district data systems. Some informal offers were likely never recorded in the data systems because they were turned down by the candidate, and thus these acceptance rates are an upper bound.

A second shortcoming is that we cannot observe teacher hiring or performance in DC charter schools. Although they enroll close to half of local students, charter schools are governed by a separate authority, the DC Public Charter School Board. We are also unable to observe if applicants take a job in another school district, such as a Virginia or Maryland district.

Table 3 presents summary statistics for applicants' SAT scores, undergraduate GPA and college selectivity (using a categorical ranking developed from Barron's Profiles of American Colleges (2009)), teaching experience, and other background measures.[36] The first set of columns of Table 3 is based on all applicants, while the second is based on those who are hired.

One third of all applicants have no prior full-time teaching experience, while another third have between one and five years and the remaining third have more than five years. Among those hired, there are noticeably fewer rookie teachers (28 percent). Average self-reported undergraduate GPA (3.40) and composite SAT/ACT scores (1149) are nearly identical regardless of hiring outcome, while college selectivity is slightly higher among hired applicants

[36] Note that we do not have data on the applicant's race or gender; the district is not permitted to require that applicants provide this information.

(2.9 vs. 2.8 on a scale from 1-5).[37] A small minority of applicants attended undergraduate or graduate school in Washington DC (12 percent), but they are overrepresented among those hired (17 percent). About half of applicants report having received a master's degree, in education or any field, or a higher degree.

Our analysis of application measures focuses on three composites drawn from the stages that were common for all cohorts of applicants: (i) a pedagogical content knowledge (PCK) score, (ii) an interview score, and (iii) an audition score. Each of the three is a rescaled composite of the scores collected by DCPS, generated by taking averages or using factor analysis to combine measures when appropriate.[38] All three measures are standardized to have mean zero and standard deviation of one within each year of application.

Table 4 shows the pairwise correlations among applicants' background characteristics and application performance scores. Academic achievement measures such as undergraduate GPA, SAT/ACT score, and college selectivity all have modest positive correlations, as one would expect. Academic achievement is slightly negatively correlated with years of teaching experience, and has small positive correlations with application scores, particularly the subject – specific written assessment.

---

[37] The SAT scores reported by our sample are somewhat higher that the national average, which would have been just above 1000 for cohorts who, like most of our sample, graduated high school in the late 1990s and early 2000s.

[38] To obtain the PCK score we first standardize (mean zero, standard deviation one within years) the subject-specific essay scores on content and writing quality, as well as the KMT score. Our "PCK score" is the average of all standardized scores available for a teacher. For 2011 and 2012 applicants, our "interview score" is the average of two component scores, each standardized: (a) the mean of the applicant's TLF scores for the mini-lesson, and (b) the mean of the applicant's behavioral interview questions scores. For 2013 applicants, we do not have separate scores for the mini-lesson and interview questions, but we have scores on several components (e.g., "instructional expertise," "communication skills") as well as several binary judgments (i.e., "outstanding," "no reservations," "reservations") which we combine using factor analysis to create the 2013 interview score. For each of the three years, a factor analysis on the components of the audition score yields just one factor, and we use the factor analysis weights in each year to construct our audition score. We get virtually identical results if we use a simple unweighted average of the component scores within the audition measure.

Interestingly, while the correlations among application scores themselves are positive, they are all fairly low in magnitude, with the highest correlation between the interview and audition scores (0.22). These correlations suggest the potential for each stage in the application process to be capturing distinct information about teaching applicants, rather than repeatedly measuring the same characteristics and skills. Of course, low correlations also may indicate a considerable amount of noise in each score.

The bottom section of Table 4 shows pairwise correlations for the additional measures collected for the 2011 application cohort. In general, these measures are not at all highly correlated with any of the other application performance measures. There is a modest correlation between extraversion and interview and audition scores (0.14 and 0.13, respectively) and the Haberman score has small positive correlations of roughly 0.2 with the academic achievement measures and the subject-specific (PCK) written assessment.

4. Predicting Hiring with Application Characteristics

To examine the relationship between applicant characteristics and the likelihood of being hired, we estimate a series of linear probability models of the form:

$$(1.1) \quad H_i = \beta \boldsymbol{X_i} + \delta P_i + \varepsilon_i$$

where H_{it} is an indicator for hire into DCPS as a teacher, $\boldsymbol{X_i}$ is a vector of teacher characteristics, and P_i is an indicator for passing to the end of the Teach DC process in order to be placed in the recommended pool. We interact P_i with the year of application allowing δ to differ by year.[39]

[39] In 2011 it appears that the recommended pool was extended to applicants passing the interview stage. In addition to the indicator for passing the final audition stage, we include an indicator for passing the interview for applicants in the 2011 cohort. Thus, the vector P_i includes four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the

The coefficients of interest are contained in the vector β—to what extent do applicant characteristics predict hire into DCPS, controlling for whether the applicant was listed in the recommended pool of candidates (which strongly predicts hiring). We present results with and without the set of indicators P_i in order to examine whether the relationship between hire and a given characteristics is driven by a mechanical relationship with placement on the TeachDC list of recommended candidates.

Because the availability of teaching positions and the supply of candidates may vary widely by subject area and over time, we present results that include fixed effects for the subject area and grade level for which the applicant applied, interacted with the application year. Not all candidates have complete data for all characteristics, and we set missing values to zero and include a set of missing variable indicator flags into the regression. We base our statistical inferences off of heteroskedasticity-robust standard errors. We have also estimated all of the analyses using Logit models, and obtain virtually identical results.

Before presenting these results, it is important to note that DCPS principals have reasonably strong incentives to hire effective teachers. Since the school year 2012-13, principal performance in DCPS has been evaluated under the IMPACT system, parallel to the evaluation for teachers. Multiple criteria, including student test scores and rubric-based evaluations by supervisors, generate an overall performance score for each principal, and low scoring principals are dismissed while high scoring principals can receive substantial bonus payments.

Table 5 presents the results on teacher hiring. We begin by discussing the background characteristics, each of which is first entered separately (Columns 1 and 2) and then

audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011. The left out category is all other applicants.

simultaneously (Columns 3 and 4). The specifications in Columns 2 and 4 include controls reaching the recommended pool. Several robust patterns emerge. Applicants with no prior teaching experience are less likely to be hired by DCPS schools than individuals with prior experience. Depending on the comparison group and specification, rookie applicants have roughly a 3-5 percentage point lower probability of being hired in DCPS. In considering the role of experience in hiring, it is important to note that schools do not bear the financial burden of paying higher salaries to teachers with more experience.

For the most part, teachers with better academic credentials appear to be no more or less likely to be hired into DCPS. Without controls for reaching the TeachDC recommended pool, the coefficients on undergraduate GPA and SAT/ACT score are both close to zero and statistically insignificant, while the coefficient on college selectivity is positive but small (about 1 percentage point for each point on the Barron's scale). However, when we control for reaching the recommended pool, the coefficient on college selectivity goes to zero, and those for undergraduate GPA and SAT/ACT score become negative and significant, though small in magnitude (about 1-1.5 percentage points). The pattern is similar for the coefficients on having a master's degree, but they are never statistically significant.

These results are consistent with two interpretations. First, school principals may not put positive weight on these basic academic achievement measures when making hiring decisions, i.e., a demand side story. The notion that principals do not place positive weight on academic credentials is consistent with prior studies (Ballou 1996, Hinrichs 2014). Second, it may be that principals do place positive weight on these characteristics, but that applicants with better academic backgrounds are less likely to accept job offers from DCPS schools. We cannot definitively separate supply and demand explanations given the limitations in our data on offers.

We do, however, observe many offers that were declined (one-third of offers in 2012 and 2013 did not result in a hire) permitting a partial empirical test. In Appendix Table A3 we present results like Table 5 separately for the outcome "hired" and the outcome "offered", restricted to 2012 and 2013 applicants. The pattern of point estimates is similar regardless of whether one uses offer or hire as the dependent variable. The similarity suggests the results in Table 5 are a demand side story, not applicants' choices about supply.

We now shift focus to the application scores—PCK, interview, and audition—reported in Table 5. Not surprisingly, each of the three application scores is positively associated with the likelihood of being hired when we do not control for reaching the recommended pool (Columns 1 and 3), with coefficients rising monotonically as we move to the later stages of the TeachDC selection process. A one standard deviation increase in applicant score is associated with increases in the likelihood of being hired of 6.0, 10.8, and 15.8 percentage points for the PCK, interview, and audition, respectively.

These effects are quite large, given the baseline hiring rate of roughly 13 percent, but is likely be driven by the effect of arriving into the recommended candidate pool. Indeed, when we include fixed effects for reaching the recommended pool (Columns 2 and 4), the coefficient on the PCK written test goes essentially to zero, while those on the interview and audition drop by about 70 percent. This suggests that principals did not rely heavily on the information collected in the application process beyond the recommendation and that the factors that the principals did rely on were not highly correlated with these scores (conditional on the other factors). In Appendix Table A3 we show that the coefficients on the three application scores are quite similar when the outcome is offer instead of hire, a pattern consistent with a school principal demand interpretation of the results rather than applicant supply decisions.

In Appendix Table A4, we present hiring regressions separately for four groups of subjects: Elementary and Early Childhood, Middle and High School "Core" (i.e. English, math, science, and social studies), Special Education, and Other Subjects.[40] It is perhaps notable that prior experience appears unimportant for hiring in Other Subjects, but generally we find little evidence that any particular group of subjects is driving the results seen in Table 5.

For the 2011 cohort, we can also ask whether hiring is related to our measures of applicant personality and the Haberman teacher screener score. For interpretation, it is important to note that these measures were not made available to principals, because DCPS officials were uncertain about their usefulness, but this fact was not told to applicants in order for the data collection to reflect normal conditions. Extraversion and the Haberman Index are both positively associated with the likelihood of being hired (see Table 6), although the Haberman coefficient becomes small and insignificant once we condition on being in the TeachDC recommended pool.

5. Predicting Performance and Attrition with Application Characteristics

We now restrict our attention to TeachDC applicants that were hired by DCPS, for whom we can observe performance and attrition. Our primary measure of job performance combines the IMPACT components using weights determined by a factor analysis of the sub-scores: classroom observation, individual value-added (if available), the teacher-assessed student achievement (if available), commitment to school community, and core professionalism. This

[40] Other subjects include Health and Physical Education, Music, Art, Drama, Foreign Languages, English as a Second Language, Dual Language, and Career and Technical Education.

consistently yields just one significant "performance factor," which we standardize (mean zero, standard deviation one) within school years.[41]

To examine the relationship between applicant characteristics and teacher performance, we estimate a series of regressions of the form:

$$(1.2)\ E_{it} = \beta \boldsymbol{X_i} + \delta P_i + \Sigma_s \alpha^s D^s{}_{it} + \varepsilon_{it}$$

where E_{it} is the performance evaluation of teacher i in school year t. $D^s{}_{it}$ is a series of binary indicators for subject-year. The other variables are the same as described in equation (1.1). For variables with missing values, we set missing to zero and include a missing variable indicator flag into the regression. We observe each newly hired teacher between one and three times, so our sample is an unbalanced panel; accordingly, we include indicator variables for a teacher's second and third year in DCPS and we report heteroskedasticity-robust standard errors that are clustered by teacher.[42]

Two points are worth noting before we discuss the results. First, we do not take a strong stand on whether application measures have a causal impact on teacher performance. For example, if individuals with unobservable traits positively associated with performance (e.g., a strong work ethic) sort into selective colleges, then we might find a positive coefficient on college selectivity even if the college education itself has no causal effect on performance. Given the primary purpose of teacher selection, we do not view this as a limitation. Second, we are concerned about the possibility that selection into our sample of hired teachers could bias the estimates in equation (1.2) in a way that would confound the inferences we would like to make.

---

[41] Using a standardized version of the official IMPACT score generated by DCPS yields similar results. We prefer the performance factor because the data indicate very similar weights on each component across years, while there were considerable changes across years in weights used by IMPACT (e.g., the TAS component score is completely omitted from the calculation of IMPACT for Group 1 teachers in 2011).

[42] We have estimated models that cluster by school and by teacher and school, and obtain virtual identical results.

For example, Table 5 indicated a positive relationship between prior experience and likelihood of being hired. In this case, we are most concerned that this would lead to a negative bias in the relationship between experience and effectiveness in equation (1.2). In other words, when we estimate the relationship between prior experience and performance using the sample of teachers who were hired, we are concerned that inexperienced teachers who were nonetheless hired may have some unobservable characteristic that is positively associated with performance in the classroom. Given the large number of covariates included in our main specifications, it is difficult to sign selection bias definitively but, as we discuss below in section 5.2, we use several strategies to address potential selection bias.

5.1 Main results for teaching performance

The relationships between applicants' characteristics and scores and their teaching performance (shown in Table 7) are strikingly different than those discussed earlier for the outcome of being hired in DCPS (Table 5). In regressions where background characteristics are examined separately (Column 1), applicants reporting no prior teaching experience were less likely to be hired, but they do not perform significantly worse than those reporting 1-10 years of experience, and they have higher performance evaluations than (the small number of) TeachDC applicants who report more than 10 years of prior experience. Applicants who had tertiary education in Washington DC were also more likely to be hired, but do not have significantly higher performance evaluations. Meanwhile, applicants' academic achievement measures (undergraduate GPA, SAT/ACT scores, college selectivity), which did not predict hiring outcomes, are all significantly positively related to performance, with substantial effect sizes of

0.15 to 0.25 standard deviations. In contrast with prior studies, we do find that teachers with a graduate degree have higher performance scores, at least among newly hired teachers.

We find that the three application scores (PCK written assessment, interview, and audition) are all positive predictors of teacher performance, with effect sizes of roughly 0.3 for PCK and interview and an effect size of 0.17 for the audition when each score is entered in a separate regression (Table 7, Column 1). When all three scores are included simultaneously, the coefficient on the audition becomes smaller (0.12) but it still statistically significant (Column 4).

5.2 Accounting for selection based on hiring

To address potential selection bias due to hiring, we first include fixed effects for being in the recommended pool (Table 7 Column 2). The estimated coefficient on our key background and application score measures are quite robust; the PCK and interview effects decline by less than five percent, while the audition coefficient falls by just over 20 percent. Recall that, in Table 5, the inclusion of these controls eliminated or substantially reduced the relationship between key application measures and the likelihood of being hired. Thus, the considerable variation in characteristics among candidates *within* the recommended and non-recommended pools, and the fact that principals did *not* appear to consider this variation in selecting teachers, ironically provides us with more convincing estimates in Table 7.

As a further test, we add school fixed effects to the performance regressions, so that identification is based purely on comparisons of applicants hired into the same DCPS school. This allows us a further test of the importance of selection bias, as it may be the case that teachers with better application scores (or academic background characteristics) are hired by schools where teachers tend to receive higher evaluation scores. Recall that each school's principal and assistant principals are responsible for determining much of the evaluation score, as described earlier. These results (see Columns 3 and 6 in Table 7) are not noticeably different than those which use both within- and between-school variation for identification.

Finally, we assess the sensitivity of our results to more formal, parametric selection corrections, all of which are variants of the Heckman model. As Heckman and Navarro-Lozano (2004) explain, inclusion of the inverse mills ratio or a more general control function based on the predicted probability of being in the sample will control for selection bias, but only under a potentially restrictive set of functional form assumptions. In order to relax these assumptions, one needs an instrument – i.e., a variable that is associated with the likelihood of being in the sample but does not directly influence the primary outcome. To create an instrument, we takes advantage of the sharp cutoffs for passing through each stage of the TeachDC process in the same way that a regression discontinuity analysis leverages discontinuities in the likelihood of treatment associated with cutoffs in an assignment variable. Appendix Figures 1-3 show the relationship between an applicant's stage score and the likelihood of passing to the next stage in the application process for stages 2-4 respectively. Consistent with the selection process established by DCPS, we see a sharp jump in the likelihood that a candidate will move to the next stage exactly at the point their score passes the threshold.[43]

Table 8 presents the results using these additional selection corrections. The dependent variable is our standardized job performance factor from IMPACT evaluation component scores.

[43] The fact that this jump is not exactly one is due to several factors. First and foremost, in 2011, as a purposeful step in the application redesign process, DCPS randomly selected a small number of applicants who failed stage 1 and passed them to stage 2 in order to study the validity of the stage 1 passing score threshold. This was also done for stage 2 and stage 3. (These applicants were still subject to the same rigorous screening in later stages of the TeachDC process, and in individual school's own selection processes. School principals, who made final hiring decisions, had access to all stage scores. No applicants were randomly selected for the recommended pool or for hire.) More generally however, we see some applicants near the thresholds who pass (fail) despite their failing (passing) scores. We attribute these to misapplication of the threshold or errors in score entry, and therefore use a "fuzzy" regression discontinuity approach.

Columns 1 and 2 simply redisplay the estimates from Table 7 Columns 4 and 5 for convenient comparison. Columns 3 and 4 are estimated just as Column 1 is, except that we add a quadratic function of the predicted probability of hire. We estimate the predicted probability of hire using the specification reported in Table 5 Column 3, but with additional instruments as regressors.

For the estimates in Column 3, we use an extremely sparse set of instruments in the hire equation – namely, four indicator variables: (i) Applicants in any year who scored above the stage 4 cut-score designated by DCPS as the threshold for the recommended pool; (ii) Applicants in 2011 who scored above the stage 3 cut-score; we assume these applicants were also placed in the recommended pool as discussed in the text; (iii) Applicants in 2011 who scored below the stage 2 cut-score but were nevertheless randomly selected to move on to stage 3; (iv) Applicants in 2011 who applied in the first weeks of the recruitment season, all of whom were allowed to move on to stage 3 regardless of their scores in stage 2 or 1. For the most part, the results in Column 3 are nearly identical to those in Column 2. One exception is the audition score, where the estimate using the control function is about 30 percent lower than the specification that merely controls for recommended-pool by year fixed effects. The point estimates on the interview and PCK measures drop somewhat as well, but both remain statistically significant and substantively important.

For the estimates in Column 4, we include a much more comprehensive set of instruments that are intended to replicate the RD intuition illustrated in Appendix Figures 1-3. Specifically, the added instruments include five indicator variables: (i)-(iii) Applicants in any year who scored above the cut-score in stage 2, 3, and 4 respectively. And again for 2011 (iv) applicants randomly selected to advance or (v) early applicants automatically advanced. We also allow the slope on each stage score to be different above and below the stage cut-score, and include fixed effects for the highest stage an applicant was invited to complete. All these added coefficients are allowed to vary by year. The results are virtually identical to those in Column 3.

In the previous discussion of hiring we offered evidence consistent with a demand side explanation—school principals' choices—rather than a supply explanation. To continue that line of analysis here we repeat our analysis using the outcome "offer" instead of "hire" in the first stage. Our goal is a test of whether supply-side selection explains the relationships between application data and on-the-job performance. The results, using data from 2012 and 2013, are provided in Appendix Table A5. In short, the pattern of results is again quite similar to the pattern in Table 8, consistent with a demand side story.

Together these specification tests suggest that selection bias is not an important factor in our context. In all remaining specifications we follow the approach taken in Columns 5 and 6 of Table 7 and include recommended pool by year fixed effects.

5.3 Heterogeneity and Sensitivity

We explore heterogeneity in our performance predictions across types of schools using two sample splits: (1) elementary vs. middle and high schools and (2) below and above median student poverty. We find some evidence that prior experience and applicants' audition scores are stronger predictors of performance in middle and high schools and in higher poverty schools (Appendix Table A6), but broadly speaking our estimates are fairly stable.

We also examine the component parts of our overall performance measure to check if our results vary significantly across the particular components of the IMPACT evaluation (see Appendix Table A7). In general, we find consistent effects across all of the non-IVA component measures—we examine teachers with IVA separately in order to isolate differences in outcomes from differences in sample. We see that prior teaching experience appears to predict TLF observation scores, but not the other performance measures.

When we turn to teachers with IVA estimates, our sample size falls by roughly 80 percent. These results are therefore far less precise and should be taken with a great deal of caution. Nevertheless, some interesting patterns emerge. First, the application scores are not significant predictors of IVA (see Appendix Table A8). While we cannot rule out meaningful effects, the evidence on the long-term effects of high value-added teachers is strong, whereas we know little about the ultimate effects of teachers do better on the other performance measures such as TLF observation scores. Correlations for DCPS teachers between value-added and each of the other components of IMPACT are modest, about 0.15 to 0.30, but all positive and significant. Thus, one interpretation is that the value-added measures are sufficiently noisy that we cannot detect small positive effects in our small sample, but an equally plausible interpretation is that the application measures are not good predictors of teachers' impacts on high stakes standardized tests. We hope to address this issue in the future by incorporating data for more cohorts of applicants and years of performance.

Second, prior teaching experience, a graduate degree, and college selectivity significantly predict IVA scores (see Appendix Table A8). These measures are also predictive of classroom observation scores for this sample. However, since we have not assessed the robustness of these findings to correct inference for multiple hypotheses, we regard them as suggestive and in need of further investigation.

Finally we examine the additional measures available for the 2011 cohort: self-reported personality traits and the Haberman test score. The most interesting result to emerge is that coefficient on the Haberman score is large, positive, and significantly associated with teacher performance (Table 9). Specifically, a one standard deviation increase in an applicant's score on the Haberman Index is associated with a 0.27 standard deviation increase in measured effectiveness, even after controlling for reaching the recommended pool.

5.4 Results for attrition

Hiring an effective teacher will be more beneficial when this individual stays employed in the school or district for a significant period of time. We therefore examine attrition from DCPS and from an individual school as additional outcomes of interest. Because we can only observe one year of attrition for the 2013 cohort, we focus on attrition after the first year and pool the three cohorts together. As above, we separately analyze background characteristics and application scores.

We find that self-reported prior experience and academic credentials are not significant predictors of attrition in our basic specification (Columns 1 and 3 of Table 10), while having attended undergraduate or graduate school in Washington DC is strongly negatively related to attrition.[44] Because teachers who perform poorly under the IMPACT evaluation system are forced to leave DCPS, we also estimate specifications that include indicators for the teacher's IMPACT performance level (Columns 2 and 4) so as to rule out any mechanical effects driven by correlations with performance. This change in specification slightly attenuates the effect of attending school in DC, but also raises the coefficients on academic achievement and leads them to become statistically significant in some cases. Teachers with higher application scores, in

[44] Given these results on attrition, we check the sensitivity of our performance regression results to omitting controls for teachers' second and third year in DCPS, since IMPACT evaluations do improve with experience. The coefficient for applicants from the DC area increases very slightly, by about 0.01 standard deviations, and remains statistically insignificant.

contrast, are never found to be more likely to leave after the first year. Indeed, the coefficients on PCK, interview, and audition are almost all negative.

6. Discussion and Conclusions

We study the relationship among applicant characteristics, hiring outcomes, and teacher performance in Washington DC Public Schools (DCPS). We find that several background characteristics (e.g., undergraduate GPA) as well as screening measures (e.g., applicant performance on a mock teaching lesson) strongly predict teacher effectiveness. Interestingly, we find that these measures are only weakly, if at all, associated with the likelihood of being hired.

Our results suggests that there exists considerable scope for improving teacher quality through the selection process. To summarize that scope graphically we plot the distributions of predicted first-year performance separately for applicants hired and not hired in Figure 1. We estimate predicted first-year performance for each applicant based on all background characteristics and application scores, applying coefficients estimated with a specification identical to Table 7 Column 4 but limiting the sample to new hires in their first year at DCPS.[45] In this estimation we use a leave-one-out procedure so that the outcome for an individual teacher does not influence his or her own predicted score.[46] As Figure 1 shows, applicants who are hired by DCPS do have higher average predicted performance. Still, there is substantial overlap in the distributions; there are many applicants who are not hired but whose predicted performance exceeds the average of those hired.

[45] As in Table 7 Column 4 the specification includes subject-taught by year fixed effects. We do not include any between subject or year variation in our predicted performance measure. Practically, we do not include the fixed effects coefficients in the prediction.

[46] Specifically, to obtain the predicted value for teacher *i*, we estimate our model using all observations except for those from teacher *i*. Using the coefficients from this regression and teacher *i*'s Xs, we calculate the predicted value for teacher *i*.

To compliment Figure 1, we summarize how the combined application measures predict the probability of being hired in Figure 2. Figure 2 plots the actual proportion hired separately for the 20 vingtiles of predicted first-year performance (as constructed in Figure 1). The probability of being hired is roughly 10 percent for applicants in the bottom third of the predicted performance distribution, and only increases slightly through the next third. In the top third of the predicted performance distribution do we see a sharp increase in the proportion hired, but even among the top 5 percent of applicants only 30 percent ended up working in the DCPS. Assuming that this is at least partly driven by demand on the part of DCPS schools, this suggests considerable scope for improving teacher quality through the hiring process.

To explore the relationship between the application measures and teacher performance in the classroom, Figure 3 presents box plots of actual performance for each vingtile of predicted first-year performance. This is the same predicted performance measure in Figures 1 and 2, except Figure 3 uses only observations on new hires. As suggested by the earlier regression results, we see a positive relationship between predicted and actual performance. Interestingly, it appears that the relationship is somewhat steeper at the top and bottom of the predicted performance distribution. In order to gain a better sense of the magnitude of these performance differences, Figure 4 plots kernel densities of actual performance separately by quartile of predicted performance. Teachers in the top quartile of predicted effectiveness score roughly two-thirds of a standard deviation higher in actual effectiveness than their peer applicants who scored in the bottom quartile. This illustrates that the predictions captured by the application measures incorporate considerable information regarding actual effectiveness.

Collecting and scoring detailed information about applicants does create additional costs, but those costs appear to be small relative to the potential gains from improved selection. The primary marginal cost is the labor of the DCPS school administrators who conduct and score the interviews and teaching auditions. Each administrator spends about one hour per interview and one hour per audition. DCPS budgets $63K per year for interviews and auditions at $34 per hour of administrator work. The screening steps leading up to the interviews and auditions is much less costly at $7,500 per year total. In addition, DCPS budgets $133K per year for the staff who manage recruitment and screening. Some of that management cost would be required even without the additional screening measures, so the marginal cost is something less than $133K. Altogether, approximate total marginal costs are between $70-200K per year, or between $370-1,070 per new hire.[47] This cost is likely to be quite small relative to the anticipated long-run benefits to future students of hiring more effective teachers (Chetty et al. 2014a,b, Jackson 2013, 2014), and small relative to the costs of proposals for screening after hiring—removing low-performing teachers (Staiger and Rockoff 2010, Rothstein 2015).

"Hire the right employees" is an intuitive goal for managers in all sectors, including school principals. How to go about improving selection in hiring is much less clear. This paper documents one example where atypically-detailed assessments of applicants' backgrounds and skills can improve hiring decisions. This new empirical evidence is an important contribution to the small but growing economics literature on employee selection and, for the literature on teachers in particular, is an encouraging contrast to many failed attempts to explain differences in teacher performance. Yet the fact that the applicant assessments weak predictors of hiring decisions suggest that the schools we study do not fully capture the potential gains from teacher selection. Providing useful information to decision makers, such as school principals, may

[47] This cost per new hire divides the total cost by 190 new hires—the average annual number of new hires during the study period who completed all stages of the TeachDC screening process. The per hire cost would be lower if we count new hires who completed only part of the TeachDC process, but, as discussed earlier, it is difficult to know exactly how the TeachDC measures influenced those hires.

therefore be an important part of the hiring, and we plan to study the impact of such information in future work.

References

Abernathy, Tammy V., Al Forsyth and Judith Mitchell. 2001. “The Bridge from Student to Teacher: What Principals, Teacher Education Faculty, and Students Value in a Teaching Applicant,” *Teacher Education Quarterly*, 28(4): 109-119.

Adnot, Melinda, Thomas Dee, Veronica Katz, and James Wyckoff. 2016 “Teacher Turnover, Teacher Quality, and Student Achievement in DCPS.” NBER Working Paper No. 21922.

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.

Arvey, Richard D., and James E. Campion. 1982. “The Employment Interview: A Summary and Review of Recent Research.” Personnel Psychology 35(2): 281-322.

Baker, A. and Santora, M. (2013, January 18). “No Deal on Teacher Evaluations; City Risks Losing $450 Million.” The New York Times, p. A1.

Ballou, D. (1996) “Do Public Schools Hire the Best Applicants?” *Quarterly Journal of Economics*, 111(1): 97-133.

Balter, Dana and William D. Duncombe. 2005. “Teacher Hiring Practices in New York State Districts,” Report prepared for the Education Finance Research Consortium.

Barrick, M. R. and Mount, M. K. (1991) “The Big Five Personality Dimensions and Job Performance: A meta-analysis,” Personnel Psychology, 44(1), 1-26.

Barron’s Profiles of American Colleges, 28th Edition. (2009) Hauppauge, NY: Barron’s Educational Series.

Bolton, D.L. 1969. “The Effect of Various Information Formats on Teacher Selection Decisions,” *American Educational Research Journal*, 6(3): 329-347

Boyd, Donald, et al. "The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools." *Journal of Policy Analysis and Management* 27.4 (2008): 793-818.

Boyd, Don, Hamp Lankford, Susanna Loeb, Matthew Ronfeldt, and Jim Wyckoff (2011), “The Role of Teacher Quality in Retention and Hiring: Using Applications to Transfer to Uncover Preferences of Teachers and Schools,” *Journal of Policy Analysis and Management* 30:1, 88-110.

Brophy, Jere, and Thomas L. Good. 1986. Teacher Behavior and Student Achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching*. 3rd ed., 238-375, New York: Simon and Schuster.

Burks, S. V., Cowgill, B., Hoffman, M., & Housman, M. 2015). “The value of hiring through employee referrals.” *Quarterly Journal of Economics* 130(2): 805-839.

Chetty, Raj, John N. Friedman & Jonah E. Rockoff, 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9), pages 2593-2632.

Chetty, Raj, John N. Friedman & Jonah E. Rockoff, 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104(9), pages 2633-79

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." Journal of Public Economics 92(5): 1352-1370.

Costa, Paul T., and Robert R. McCrae. "Neo PI-R professional manual." (1992): 653-65.

Dee, Thomas, and James Wyckoff. Incentives, selection, and teacher performance: Evidence from IMPACT. No. w19529. National Bureau of Economic Research, 2013.

Dobbie W. 2011. Teacher Characteristics and Student Achievement: Evidence from Teach for America. Harvard University Working Paper.

Dolton, Peter and Wilbert van der Klaauw. 1995. "Leaving Teaching in the UK: A Duration Analysis," *The Economic Journal* 105(429):431-444.

Dolton, Peter and Wilbert van der Klaauw. 1999. "The Turnover of Teachers: A Competing Risks Explanation," *The Review of Economics and Statistics*, 81(3): 543-550.

Donovan, JJ, Dwight S.A., Schneider, D. (2014) "The Impact of Applicant Faking on Selection Measures, Hiring Decisions, and Employee Performance," *Journal of Business Psychology* 29:479–493.

Duckworth, Angela Lee, and Patrick D. Quinn. "Development and validation of the Short Grit Scale (GRIT–S)." *Journal of personality assessment* 91.2 (2009): 166-174.

Gallagher, H. Alix. 2004. "Vaughn Elementary's Innovative Teacher Evaluation System: Are Teacher Evaluation Scores Related to Growth in Student Achievement?" *Peabody Journal of Education*, 79(4): 79-107.

Goldhaber, D., Grout, C., and Huntington-Klein, N. (2014). Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools. CEDR Working Paper 2014-9. University of Washington, Seattle, WA.

Goldhaber and Theobald (2013) "Managing the Teacher Workforce in Austere Times: The Determinants and Implications of Teacher Layoffs," *Education Finance and Policy* 8(4):494-527

Gordon, R., Kane, T.J., Staiger, D.O. (2006) "Identifying Effective Teachers Using Performance on the Job" Hamilton Project Discussion Paper 2006-01.

Haberman, M. (1993). Predicting the Success of Urban Teachers (The Milwaukee Trials). Action in Teacher Education, 15(3), pp.1-5.

Hanushek, E.A. (2011) "The Economic Value of Higher Teacher Quality" *Economics of Education Review* 30(3):466-479

Hanushek Eric A., John F. Kain, and Steven G. Rivkin (2004) "Why Public Schools Lose Teachers," *Journal of Human Resources* 39(2): 326-354.

Hansen, Cecil Ray, 1976. Practices and Procedures Used by Selected Utah Public School Districts in the Recruitment and Selection of Teachers. Doctoral Dissertation, Brigham Young University.

Harris, D.N and Sass, T.R. (2011) "Teacher training, teacher quality and student achievement," *Journal of Public Economics* 95(7–8):798-812.

Heckman, James and Salvador Navarro-Lozano (2004). "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics*, 86(1): 30-57.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105, 11–30.

Hinrichs, Peter 2014. "What Kind of Teachers Are Schools Looking For? Evidence from a Randomized Field Experiment," Federal Reserve Bank of Cleveland Working Paper 14-36.

Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li. 2015 "Discretion in Hiring." NBER Working Paper 21709.

Holtzapple, Elizabeth. 2003. "Criterion-related validity evidence for a standards-based teacher evaluation system." *Journal of Personnel Evaluation in Education* 17(3): 207-219.

Hunter, John E., and Ronda F. Hunter. 1984 "Validity and utility of alternative predictors of job performance." *Psychological Bulletin* 96(1): 72-98.

John, O.P., Donahue, E.M., and Kentle, R. L. (1991). The "Big Five" Inventory—Versions 4a and 54. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.

Jackson, C.K. (2013) "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina," NBER Working Paper No. 18624.

Jackson, C.K. (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. 2009. "Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation." *Journal of Labor Economics* 27(2): 213-256.

Jacob, Brian A. (2011). "Do Principals Fire the Worst Teachers?" *Educational Evaluation and Policy Analysis*. 33(4): 403-434.

Jacob, Brian A. (2013). "The Effect of Employment Protection on Worker Effort: Evidence from Public Schooling." *Journal of Labor Economics*. 31(4): 727-761.

Jacob, Brian and Elias Walsh (2011). "What's in a Rating?" *Economics of Education Review*. 30(3): 434-448.

Jacobson, Stephen L. 1989. "The effects of pay incentives on teacher absenteeism." *Journal of Human Resources* 24(2): 280-286.

Kane, Thomas J., and Douglas O. Staiger. 2005. "Using Imperfect Information to Identify Effective Teachers." Unpublished manuscript, April 2005.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27: 615-631

Kane, Thomas J., Douglas O. Staiger, and Dan McCaffrey. 2012. "Gathering Feedback for Teaching," Bill and Melinda Gates Foundation Research Paper. Accessed in May 2015 at http://www.metproject.org/reports.php.

Kane, Thomas J., Eric S. Taylor, John H. Tyler, and Amy L. Wooten. 2011. "Identifying effective classroom practices using student achievement data." *Journal of Human Resources* 46(3): 587-613.

Kimball, Steven M., Brad White, Anthony T. Milanowski, and Geoffrey Borman. 2004. "Examining the Relationship Between Teacher Evaluation and Student Assessment Results in Washoe County." *Peabody Journal of Education*, 79(4): 54-78.

Liu, E. and Kardos, S.M. (2002) "Hiring and professional culture in New Jersey Schools," Cambridge: Project on the Next Generation of Teachers at the Harvard Graduate School of Education.

Liu, E. and Moore Johnson, S. (2006) "New teachers' experiences of hiring: Late, rushed, and information poor," *Educational Administration Quarterly*, 42(3): 324-360.

Luthy TF (1982) Policies and procedures for selection of Missouri school teachers. Doctoral Dissertation, University of Missouri, Columbia.

Mason, Richard W. and Mark P. Schroeder. 2010. "Principal Hiring Practices: Toward a Reduction of Uncertainty," *The Clearing House*, 83(5): 186–193.

McDaniel, Michael A., Deborah L. Whetzel, Frank L. Schmidt, and Steven D. Maurer. 1994. "The validity of employment interviews: A comprehensive review and meta-analysis." *Journal of Applied Psychology* 79(4): 599-616.

Meriam, Junius Lathrop. 1906. "Normal School Education and Efficiency in Teaching," Columbia University Contributions to Education No. 1.

Milanowski, Anthony. 2004. "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati." *Peabody Journal of Education*, 79(4): 33-53.

Morsh, Joseph E. and Eleanor W. Wilder (1954) "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952" Air Force Personnel and Training Research Center, Research Bulletin 54-44.

Motowidlo, Stephen J., Marvin D. Dunnette, and Gary W. Carter. 1990. "An alternative selection procedure: The low-fidelity simulation." *Journal of Applied Psychology* 75(6): 640-647.

Mueller-Hanson, Rose, Eric D. Heggestad, and George C. Thornton III. 2003. "Faking and Selection: Considering the Use of Personality from Select-In and Select-Out Perspectives," *Journal of Applied Psychology*, 88(2): 348–355

Nalley, B. J. (1971) A descriptive survey of the recruitment and selection process of teachers for the District of Columbia and a comparison of procedures used in selected school systems of comparable size. Doctoral dissertation, George Washington University, Dissertation Abstracts International, 32, 3626A.

Neely, Melvin E. (1957) “A Survey of Present Procedures in the Selection of Teacher Personnel” Doctoral Dissertation, University of Kansas.

Ost, Ben. 2014. "How Do Teachers Improve? The Relative Importance of Specific and General Human Capital." *American Economic Journal: Applied Economics*, 6(2): 127-51.

Oyer, Paul, and Scott Schaefer. 2011. “Personnel economics: hiring and incentives.” Handbook of Labor Economics 4: 1769-1823.

Papay, John P., Matthew A. Kraft. 2015. “Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement,” *Journal of Public Economics*, 130: 105–119

Place, Andrew W. and Theodore J. Kowalski. 1993. “Principal Ratings of Criteria Associated with Teacher Selection,” *Journal of Personnel Evaluation in Education* 7(): 291-300.

Pounder, Diana G. 1989. “Improving the Predictive Validity of Teacher Selection Decisions: Lessons from Teacher Appraisal,” *Journal of Personnel Evaluation in Education* 2(2):141-150.

Rice, Jennifer K. 2013. “Learning from experience? Evidence on the impact and distribution of teacher experience and the implications for teacher policy,” *Education Finance and Policy*, 8(3): 332–348.

Rivkin, Steven. G., Eric. A. Hanushek, and John F. Kain. 2005. "Teachers, Schools and Academic Achievement." *Econometrica* 73: 417-458.

Rockoff J.E. (2004) “The impact of Individual Teachers on Student Achievement: Evidence from Panel Data,” *American Economic Review Papers and Proceedings* 94(2):247-252

Rockoff J.E., Jacob B.J., Kane T.J., Staiger D.O. (2011) Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*. 6(1):43-74.

Rockoff J.E., Speroni C. 2010. Subjective and Objective Evaluations of Teacher Effectiveness. *American Economic Review* 100(2): 261–66

Rockoff, Jonah E., Douglas O. Staiger, Eric Taylor, and Thomas J. Kane. 2012. “Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools.” American Economic Review 102(7): 3184-3213.

Rothstein J. (2015) “Teacher Quality Policy When Supply Matters,” *American Economic Review* 105(1):100-130.

Rutledge, Stacey A., Douglas N. Harris, Cynthia T. Thompson, W. Kyle Ingle. 2008. “Certify, Blink, Hire: An Examination of the Process and Tools of Teacher Screening and Selection,” *Leadership and Policy in Schools*, 7(3): 237–263.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner. 2007. "Race, poverty, and teacher mobility." *Economics of Education Review* 26(2): 145-159.

Schacter, John, and Yeow M. Thum. 2004. “Paying for High- and Low- Quality Teaching.” *Economics of Education Review*, 23(4): 411-440.

Schmitt, Neal. 1976. “Social and Situational Determinants of Interview Decisions: Implications for the Employment Interview.” *Personnel Psychology* 29(1): 79-101.

Schmutte, Ian M. 2015. “Job Referral Networks and the Determination of Earnings in Local Labor Markets” *Journal of Labor Economics*, 33(1)

Staiger DO, Rockoff JE. 2010. Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives* 24: 97-117

Stanton, Christopher, and Catherine Thomas. (2013) “Landing the first job: the value of intermediaries in online hiring.” Unpublished manuscript.

Strauss, R. 1998. “Teacher Preparation and Selection in Pennsylvania,” Research Report to the Pennsylvania State Board of Education.

Wede, Richard J. 1996 “Teacher Selection: Use of Demonstration Lessons” Doctoral Dissertation, Drake University.

Young, I.P., Pounder, D.G. 1986. “Salient Factors Affecting Decision Making in Simulated Teacher Selection Interviews,” Journal of Educational Equity and Leadership, 5(3):216-233

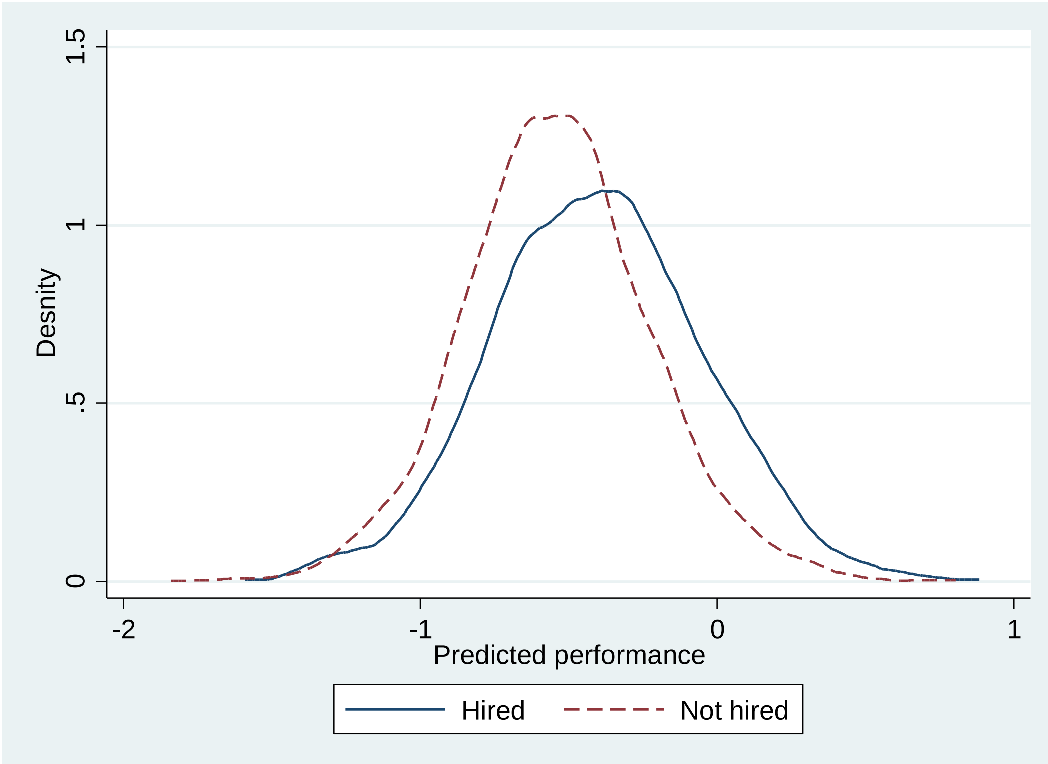


Figure 1—Predicted performance for applicants hired and not hired

Note: Kernel densities estimated separately for applicants hired and not hired of predicted performance in the first year. Predicted performance is estimated as follows: First, using the sample of new hires in their first year at DCPS, fit a regression similar to Tables 7 Column 4. The dependent variable is the first predicted factor from a factor analysis of IMPACT evaluation component scores from a teacher's first year at DCPS. The covariates include all of the background characteristics and application scores. The specification also includes subject-taught by year fixed effects. Second, the estimated coefficients from that regression are applied to the applicant sample. This predicted performance measure does not include differences between the subject-taught by year fixed effect groups.

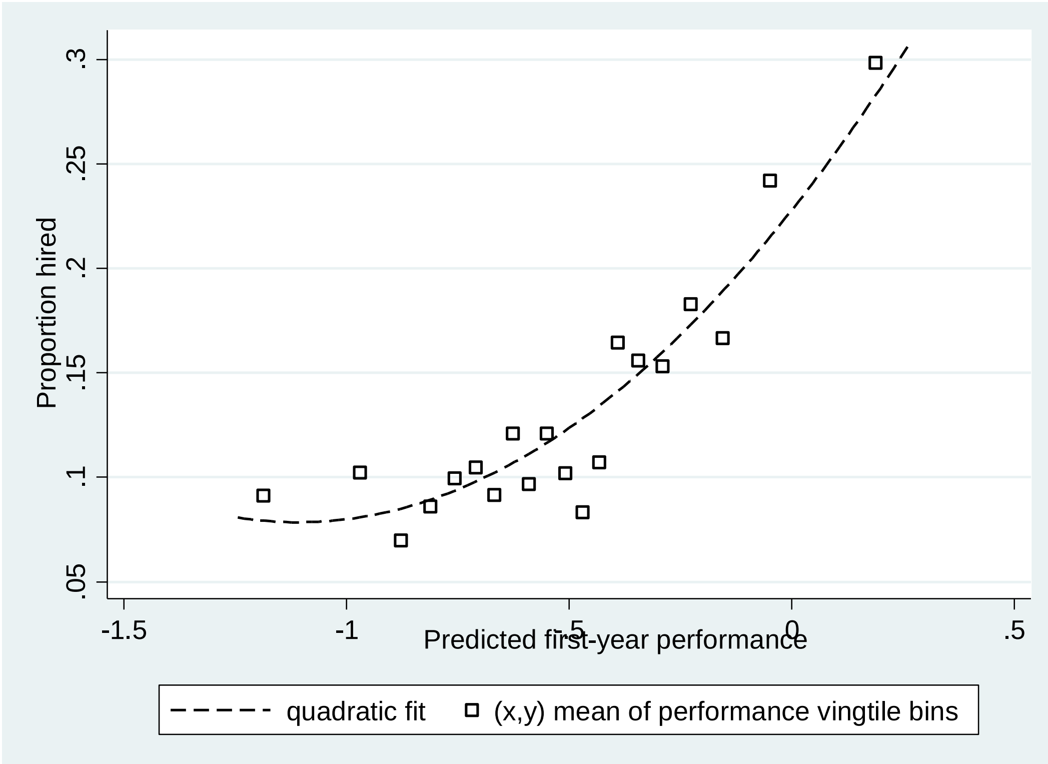


Figure 2—Relationship between predicted performance and hiring

Note: The dashed line represents results of a regression with 7,442 observations (all applicants). In each case the binary outcome being hired by DCPS is regressed on predicted first-year job performance, and year-by-subject-applied fixed effects. Each square marks the (x,y) mean for 20 bins. Each bin is a vingtile of predicted performance. The x-axis is limited to the 2nd through 98th percentiles of predicted performance.

Predicted performance is estimated as follows: First, using the sample of new hires in their first year at DCPS, fit a regression similar to Tables 7 Column 4. The dependent variable is the first predicted factor from a factor analysis of IMPACT evaluation component scores from a teacher's first year at DCPS. The covariates include all of the background characteristics and application scores. The specification also includes subject-taught by year fixed effects. Second, the estimated coefficients from that regression are applied to the applicant sample. This predicted performance measure does not include differences between the subject-taught by year fixed effect groups.

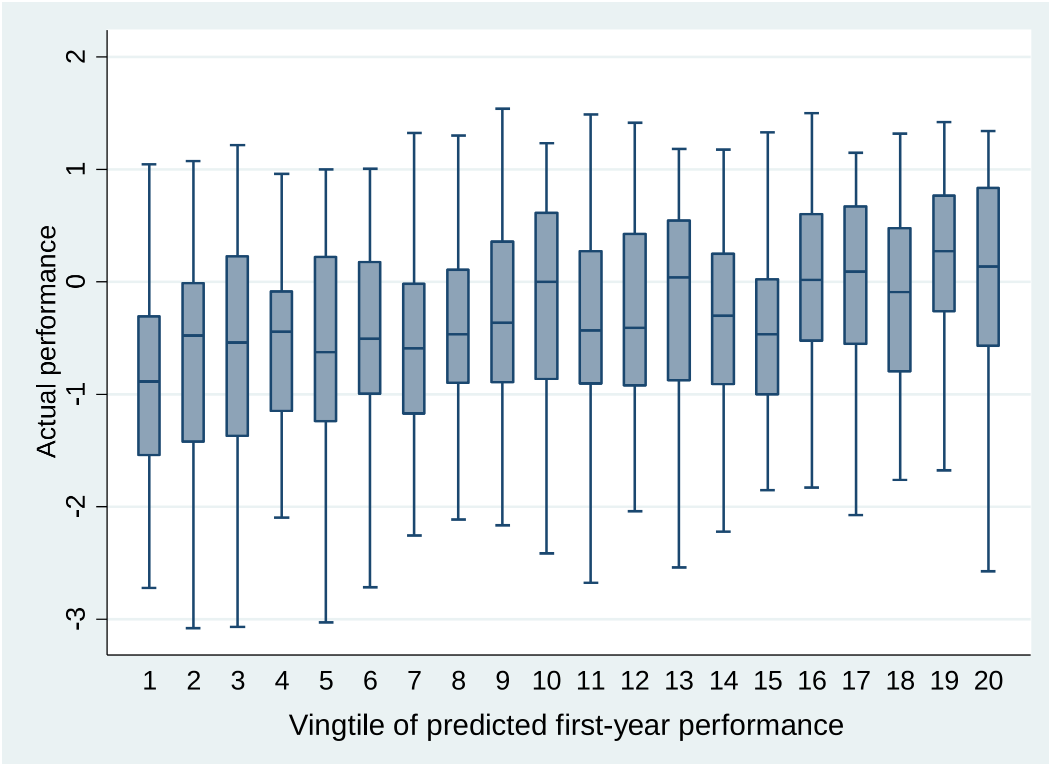


Figure 3—Relationship between predicted performance and actual performance

Note: Box-plots of actual performance for each vingtile of predicted performance. Actual performance is the first predicted factor from a factor analysis of IMPACT evaluation component scores. Predicted performance is the fitted value obtained after the following regression: Using the sample of new hires, fit a regression similar to Table 7 Column 4. The dependent variable is the first predicted factor of IMPACT scores. The covariates include all of the background characteristics and application scores. The specification also includes subject-taught by year fixed effects.

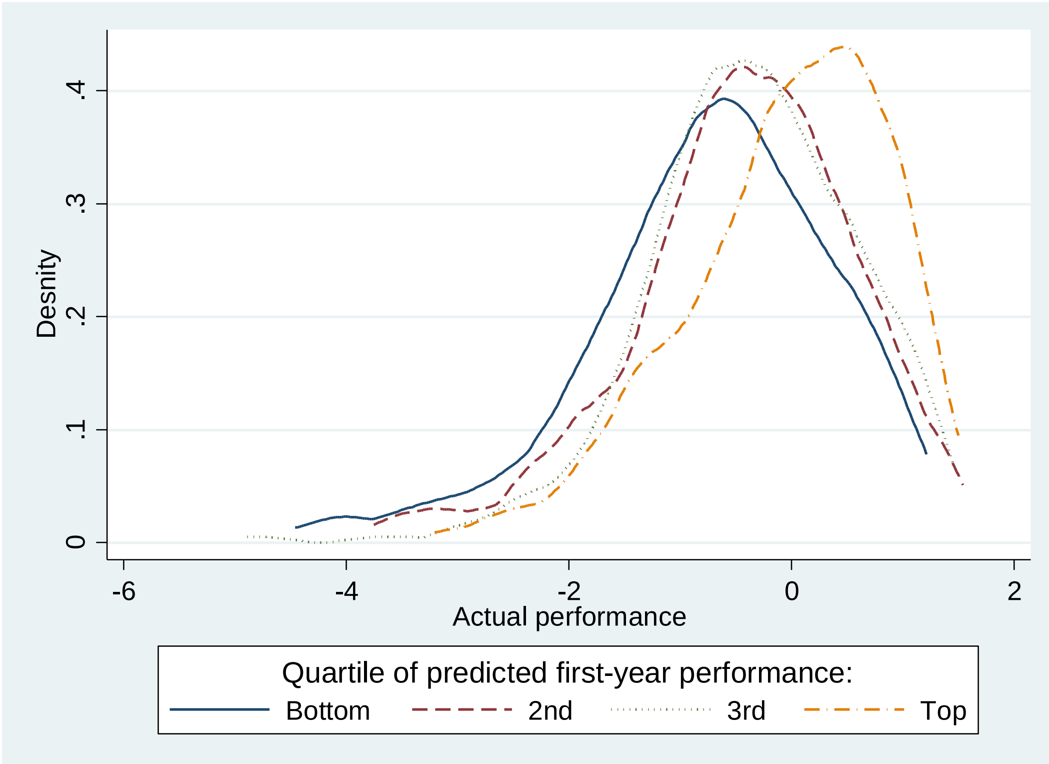


Figure 4—Relationship between screening measures and performance

Note: Kernel densities estimated separately by quartile of predicted performance using teacher-by-year observations. Predicted performance is the fitted value obtained after the following regression: Using the sample of new hires, fit a regression similar to Table 7 Column 4. The dependent variable is the first predicted factor of IMPACT scores. The covariates include all of the background characteristics and application scores. The specification also includes subject-taught by year fixed effects.

Table 1—DCPS teacher characteristics

	Hired before 2011		New hires, first year on the job			
	First year in data		Non TeachDC		TeachDC	
	Obs.	Mean (st.dev.)	Obs.	Mean (st.dev.)	Obs.	Mean (st.dev.)
Female	2,920	0.76	842	0.75	927	0.75
Race/ethnicity	2,704		380		823	
Black		0.60		0.39		0.44
White		0.32		0.42		0.47
Hispanic		0.04		0.11		0.05
Asian		0.04		0.08		0.01
Other		0.01		0.00		0.03
Age	2,914	42.32	820	29.97	900	31.41
School type	2,917		839		926	
Education center		0.17		0.19		0.18
Elementary school		0.46		0.38		0.44
Middle school		0.09		0.17		0.15
High school		0.25		0.23		0.20
Other		0.03		0.03		0.03
Final IMPACT score	2,920	315.04	842	290.50	930	297.95
		(45.36)		(48.51)		(46.87)
Not in DCPS next year	2,920	0.19	842	0.21	930	0.22
Not in same school next year	2,920	0.26	842	0.29	930	0.31

Note: Authors' calculations. Sample restricted to DCPS teachers with IMPACT scores. Calculations based on one observation per teacher, the first year they appear in the data.

Table 2—Applicant progress through TeachDC process

		2011 applicants		2012 applicants		2013 applicants	
		#	Fraction hired	#	Fraction hired	#	Fraction hired
Eligible but Initial Stage Incomplete:		174	0.13	787	0.10	1,041	0.02
General Essay	Failed this Stage:	228	0.04				
	Incomplete Next Stage:	362	0.09				
Content Knowledge	Failed this Stage:	530	0.06	622	0.05	260	0.02
	Incomplete Next Stage:	314	0.09	304	0.09	150	0.06
Interview + Sample Lesson	Failed this Stage:	239	0.09	283	0.05	184	0.03
	Incomplete Next Stage:	269	0.32	39	0.18	302	0.09
Teaching Audition	Failed this Stage:	80	0.31	100	0.01	156	0.04
	Passed Stage:	164	0.48	392	0.42	462	0.52

Note: Authors' calculations. "Stage reached" is the highest stage in which the data include a score or pass/fail determination.

Table 3—Characteristics of TeachDC applicants, 2011-2013

	All applicants		Applicants hired	
	Obs.	Mean (st.dev.)	Obs.	Mean (st.dev.)
Hired	7,442	0.13	982	1
Prior teaching experience	7,314		978	
Novice		0.33		0.28
1 to 2		0.17		0.19
3 to 5		0.18		0.20
6 to 10		0.17		0.20
11 or more		0.14		0.14
Undergraduate GPA	7,112	3.40	939	3.42
		(0.43)		(0.44)
SAT math+verbal (or ACT equivalent)	4,600	1148.72	674	1148.75
		(175.15)		(168.58)
Undergraduate college Barron's ranking	6,588	2.81	907	2.91
		(1.24)		(1.26)
Master's degree or higher	7,442	0.51	982	0.54
Location of undergrad or grad school	7,076		940	
DC		0.12		0.17
Maryland or Virginia		0.28		0.28
Outside DC, MD, VA		0.60		0.55

Note: Authors' calculations. Excluding applicants who were not eligible for a teaching license in DC.7,442 total observations. Location indicators are mutually exclusive, applicants with multiple locations coded based on location nearest DC.

Table 4—Pairwise correlations of applicant characteristics and scores

		SAT	GPA	Barr.	Exper.	PCK	Interv.	Aud.	Essay	Personality questions: Extrov.	Personality questions: Pos.	Personality questions: Neg.	Haber-man
2011-2013 applicants	SAT M+V (or ACT equiv)	1											
	Undergraduate GPA	0.31	1										
	College Barron's ranking	0.34	0.13	1									
	Years of experience	-0.05	-0.10	-0.14	1								
	PCK written test	0.22	0.16	0.19	-0.11	1							
	Interview	0.13	0.12	0.08	-0.02	0.10	1						
	Audition	0.08	0.06	0.03	0.04	0.10	0.22	1					
2011 applicants only	General essay	0.19	0.15	0.23	-0.17	0.19	0.17	0.04	1				
	Personality questions												
	Extroversion	0.07	0.02	0.06	-0.12	0.06	0.14	0.13	0.07	1			
	Positive spin	-0.04	0.01	-0.01	0.04	-0.02	0.04	-0.01	0.05	0.28	1		
	Negative spin	-0.05	0.01	-0.04	0.04	-0.02	0.04	-0.04	0.01	0.26	0.70	1	
	Haberman total score	0.21	0.20	0.19	-0.14	0.20	0.12	0.01	0.25	0.13	0.11	0.07	1

Note: Pairwise correlations of applicant characteristics and scores. Maximum observations for a cell is 7,442, see Table 2.

Table 5—Hiring

	Characteristics separately		Characteristics simultaneously	
	(1)	(2)	(3)	(4)
Years prior experience				
1 to 2	0.029*	0.023*	0.027*	0.025*
	(0.012)	(0.011)	(0.011)	(0.011)
3 to 5	0.027*	0.024*	0.024*	0.025*
	(0.012)	(0.011)	(0.011)	(0.011)
6 to 10	0.040**	0.042**	0.041**	0.042**
	(0.012)	(0.011)	(0.011)	(0.011)
11 or more	0.006	0.026*	0.027*	0.029*
	(0.013)	(0.012)	(0.012)	(0.012)
Undergrad GPA (std)	0.006	-0.014**	-0.010*	-0.011*
	(0.004)	(0.004)	(0.004)	(0.004)
SAT math+verbal (std)	0.000	-0.016**	-0.016**	-0.015**
	(0.005)	(0.005)	(0.005)	(0.005)
Barron's Rank (linear 0-5)	0.009*	-0.002	0.001	-0.000
	(0.004)	(0.003)	(0.004)	(0.003)
Master's degree or higher	0.009	-0.001	-0.006	-0.007
	(0.008)	(0.007)	(0.008)	(0.008)
Location of undergrad or grad school				
DC	0.056**	0.057**	0.055**	0.057**
	(0.013)	(0.012)	(0.012)	(0.012)
Maryland or Virginia	0.014	0.026**	0.022**	0.025**
	(0.009)	(0.008)	(0.009)	(0.008)
PCK written test (std)	0.060**	0.008+	0.015**	0.008
	(0.005)	(0.005)	(0.006)	(0.005)
Interview (std)	0.108**	0.028**	0.058**	0.024**
	(0.007)	(0.007)	(0.007)	(0.007)
Audition (std)	0.158**	0.053**	0.148**	0.050**
	(0.010)	(0.012)	(0.010)	(0.012)
Recommended-pool by year FE		√		√
Adjusted R-squared			0.167	0.207
F-statistic subject-applied by year FE			1.65	1.26
p-value			0.000	0.052
F-statistic recommended-pool by year FE				91.8
p-value				0.000

Note: Estimates from linear regressions with 7,442 observations, where being hired is the dependent variable. In columns 1-2 each group of coefficients separated by a solid line are estimates from a separate regression. Columns 3-4 each report estimates from a single regression. Each specification includes year-by-subject-applied fixed effects. Location indicators are mutually exclusive, applicants with multiple locations coded based on location nearest DC. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate.
+ indicates p < 0.10, * 0.05, and ** 0.01

Table 6—Additional characteristics from 2011 applicants and hiring

	Characteristics separately		Characteristics simultaneously	
	(1)	(2)	(3)	(4)
Positive spin factor (std)	-0.015	-0.013	-0.016	-0.013
	(0.012)	(0.011)	(0.012)	(0.011)
Negative spin factor (std)	0.006	0.011	0.006	0.010
	(0.011)	(0.011)	(0.011)	(0.011)
Big Five Index: Extroversion (std)	0.030**	0.019**	0.028**	0.018*
	(0.008)	(0.007)	(0.008)	(0.007)
Haberman total score (std)	0.018*	0.005	0.012	0.002
	(0.007)	(0.007)	(0.008)	(0.007)
General teaching essay (std)	0.018*	0.001	0.015+	0.001
	(0.008)	(0.008)	(0.009)	(0.008)
Recommended-pool FE		√		√
Number of observations			2,360	2,360
Adjusted R-squared			0.027	0.133
F-statistic subject-applied FE			1.95	1.12
p-value			0.002	0.300
F-statistic recommended-pool FE				143
p-value				0.000

Note: Estimates from an LPM with 2,360 observations (all from 2011) where being hired is the dependent variable. In columns 1-2 each group of coefficients separated by a solid line are estimates from a separate regression. Columns 3-4 each report estimates from a single regression. Each specification includes year-by-subject-applied fixed effects. The recommended-pool FE include two mutually exclusive indicators: (i) applicants who pass the audition in 2011, and (ii) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate.
+ indicates p < 0.10, * 0.05, and ** 0.01

Table 7—Job performance

	Characteristics separately			Characteristics simultaneously		
	(1)	(2)	(3)	(4)	(5)	(6)
Years prior experience						
1 to 2	0.081	0.100	0.024	0.100	0.098	0.058
	(0.089)	(0.090)	(0.077)	(0.082)	(0.083)	(0.073)
3 to 5	0.121	0.101	0.109	0.163+	0.156+	0.182*
	(0.101)	(0.099)	(0.083)	(0.088)	(0.088)	(0.078)
6 to 10	0.017	0.034	0.080	0.069	0.064	0.115
	(0.093)	(0.092)	(0.083)	(0.086)	(0.087)	(0.083)
11 or more	-0.255*	-0.214+	-0.163	-0.104	-0.109	-0.054
	(0.115)	(0.117)	(0.112)	(0.109)	(0.109)	(0.105)
Undergrad GPA (std)	0.259**	0.243**	0.185**	0.181**	0.185**	0.160**
	(0.035)	(0.037)	(0.036)	(0.034)	(0.035)	(0.035)
SAT math+verbal (std)	0.173**	0.155**	0.092*	0.019	0.019	-0.011
	(0.040)	(0.039)	(0.037)	(0.038)	(0.038)	(0.038)
Barron's Rank (linear 0-5)	0.155**	0.149**	0.098**	0.108**	0.109**	0.088**
	(0.029)	(0.029)	(0.027)	(0.027)	(0.027)	(0.025)
Master's degree or higher	0.230**	0.215**	0.173**	0.116+	0.112+	0.079
	(0.065)	(0.064)	(0.056)	(0.062)	(0.062)	(0.055)
Location of undergrad or grad school						
DC	-0.030	0.020	0.045	-0.016	-0.020	0.037
	(0.096)	(0.094)	(0.079)	(0.085)	(0.087)	(0.081)
Maryland or Virginia	-0.131+	-0.085	-0.004	-0.076	-0.075	-0.011
	(0.073)	(0.074)	(0.067)	(0.069)	(0.069)	(0.064)
PCK written test (std)	0.279**	0.260**	0.204**	0.182**	0.186**	0.150**
	(0.056)	(0.056)	(0.052)	(0.052)	(0.053)	(0.050)
Interview (std)	0.316**	0.298**	0.270**	0.271**	0.283**	0.257**
	(0.051)	(0.055)	(0.051)	(0.048)	(0.051)	(0.049)
Audition (std)	0.174**	0.149*	0.152*	0.118*	0.104	0.107+
	(0.062)	(0.066)	(0.064)	(0.059)	(0.065)	(0.062)
Recommended-pool by year FE		√	√		√	√
School FE			√			√
Adjusted R-squared				0.176	0.176	0.358
F-statistic recommended-pool by year FE					0.681	0.980
p-value					0.665	0.438
F-statistic school FE						8.932
p-value						0.000

Note: Estimates from least squares regressions with 1,581 teacher-by-year observations, and 917 unique teachers. The dependent variable is job performance measured by the first predicted factor from a factor analysis of IMPACT evaluation component scores, standardized. In columns 1-3 each group of coefficients separated by a solid line are estimates from a separate regression. Columns 4-6 each report estimates from a single regression. Each specification includes year-by-subject-taught fixed effects, and indicators for second year in the district and third year in the district. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses.

+ indicates p < 0.10, * 0.05, and ** 0.01

Table 8—Robustness to parametric selection correction

	Table 7 Columns 4 & 5		With control function	
	(1)	(2)	(3)	(4)
Years prior experience				
1 to 2	0.100	0.098	0.067	0.066
	(0.082)	(0.083)	(0.084)	(0.084)
3 to 5	0.163+	0.156+	0.141	0.131
	(0.088)	(0.088)	(0.088)	(0.088)
6 to 10	0.069	0.064	0.023	0.009
	(0.086)	(0.087)	(0.088)	(0.088)
11 or more	-0.104	-0.109	-0.144	-0.155
	(0.109)	(0.109)	(0.110)	(0.109)
Undergrad GPA (std)	0.181**	0.185**	0.185**	0.191**
	(0.034)	(0.035)	(0.034)	(0.034)
SAT math+verbal (std)	0.019	0.019	0.034	0.040
	(0.038)	(0.038)	(0.039)	(0.039)
Barron's Rank (linear 0-5)	0.108**	0.109**	0.106**	0.109**
	(0.027)	(0.027)	(0.027)	(0.027)
Master's degree or higher	0.116+	0.112+	0.134*	0.134*
	(0.062)	(0.062)	(0.063)	(0.062)
Location of undergrad or grad school				
DC	-0.016	-0.020	-0.059	-0.069
	(0.085)	(0.087)	(0.088)	(0.087)
Maryland or Virginia	-0.076	-0.075	-0.098	-0.096
	(0.069)	(0.069)	(0.069)	(0.069)
PCK written test (std)	0.182**	0.186**	0.167**	0.175**
	(0.052)	(0.053)	(0.053)	(0.051)
Interview (std)	0.271**	0.283**	0.237**	0.232**
	(0.048)	(0.051)	(0.054)	(0.053)
Audition (std)	0.118*	0.104	0.084	0.068
	(0.059)	(0.065)	(0.066)	(0.059)
Predicted probability of hire			1.411*	-0.017
			(0.708)	(0.681)
Predicted probability of hire ^ 2			-1.154	0.865
			(0.781)	(0.807)
Recommended-pool by year FE		√		
F-statistics excluded instruments			23.37	2.30

Note: Estimates from least squares regressions with 1,581 teacher-by-year observations, and 917 unique teachers. The dependent variable is our standardized job performance factor from IMPACT evaluation component scores. Columns 1 and 2 simply repeat the estimates in Table 7 Columns 4 and 5 for convenient comparison. Columns 3 and 4 are estimated just as Column 1 is, except that we add a quadratic function of the predicted probability of hire. The predicted probability of hire is estimated using the specification reported in Table 5 Column 3 (all characteristic and score regressors and subject-applied by year fixed effects, but no recommended-pool by year fixed effects) but with additional instruments added as regressors. The hiring prediction regressions include 7,442 observations.

For the estimates reported above in Column 3, the instruments in the hire equation are four indicator variables: (i) Applicants in any year who scored above the stage 4 cut-score designated by DCPS as the threshold for the recommended pool. (ii) Applicants in 2011 who scored above the stage 3 cut-score; we assume these applicants were also placed in the recommended pool as discussed in the text. (iii) Applicants in 2011 who scored below the stage 2 cut-score but were nevertheless randomly selected to move on to stage 3. (iv) Applicants in 2011 who applied in the first weeks of the recruitment season. All of these early applicants were allowed to move on to stage 3 regardless of their scores in stage 2 or 1.

For the estimates in Column 4, the added instruments include five indicator variables: (i)-(iii) Applicants in any year who scored above the cut-score in stage 2, 3, and 4 respectively. And again for 2011 (iv) applicants randomly selected to advance or (v) early applicants automatically advanced. We also allow the slope on each stage score to be different above and below the stage cut-score, and include fixed effects for the highest stage an applicant was invited to complete. All these added coefficients are allowed to vary by year.

Clustered (teacher) standard errors in parentheses.

+ indicates $p < 0.10$, * 0.05, and ** 0.01

Table 9—Additional characteristics from 2011 applicants and teacher job performance

	(1)	(2)
Positive spin factor (std)	0.032	0.015
	(0.061)	(0.062)
Negative spin factor (std)	-0.000	0.023
	(0.074)	(0.074)
Big Five Index: Extroversion (std)	-0.001	-0.016
	(0.059)	(0.059)
Haberman total score (std)	0.291**	0.270**
	(0.054)	(0.054)
General teaching essay (std)	0.222**	0.187**
	(0.070)	(0.069)
Recommended-pool by year FE		√

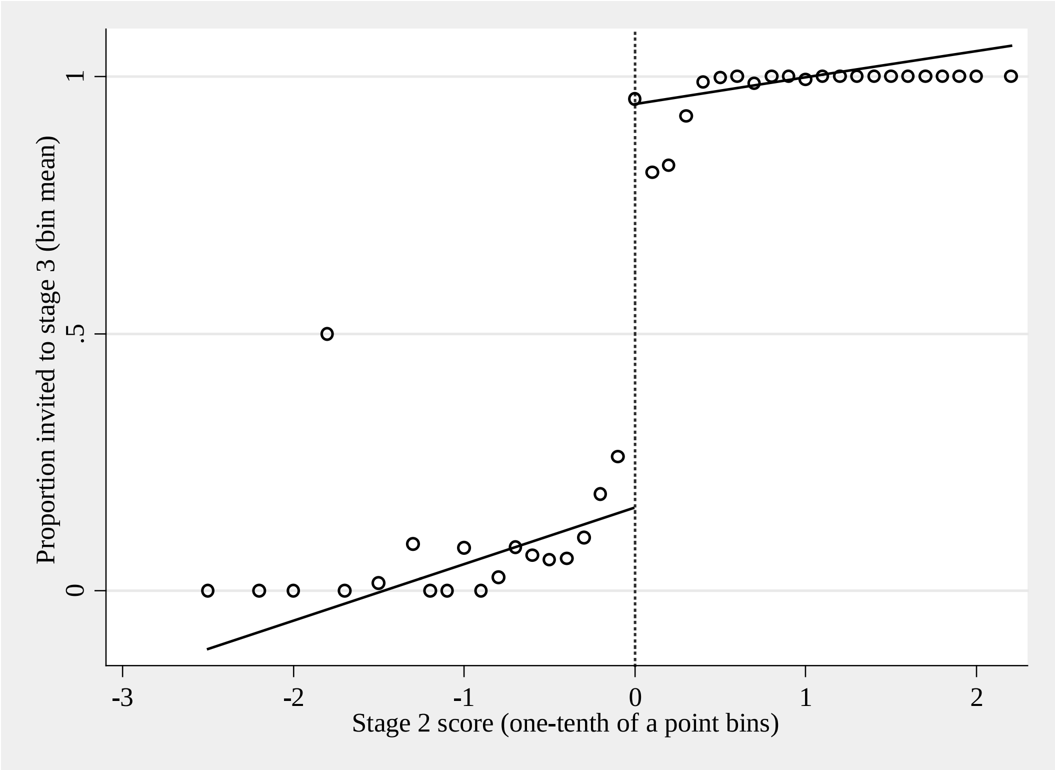
Note: Estimates from least squares regressions with 744 teacher-by-year observations, and 314 unique teachers (hired in 2011 only). The dependent variable is job performance measured by the first predicted factor from a factor analysis of IMPACT evaluation component scores, standardized. Each group of coefficients separated by a solid line are estimates from a separate regression. Each specification includes year-by-subject-taught fixed effects, and indicators for second year in the district and third year in the district. The recommended-pool FE include two mutually exclusive indicators: (i) applicants who pass the audition in 2011, and (ii) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses.
+ indicates p < 0.10, * 0.05, and ** 0.01

Table 10—Attrition after first year

	Leave DCPS		Leave school	
	(1)	(2)	(3)	(4)
Years prior experience (novice omitted)				
1 to 2	0.011	0.006	0.005	-0.001
	(0.043)	(0.040)	(0.047)	(0.045)
3 to 5	-0.029	0.010	0.013	0.056
	(0.041)	(0.040)	(0.046)	(0.045)
6 to 10	-0.003	0.010	0.035	0.051
	(0.043)	(0.039)	(0.049)	(0.045)
11 or more	-0.018	-0.029	-0.004	-0.019
	(0.049)	(0.047)	(0.056)	(0.054)
Undergrad GPA (std)	0.013	0.030+	-0.002	0.018
	(0.017)	(0.016)	(0.019)	(0.019)
SAT math+verbal (std)	0.030	0.033+	0.035+	0.041*
	(0.018)	(0.019)	(0.020)	(0.020)
Barron's Rank (linear 0-5)	0.000	0.004	-0.001	0.005
	(0.013)	(0.012)	(0.015)	(0.014)
Master's degree or higher	0.050+	0.058*	0.040	0.048
	(0.030)	(0.028)	(0.034)	(0.032)
Location of undergrad or grad school				
DC	-0.148**	-0.124**	-0.205**	-0.184**
	(0.037)	(0.035)	(0.040)	(0.038)
Maryland or Virginia	-0.052	-0.044	-0.053	-0.047
	(0.034)	(0.032)	(0.039)	(0.037)
PCK written test (std)	-0.019	-0.010	-0.037	-0.029
	(0.020)	(0.019)	(0.024)	(0.023)
Interview (std)	-0.015	0.005	-0.031	-0.007
	(0.021)	(0.021)	(0.025)	(0.024)
Audition (std)	-0.044	-0.037	-0.051	-0.043
	(0.030)	(0.027)	(0.032)	(0.030)
IMPACT rating FE		√		√
Adjusted R-squared	0.024	0.123	0.026	0.115
DCPS mean for outcome	0.199	0.199	0.276	0.276

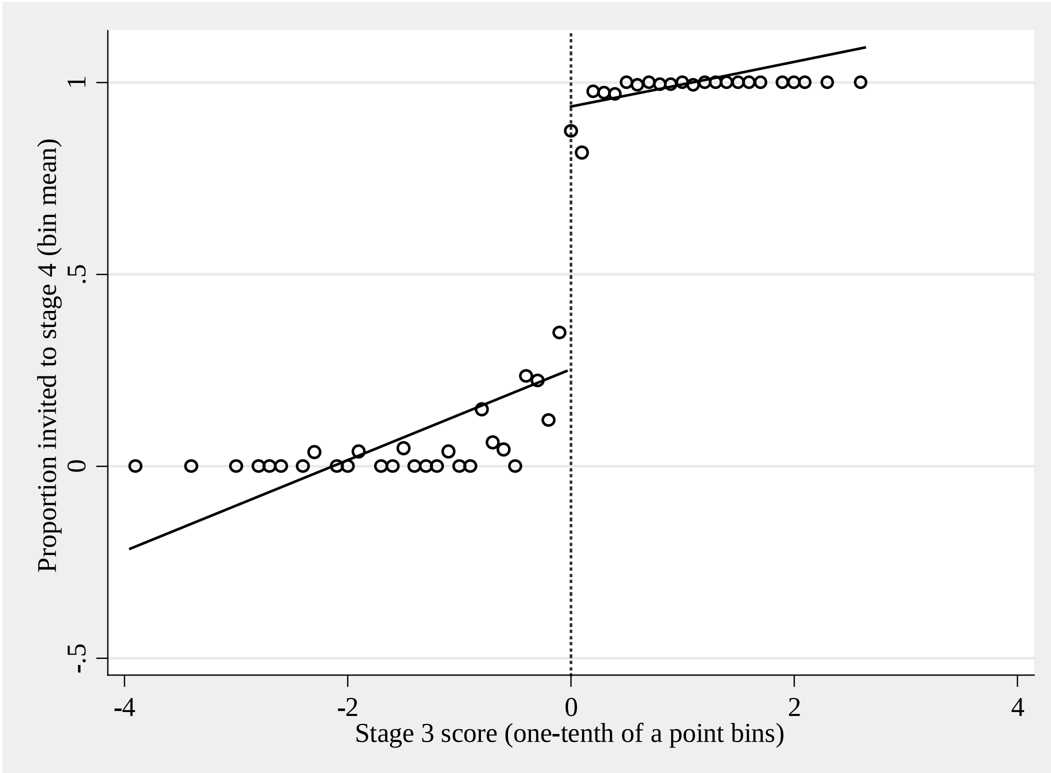
Note: Estimates from least squares regressions with 902 teacher observations. The dependent variable in Columns 1 and 2 is an indicator for having left DCPS after their first year, while in Columns 3 and 4 it is an indicator for leaving DCPS or the school in which they taught during their first year. Each specification includes year-by-subject-taught fixed effects and recommended-pool by year fixed effects. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate.

+ indicates $p < 0.10$, * 0.05, and ** 0.01



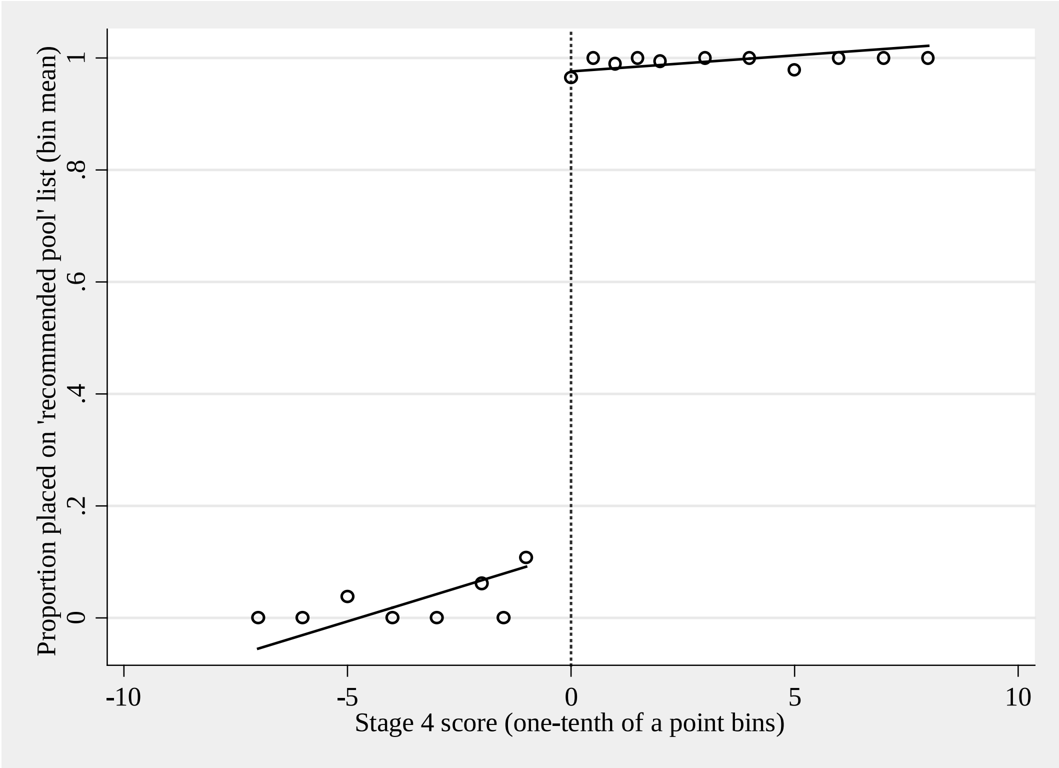
Appendix Figure 1—Applicant stage 2 score and likelihood of being invited to stage 3

Note: Each marker represents a bin of applicant observations, grouped by stage 2 score into one-tenth of a point bins. The y-axis measures the bin mean of an indicator = 1 if the applicant was invited to complete stage 3. The x-axis has been re-centered so that the official cut-off for being invited to stage 3 is set to zero. The plot also shows the bi-variate linear fit from a regression using applicant observations, not bin observations.



Appendix Figure 2—Applicant stage 3 score and likelihood of being invited to stage 4

Note: Each marker represents a bin of applicant observations, grouped by stage 3 score into one-tenth of a point bins. The y-axis measures the bin mean of an indicator = 1 if the applicant was invited to complete stage 4. The x-axis has been re-centered so that the official cut-off for being invited to stage 4 is set to zero. The plot also shows the bi-variate linear fit from a regression using applicant observations, not bin observations.



Appendix Figure 3—Applicant stage 4 score and likelihood of being placed on the recommended pool list

Note: Each marker represents a bin of applicant observations, grouped by stage 4 score into one-tenth of a point bins. The y-axis measures the bin mean of an indicator = 1 if the applicant was placed on the recommended pool list. The x-axis has been re-centered so that the official cut-off for being in the recommended pool is set to zero. The plot also shows the bi-variate linear fit from a regression using applicant observations, not bin observations.

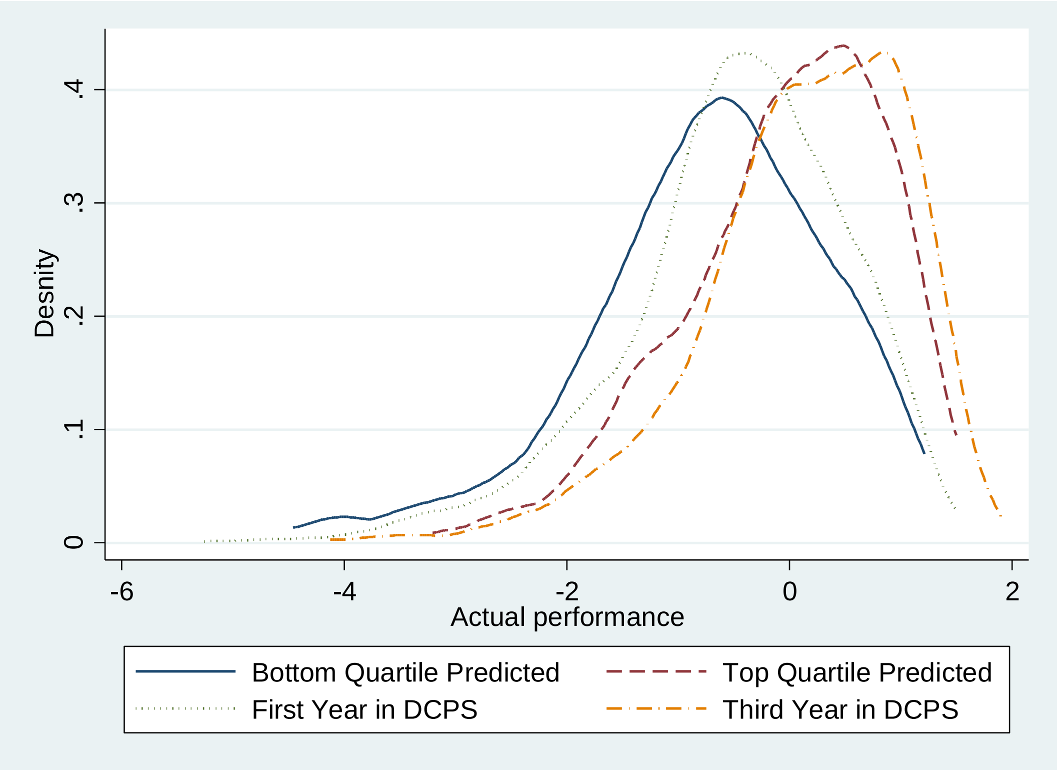


Figure 4—Relationship between screening measures and performance

Note: Kernel densities estimated separately for the top and bottom quartile of predicted performance, and for teachers in their first and third years working in DCPS as teachers. Predicted performance is the fitted value obtained after the following regression: Using the sample of new hires, fit a regression similar to Table 7 Column 4. The dependent variable is the first predicted factor of IMPACT scores. The covariates include all of the background characteristics and application scores. The specification also includes subject-taught by year fixed effects.

Appendix Table A1—Factor analysis of personality and grit questions

	Positive Spin	Negative Spin	Extraversion
+ BFI Q3	**0.43**	-0.10	0.09
+ BFI Q8	**0.44**	-0.04	-0.02
+ BFI Q15	**-0.41**	-0.13	0.06
+ BFI Q17	**0.55**	0.01	-0.06
+ BFI Q23	**0.39**	0.22	0.00
+ BFI Q26	**0.45**	0.00	-0.01
+ BFI Q28	**0.48**	-0.02	0.03
+ Grit Q2	**0.26**	0.17	0.00
+ Grit Q4	**0.55**	-0.06	-0.05
+ Grit Q7	**0.41**	0.28	-0.11
+ Grit Q8	**0.60**	0.11	-0.11
- BFI Q5	0.09	**0.43**	0.01
- BFI Q11	0.09	**0.46**	-0.03
- BFI Q12	0.20	**0.17**	0.11
- BFI Q22	0.00	**0.69**	0.01
- BFI Q24	0.17	**0.20**	-0.07
- BFI Q25	0.20	**0.26**	0.02
- BFI Q29	-0.07	**-0.42**	-0.17
- Grit Q1	-0.04	**0.64**	0.01
- Grit Q3	-0.04	**0.62**	0.08
- Grit Q5	-0.01	**0.51**	0.04
- Grit Q6	0.08	**0.54**	-0.06
Extraversion Q1	0.04	-0.18	**0.60**
Extraversion Q4	-0.15	0.04	**0.60**
Extraversion Q7	**0.40**	-0.05	**0.34**
Extraversion Q10	**0.43**	-0.06	**0.34**
Extraversion Q13	-0.11	0.07	**0.71**
Extraversion Q16	0.19	-0.01	**0.34**
Extraversion Q19	-0.16	0.25	**0.62**
Extraversion Q21	0.26	-0.08	**0.59**

Note: This table presents the results of a factor analysis on items used to assess the Big Five personality traits as well as Duckworth's Grit measure, where factors having an eigenvalue greater than 1 were retained. Factor weights are given with a Promax Rotation. Items with "+" are positively scored, those with "-" are negatively scored. BFI refers to big five items that do not pertain to extraversion. Note that items BFI 15 and 29 relate to neuroticism, and hence are inversely related to the underlying factors.

Appendix Table A2—IMPACT component weights

	2011-12		2012-13 and 2013-14	
	Group 1	Group 2	Group 1	Group 2
Individual value-added	0.50		0.35	
Teacher assessed student learning		0.10	0.15	0.15
Teaching and learning framework	0.35	0.75	0.40	0.75
Commitment to school community	0.10	0.10	0.10	0.10
School value-added	0.05	0.05		

Source: DCPS.

Appendix Table A3—Hiring and offers, 2012 and 2013

	Hired				Offered			
	Characteristics separately		Characteristics simultaneously		Characteristics separately		Characteristics simultaneously	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years prior experience								
1 to 2	0.030*	0.020	0.024+	0.023+	0.048**	0.040**	0.042**	0.042**
	(0.014)	(0.013)	(0.013)	(0.013)	(0.016)	(0.015)	(0.015)	(0.015)
3 to 5	0.027+	0.023+	0.023+	0.024+	0.071**	0.067**	0.068**	0.069**
	(0.014)	(0.012)	(0.013)	(0.013)	(0.016)	(0.015)	(0.015)	(0.015)
6 to 10	0.051**	0.047**	0.046**	0.046**	0.093**	0.089**	0.088**	0.089**
	(0.014)	(0.012)	(0.013)	(0.013)	(0.016)	(0.015)	(0.015)	(0.015)
11 or more	0.021	0.037**	0.035*	0.037**	0.067**	0.082**	0.081**	0.084**
	(0.015)	(0.013)	(0.014)	(0.014)	(0.017)	(0.016)	(0.017)	(0.017)
Undergrad GPA (std)	0.015**	-0.009+	-0.004	-0.006	0.013*	-0.008	-0.003	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
SAT math+verbal (std)	0.002	-0.012*	-0.014*	-0.012*	0.001	-0.013+	-0.017*	-0.014*
	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Barron's Rank (linear 0-5)	0.008+	-0.004	-0.001	-0.002	0.009+	-0.002	0.003	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Master's degree or higher	0.019*	0.002	-0.004	-0.005	0.004	-0.011	-0.010	-0.011
	(0.010)	(0.008)	(0.010)	(0.009)	(0.011)	(0.010)	(0.011)	(0.011)
Location of undergrad or grad school								
DC	0.046**	0.046**	0.046**	0.046**	0.077**	0.076**	0.073**	0.072**
	(0.014)	(0.013)	(0.013)	(0.013)	(0.016)	(0.015)	(0.015)	(0.015)
Maryland or Virginia	0.015	0.028**	0.024*	0.027**	0.027*	0.039**	0.039**	0.041**
	(0.011)	(0.010)	(0.010)	(0.010)	(0.013)	(0.012)	(0.012)	(0.012)
PCK written test (std)	0.063**	0.009	0.017*	0.007	0.068**	0.016*	0.026**	0.015*
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Interview (std)	0.096**	0.022**	0.041**	0.018*	0.102**	0.030**	0.051**	0.024*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Audition (std)	0.177**	0.051**	0.170**	0.050**	0.183**	0.069**	0.171**	0.062**
	(0.011)	(0.013)	(0.011)	(0.013)	(0.013)	(0.015)	(0.013)	(0.015)
Highest-stage-reached by year FE								
Recommended-pool by year FE		√		√		√		√
Adjusted R-squared			0.191	0.239			0.154	0.190
F-statistic subject-applied by year FE			1.78	1.34			1.82	1.61
p-value			0.000	0.048			0.000	0.003
F-statistic recommended-pool by year FE				158				111
p-value				0.000				0.000

Note: Estimates from linear regressions with 7,442 observations, where being hired or being offered a job is the dependent variable. In columns 1-2 each group of coefficients separated by a solid line are estimates from a separate regression. Columns 3-4 each report estimates from a single regression. Each specification includes year-by-subject-applied fixed effects. Location indicators are mutually exclusive, applicants with multiple locations coded based on location nearest DC. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate.

+ indicates p < 0.10, * 0.05, and ** 0.01

Appendix Table A4—Hiring by subject area

	Elementary and ECE	MS/HS core	Special education	Other
	(1)	(2)	(2)	(4)
Years prior experience				
1 to 2	0.045*	0.043*	0.064	-0.025
	(0.019)	(0.020)	(0.040)	(0.030)
3 to 5	0.027	0.059**	0.021	-0.018
	(0.020)	(0.019)	(0.037)	(0.030)
6 to 10	0.065**	0.045*	0.055	-0.004
	(0.021)	(0.021)	(0.038)	(0.030)
11 or more	0.032	0.014	0.041	0.008
	(0.023)	(0.023)	(0.042)	(0.031)
Undergrad GPA (std)	-0.011	-0.009	-0.009	-0.012
	(0.008)	(0.008)	(0.014)	(0.012)
SAT math+verbal (std)	-0.008	-0.016+	-0.045**	-0.011
	(0.009)	(0.009)	(0.016)	(0.014)
Barron's Rank (linear 0-5)	0.005	-0.001	-0.010	-0.014
	(0.006)	(0.006)	(0.012)	(0.010)
Master's degree or higher	0.013	-0.006	-0.008	-0.017
	(0.014)	(0.014)	(0.028)	(0.022)
Location of undergrad or grad school				
DC	0.079**	0.036	0.038	0.059+
	(0.021)	(0.023)	(0.038)	(0.030)
Maryland or Virginia	0.022	0.026+	0.002	0.039+
	(0.015)	(0.016)	(0.028)	(0.023)
PCK written test (std)	0.009	-0.005	0.014	0.025+
	(0.009)	(0.009)	(0.018)	(0.014)
Interview (std)	0.023+	0.033*	0.019	0.012
	(0.013)	(0.013)	(0.022)	(0.020)
Audition (std)	0.036+	0.071**	0.055	0.033
	(0.020)	(0.021)	(0.043)	(0.032)
Adjusted R-squared	0.218	0.223	0.231	0.177
Number of observations	2,441	2,059	710	1,208

Note: Estimates from linear regressions where being hired is the dependent variable. Each specification includes year-by-subject-applied fixed effects and recommended-pool by year fixed effects. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. ECE stands for early childhood education. "Core" middle school (MS) and high school (HS) subjects include English, math, science, and social studies. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate.
+ indicates p < 0.10, * 0.05, and ** 0.01

Appendix Table A5—Robustness to parametric selection correction, using job offer in first-stage in place of hired, 2012 and 2013

	Like Table 7 Columns 4 & 5		With control function	
	(1)	(2)	(3)	(4)
Years prior experience				
1 to 2	-0.027	-0.031	-0.073	-0.062
	(0.110)	(0.109)	(0.111)	(0.113)
3 to 5	0.053	0.032	-0.023	-0.011
	(0.123)	(0.124)	(0.125)	(0.130)
6 to 10	-0.033	-0.042	-0.134	-0.115
	(0.120)	(0.120)	(0.126)	(0.128)
11 or more	-0.262+	-0.276+	-0.368*	-0.353*
	(0.146)	(0.147)	(0.149)	(0.155)
Undergrad GPA (std)	0.240**	0.239**	0.236**	0.244**
	(0.054)	(0.054)	(0.054)	(0.054)
SAT math+verbal (std)	0.003	0.002	0.029	0.024
	(0.055)	(0.054)	(0.057)	(0.056)
Barron's Rank (linear 0-5)	0.048	0.050	0.042	0.046
	(0.035)	(0.035)	(0.035)	(0.035)
Master's degree or higher	0.017	0.014	0.054	0.033
	(0.087)	(0.087)	(0.086)	(0.086)
Location of undergrad or grad school				
DC	-0.030	-0.036	-0.095	-0.073
	(0.111)	(0.112)	(0.112)	(0.110)
Maryland or Virginia	-0.068	-0.054	-0.099	-0.081
	(0.094)	(0.095)	(0.094)	(0.093)
PCK written test (std)	0.144+	0.148+	0.123	0.132
	(0.081)	(0.081)	(0.081)	(0.081)
Interview (std)	0.202**	0.207**	0.167*	0.183**
	(0.067)	(0.068)	(0.069)	(0.068)
Audition (std)	0.157+	0.149+	0.106	0.108
	(0.084)	(0.083)	(0.083)	(0.085)
Predicted probability of job offer			2.298*	-0.315
			(1.092)	(0.918)
Predicted probability of job offer ^ 2			-1.847	1.019
			(1.223)	(1.021)
Recommended-pool by year FE		√		
F-statistics excluded instruments			23.62	4.43

Note: Estimates from least squares regressions with 1,581 teacher-by-year observations, and 917 unique teachers. The dependent variable is our standardized job performance factor from IMPACT evaluation component scores. Columns 1 and 2 show estimates like in Table 7 Columns 4 and 5 but restricted to the 2012 and 2013 cohorts. Columns 3 and 4 are estimated just as Column 1 is, except that we add a quadratic function of the predicted probability of hire. The predicted probability of hire is estimated using the specification reported in Table 5 Column 3 (all characteristic and score regressors and subject-applied by year fixed effects, but no recommended-pool by year fixed effects) but with additional instruments added as regressors. The hiring prediction regressions include 7,442 observations.

For the estimates reported above in Column 3, the instruments in the hire equation are four indicator variables: (i) Applicants in any year who scored above the stage 4 cut-score designated by DCPS as the threshold for the recommended pool. (ii) Applicants in 2011 who scored above the stage 3 cut-score; we assume these applicants were also placed in the recommended pool as discussed in the text. (iii) Applicants in 2011 who scored below the stage 2 cut-score but were nevertheless randomly selected to move on to stage 3. (iv) Applicants in 2011 who applied in the first weeks of the recruitment season. All of these early applicants were allowed to move on to stage 3 regardless of their scores in stage 2 or 1.

For the estimates in Column 4, the added instruments include five indicator variables: (i)-(iii) Applicants in any year who scored above the cut-score in stage 2, 3, and 4 respectively. And again for 2011 (iv) applicants randomly selected to advance or (v) early applicants automatically advanced. We also allow the slope on each stage score to be different above and below the stage cut-score, and include fixed effects for the highest stage an applicant was invited to complete. All these added coefficients are allowed to vary by year.

Clustered (teacher) standard errors in parentheses.

+ indicates $p < 0.10$, * 0.05, and ** 0.01

Appendix Table A6—Job performance by school characteristics

	Elementary schools	Middle & high schools	School % FRL < median	School % FRL ≥ median
	(1)	(2)	(3)	(4)
Years prior experience				
1 to 2	0.042	0.213	-0.027	0.209+
	(0.106)	(0.147)	(0.122)	(0.111)
3 to 5	0.024	0.374*	0.112	0.214+
	(0.120)	(0.146)	(0.129)	(0.126)
6 to 10	-0.005	0.232	0.067	0.152
	(0.113)	(0.144)	(0.128)	(0.117)
11 or more	-0.206	0.149	-0.246	0.075
	(0.148)	(0.185)	(0.171)	(0.142)
Undergrad GPA (std)	0.144**	0.180**	0.199**	0.162**
	(0.046)	(0.055)	(0.053)	(0.048)
SAT math+verbal (std)	0.001	0.061	0.072	-0.038
	(0.051)	(0.065)	(0.058)	(0.052)
Barron's Rank (linear 0-5)	0.106**	0.107*	0.067	0.113**
	(0.035)	(0.049)	(0.044)	(0.036)
Master's degree or higher	0.111	0.145	0.111	0.081
	(0.080)	(0.111)	(0.097)	(0.082)
Location of undergrad or grad school				
DC	-0.065	0.011	0.092	-0.087
	(0.110)	(0.152)	(0.129)	(0.119)
Maryland or Virginia	-0.093	-0.102	-0.026	-0.097
	(0.091)	(0.124)	(0.102)	(0.096)
PCK written test (std)	0.247**	0.115	0.098	0.214**
	(0.080)	(0.077)	(0.072)	(0.079)
Interview (std)	0.264**	0.287**	0.292**	0.277**
	(0.068)	(0.076)	(0.079)	(0.069)
Audition (std)	0.087	0.266+	0.031	0.196+
	(0.077)	(0.139)	(0.078)	(0.104)
Adjusted R-squared	0.171	0.147	0.148	0.175
Number of observations	948	601	741	816
Number of teachers	570	328	413	494

Note: Estimates from least squares regressions where the dependent variable is job performance measured by our standardized IMPACT performance measure. Each specification includes year-by-subject-taught fixed effects, indicators for second year in the district and third year in the district, and recommended-pool by year fixed effects. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses.

+ indicates p < 0.10, * 0.05, and ** 0.01

Appendix Table A7—Components of job performance

	TLF class observation				
	Principal	Master Ed.	CP	CSC	TAS
	(1)	(2)	(3)	(4)	(5)
Years prior experience					
1 to 2	0.207**	0.176*	-0.009	-0.002	0.029
	(0.080)	(0.073)	(0.027)	(0.084)	(0.091)
3 to 5	0.282**	0.244**	-0.015	0.033	0.066
	(0.083)	(0.076)	(0.028)	(0.089)	(0.090)
6 to 10	0.182*	0.132	-0.048	-0.033	0.025
	(0.087)	(0.083)	(0.029)	(0.089)	(0.089)
11 or more	0.014	0.004	-0.002	-0.189+	-0.239*
	(0.110)	(0.095)	(0.035)	(0.110)	(0.117)
Undergrad GPA (std)	0.130**	0.134**	0.051**	0.164**	0.109**
	(0.033)	(0.031)	(0.012)	(0.038)	(0.036)
SAT math+verbal (std)	-0.008	0.011	0.005	0.006	0.036
	(0.036)	(0.036)	(0.012)	(0.041)	(0.038)
Barron's Rank (linear 0-5)	0.088**	0.099**	0.028**	0.083**	0.035
	(0.026)	(0.025)	(0.009)	(0.028)	(0.029)
Master's degree or higher	0.033	0.001	0.012	0.153*	0.032
	(0.059)	(0.055)	(0.021)	(0.065)	(0.064)
Location of undergrad or grad school					
DC	0.088	0.009	-0.039	-0.077	0.042
	(0.082)	(0.077)	(0.028)	(0.089)	(0.092)
Maryland or Virginia	-0.043	-0.021	-0.027	-0.101	0.031
	(0.068)	(0.063)	(0.023)	(0.072)	(0.071)
PCK written test (std)	0.149**	0.055	0.024	0.185**	0.119*
	(0.050)	(0.045)	(0.015)	(0.049)	(0.049)
Interview (std)	0.274**	0.213**	0.040*	0.229**	0.170**
	(0.047)	(0.045)	(0.018)	(0.052)	(0.056)
Audition (std)	0.095	0.080	0.050*	0.135*	-0.052
	(0.062)	(0.054)	(0.021)	(0.066)	(0.069)

Note: Estimates from least squares regressions with 1,581 observations and 917 teachers. The dependent variables are indicated in the column headers. Each column reports estimates from a single regression. Each specification includes year-by-subject-taught fixed effects, recommended-pool by year fixed effects, and indicators for second year in the district and third year in the district. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses.
+ indicates p < 0.10, * 0.05, and ** 0.01

Appendix Table A8—Components of job performance for IVA sample

		TLF class observation					Value-added		
	Overall Perform.	Princi-pal	Master Educ.	CP	CSC	TAS	Average	Math	Reading
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Years prior experience									
1 to 2	0.305	0.270	0.273	0.063	0.100	0.276	0.307	0.293	-0.000
	(0.221)	(0.214)	(0.214)	(0.081)	(0.228)	(0.266)	(0.213)	(0.280)	(0.289)
3 to 5	0.581*	0.644**	0.445*	0.145	0.300	0.340	0.501*	0.595	0.173
	(0.228)	(0.203)	(0.206)	(0.088)	(0.233)	(0.263)	(0.223)	(0.456)	(0.266)
6 to 10	0.309	0.267	0.166	-0.090	0.220	0.248	0.355+	0.532*	0.094
	(0.199)	(0.210)	(0.215)	(0.076)	(0.197)	(0.246)	(0.189)	(0.235)	(0.261)
11 or more	0.247	0.181	0.215	0.021	0.160	-0.279	0.672**	0.655*	0.643*
	(0.266)	(0.307)	(0.274)	(0.104)	(0.257)	(0.307)	(0.240)	(0.309)	(0.283)
Undergrad GPA (std)	0.097	0.047	0.073	0.048	0.107	0.096	-0.124	-0.061	-0.170
	(0.081)	(0.074)	(0.072)	(0.037)	(0.087)	(0.098)	(0.080)	(0.111)	(0.109)
SAT math+verbal (std)	-0.010	-0.064	-0.083	-0.021	-0.044	0.159	0.115	0.076	0.090
	(0.090)	(0.086)	(0.092)	(0.036)	(0.091)	(0.111)	(0.099)	(0.126)	(0.118)
Barron's rank (linear 0-5)	0.174**	0.080	0.180**	0.035	0.159**	0.023	0.060	0.188*	0.006
	(0.058)	(0.058)	(0.056)	(0.027)	(0.061)	(0.077)	(0.062)	(0.085)	(0.083)
Master's degree or higher	0.180	0.184	0.136	0.064	0.099	-0.236	0.455**	0.448+	0.575**
	(0.154)	(0.149)	(0.142)	(0.062)	(0.162)	(0.178)	(0.148)	(0.232)	(0.192)
Location of undergrad or grad school									
DC	-0.340+	-0.109	-0.322	-0.180*	-0.213	-0.383	-0.189	0.055	-0.319
	(0.183)	(0.190)	(0.206)	(0.082)	(0.183)	(0.241)	(0.179)	(0.257)	(0.239)
Maryland or Virginia	0.097	0.007	0.122	-0.004	0.097	0.009	0.109	-0.074	0.238
	(0.162)	(0.157)	(0.140)	(0.071)	(0.163)	(0.173)	(0.164)	(0.272)	(0.201)
PCK written test (std)	0.188+	0.141	0.037	0.047	0.182+	0.073	0.072	0.241+	0.008
	(0.096)	(0.104)	(0.097)	(0.036)	(0.095)	(0.101)	(0.102)	(0.137)	(0.125)
Interview (std)	0.393**	0.454**	0.411**	0.114*	0.233*	0.237+	-0.046	0.157	-0.119
	(0.109)	(0.101)	(0.102)	(0.046)	(0.118)	(0.124)	(0.091)	(0.147)	(0.115)
Audition (std)	-0.092	-0.113	0.115	-0.002	0.033	-0.234	-0.212+	-0.201	-0.171
	(0.154)	(0.159)	(0.141)	(0.051)	(0.154)	(0.159)	(0.125)	(0.169)	(0.164)
Teacher-year observations	286	286	286	286	286	286	286	157	210
Teacher observations	195	195	195	195	195	195	195	108	148

Note: Estimates from least squares regressions. The dependent variables are indicated in the column headers. Each column reports estimates from a single regression. Each specification includes year-by-subject-taught fixed effects, recommended-pool by year fixed effects, and indicators for second year in the district and third year in the district. The recommended-pool by year FE include four mutually exclusive indicators: (i) applicants who reach the recommended pool in 2013, (ii) applicants who reach the recommended pool in 2012, (iii) applicants who pass the audition in 2011, and (iv) applicants who pass the interview but not the audition in 2011 (see the text for more details). The left out category is all other applicants. When a covariate is missing for an observation, we set the value to zero and include an indicator = 1 for anyone missing that covariate. Clustered (teacher) standard errors in parentheses.
+ indicates p < 0.10, * 0.05, and ** 0.01