NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Making the Cut: The Effectiveness of Teacher Screening and Hiring in the Los Angeles Unified School District

Paul Bruno
Katharine O. Strunk

WORKING PAPER 184 • January 2018

Making the Cut: The Effectiveness of Teacher Screening and Hiring in the Los Angeles Unified School District

Paul Bruno
University of Southern California

Katharine O. Strunk
Michigan State University/CALDER

Contents

Acknowledgements…………………………………………………………………….....ii

Abstract……………………………………………………………………………………iii

I. Introduction…………………………………………………………………………1

II. Previous Literature…..……………. ……………………………………………..2

III. LAUSD's Multiple Measures Teacher Selection Process……………………………...8

IV. Data & Methods……………………………………………………………………14

V. Results………………………………………………………………………………26

VI. Discussion and Policy Implications…………………………………………………...41

References………………………………………………………….……….........44

Tables and Figures.………………………………………………………….............53

Acknowledgements

We are grateful to the Los Angeles Unified School District for providing the data necessary to conduct this research, and in particular to Vivian Ekchian, Sergio Franco, Leanne Hannah, Bryan Johnson, Cynthia Lim, Emily Mohr and Lydia Stephens for their partnership in developing the research agenda and Inocencia Cordova, Joshua Klarin, Jonathan Lesser, Jacob Guthrie and Kevon Tucker-Seeley for their assistance in obtaining and understanding the data.

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Making the Cut: The Effectiveness of Teacher Screening and Hiring in the Los Angeles Unified School District

Paul Bruno, Katharine O. Strunk

CALDER Working Paper No. 184
January 2018

Abstract

Despite evidence that many schools and districts have considerable discretion when hiring teachers and the existence of an extensive literature on teacher quality, little is known about how best to hire teachers. This is, in part, because predicting teacher quality using readily-observable teacher characteristics has proven difficult and there is very little evidence linking information collected during the teacher hiring process to teachers' outcomes once they are hired. We contribute to this literature using data from a recently-adopted teacher screening system in the Los Angeles Unified School District (LAUSD) that allows applicant records to be linked to student- and teacher-level data for those teachers who are subsequently employed in the district. We find that performance during screening, and especially performance on specific screening assessments, is significantly predictive of applicants' eventual employment in LAUSD and teachers' later contributions to student achievement, evaluation outcomes, and attendance, but not to teacher mobility or retention. However, applicants' performance on individual components of the screening process are differentially predictive of different teacher outcomes, highlighting potential trade-offs faced by districts during screening. In addition, we find suggestive evidence across time and between districts that the shift to the new teacher screening system improved hiring outcomes in LAUSD relative to other similar districts and schools.

Keywords: Teacher Quality, Teacher Hiring, Los Angeles

Introduction

Despite widespread concerns about teacher shortages, many schools and districts continue to receive more applications for open teaching positions than they have vacancies and numerous newly-certified teachers do not get hired into teaching positions at all (Cowan, Goldhaber, Hayes, & Theobald, 2016; Engel, Jacob, & Curran, 2014). This suggests that many administrators have substantial discretion when hiring new teachers, and given that teachers vary considerably in their effectiveness (e.g., Chetty, Friedman, & Rockoff, 2014a, 2014b; Hanushek & Rivkin, 2012), how this discretion is exercised may have important implications for students and schools. However, a great deal remains unknown about how teachers are – or should be – screened and hired (Strunk, Marsh, & Bruno, 2017).

Using detailed applicant data from a new district-level teacher screening system, entitled the Multiple Measure Teacher Selection Process (MMTSP), in the Los Angeles Unified School District (LAUSD), as well as teacher- and student-level administrative data on the outcomes of teachers who are hired and the students they teach, we investigate the manner in which teachers are hired in the second-largest school district in the country. These data capture many applicant characteristics that are often difficult to observe and allow for novel analyses of both the relative employment prospects of teachers in a large, urban labor market and of the potential for improving teacher quality through predictive screening.

Results indicate that LAUSD's MMTSP captures information that is of interest to school administrators, as applicants with better performance during district-level screening are more likely to be subsequently employed as teachers in the district even as the school-level leaders making final hiring decisions do not know prospective teachers' exact screening scores. Additionally, overall performance during screening is significantly and meaningfully predictive of teachers' outcomes once they are hired, including their attendance, contributions to student achievement, and final performance evaluation ratings. However, screening performance is not predictive of teacher retention and individual components of the screening assessment are differentially predictive of different teacher outcomes, implying that districts that define teacher quality narrowly during screening may face trade-offs in terms of the attributes of teachers they eventually hire. We also find evidence over time and across districts suggesting that teacher hiring outcomes may have improved since LAUSD adopted its new screening system.

The remainder of this paper proceeds as follows. First, we summarize the existing literature on teacher screening and hiring and articulate the research questions with which we contribute to extant work. We then describe the screening and administrative processes in the district, followed by a discussion of the data we employ and our empirical strategies. This is followed by a presentation of results and, finally, a discussion of their implications for school district human resource operations and our understanding of teacher quality and teacher labor markets.

Previous Literature

There is considerable evidence that teachers are substantially heterogeneous in quality. For example, teachers vary in measurable and important ways in their contributions to student achievement (e.g., Chetty et al., 2014b; Hanushek, 1992; Hanushek & Rivkin, 2012; Koedel, Mihaly, & Rockoff, 2015), with those differences explaining the largest share (among observable school inputs) of variation in student achievement outcomes (Goldhaber, 2016). Over and above their direct effects on student achievement or other student outcomes, teachers can impose costs on schools through mechanisms such as absenteeism (Clotfelter, Ladd, & Vigdor, 2009; Miller, Murnane, & Willett, 2008) and turnover (Hanushek, Rivkin, & Schiman, 2016; Milanowski & Odden, 2007; Ronfeldt, Loeb, & Wyckoff, 2013), all of which could be mitigated by better-informed hiring decisions. And school administrators often face considerable barriers to dismissing teachers once they have been hired (e.g., Associated Press, 2008;

Griffith & McDougald, 2016; Painter, 2000), which may make it preferable to screen teachers more carefully during the hiring phase, when administrative discretion is greater.

Nevertheless, it may be particularly difficult to ascertain during a hiring process what qualities will make for a successful teacher for at least two reasons. First, districts may be constrained in their hiring for a variety of legal, bureaucratic, and institutional reasons. For example, in addition to having to navigate largely predetermined instructional calendars, district administrators may be constrained by collectively-bargained labor agreements (e.g., governing transfer rights for existing teachers) and bureaucratic dysfunction (e.g., delayed information about budgets) (Levin & Quinn, 2003; Strunk, 2014). These factors may preclude additional applicant screening, and may help to explain why hiring processes are often rushed – taking place shortly before or after the start of the school year – and thus "information-poor" (Liu & Johnson, 2006).

Second, despite the well-documented variation in teacher quality discussed above, predicting teacher effectiveness using readily-observable teacher characteristics has proven challenging. Consistent with evidence from other fields indicating that cognitive ability tests are strong predictors of worker performance (Ryan & Tippins, 2004), there is evidence that cognitive ability matters for teachers (Harris & Rutledge, 2010). This includes studies finding that teachers' math or verbal abilities, licensure test scores, subject matter knowledge, and knowledge of how to teach particular content are predictive of multiple effectiveness outcomes, especially student achievement (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008; Chingos & Peterson, 2011; Clotfelter, Ladd, & Vigdor, 2006, 2007; Ehrenberg & Brewer, 1995; Goldhaber, 2007; Hill, Rowan, & Ball, 2005; Rockoff, Jacob, Kane, & Staiger, 2010; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013).[1] However, many of these teacher attributes may be difficult for district administrators to observe and with a few exceptions (Darling-Hammond, Berry, & Thoreson,

[1] The evidence on the relevance of teachers' cognitive ability is not entirely consistent; some studies do not find similar results (Hanushek, Rivkin, Rothstein, & Podgursky, 2004; Harris & Sass, 2011).

2001; Ehrenberg & Brewer, 1994) most studies find that more easily-observed educational credentials (e.g., certification or advanced degrees) that might be expected to proxy for cognitive ability appear at most weakly predictive of teacher effectiveness (Angrist & Guryan, 2008; Buddin & Zamarro, 2009; Chingos & Peterson, 2011; Clotfelter et al., 2007; Goldhaber & Brewer, 2000, 2001; Hanushek, 2003; Kane, Rockoff, & Staiger, 2008; Monk, 1994), though possible exceptions include college selectivity (Rice, 2003) and credentials or coursework related to a teacher's content area (Feng & Sass, 2013; Goldhaber & Brewer, 1998; Kukla-Acevedo, 2009; Monk, 1994; Rice, 2003; Wayne & Youngs, 2003). In addition, there is a large body of work that finds clear evidence that teachers' experience is associated with their teaching effectiveness, both in their novice years and beyond (e.g., Boyd et al., 2008; Kraft & Papay, 2015; Ladd & Sorenson, 2016; Rockoff, 2004; Wiswall, 2013).

The body of literature relating teachers' noncognitive attributes, such as personalities and values, to their effectiveness is smaller, perhaps because these characteristics are even more difficult to observe than cognitive abilities, and is somewhat mixed. Some studies do find that teachers' self-efficacy (Caprara, Barbaranelli, Steca, & Malone, 2006), grit (Duckworth et al., 2009; Robertson-Kraft & Duckworth, 2014), leadership abilities (Dobbie, 2011), and values (e.g., commitment to student learning; Metzger & Wu, 2008) predict evaluation outcomes and student achievement. However, Rockoff et al. (2010) consider a range of personality traits and outcomes and find few significant relationships.

Inconsistencies across studies may reflect differences in measures of teacher characteristics (e.g., general vs. subject-specific abilities) or effectiveness (e.g., student gain scores vs. VAMs), and even similar-seeming measures may be sensitive to district context.[2] Measures of cognitive ability may also be crude or noisy; this would explain why direct ability tests and subject-specific credentials appear to be more valuable than credentials generally, and why composite measures of cognitive and noncognitive

[2] For example, even value-added measures of teachers' contributions to student achievement appear to be somewhat sensitive to the choice of test administered to students (Lockwood et al., 2007; Papay, 2011), which will tend to vary across states and over time.

ability are more predictive than measures used individually (Dobbie, 2011; Goldhaber, 2007; Rockoff et al., 2010).

For the purpose of screening teachers prior to hire, one limitation of the research discussed so far linking teacher attributes and teacher effectiveness is that it does not typically employ data collected and utilized in screening and hiring, instead drawing inferences from administrative records or survey data assembled after teachers have been hired. This makes it difficult to know whether measures of teacher attributes would predict teacher effectiveness similarly well (or poorly) if used in districts' actual practice. At the same time, the difficulty of predicting teacher effectiveness has prompted some districts to adopt screening devices that are intended to be more rigorous, such as structured interview protocols or standardized batteries of assessments. Many districts now pre-screen candidates more carefully, provide principals with measures of past performance for transfer applicants, or require principals to justify hiring decisions (Cannata et al., 2017). These reforms are appealing for their potential to collect information about applicants that might otherwise be difficult to observe, and to ensure that that information is utilized when making hiring decisions. However, search models of matching in the labor market emphasize that such screening may also entail trade-offs if, for example, it is costly to implement or changes for the worse the composition of the applicant pool (e.g., Delfgaauw & Dur, 2007; Oyer & Schaefer, 2011). Evaluating these novel (or otherwise poorly understood) hiring processes is therefore of both theoretical and practical importance. However, to date only a small number of studies directly link information collected during screening to student and teacher outcomes after hire.

For example, Metzger and Wu (2008) conduct a meta-analysis of 24 studies of a commonly-used teacher screening instrument – Gallup's Teacher Perceiver Interview (TPI) – intended to assess a range of applicant values and beliefs about education and teaching (e.g., empathy and commitment). They find that aggregate TPI ratings are weakly or moderately predictive of some teacher outcomes, especially attendance and ratings by administrators. Additionally, TPI's predicative validity appears somewhat weaker when the assessment is administered during the hiring process, rather than for research purposes during employment, suggesting that predictors of active teacher effectiveness may not always generalize to prospective teachers during screening.

More recently, Goldhaber, Grout, and Huntington-Klein (2017) examine data from a district-level teacher screening system in Spokane, Washington, and find mixed evidence that the hiring process is sensitive to teacher quality. Applicants who are rated more highly during the screening process have higher value-added measures of effectiveness and lower attrition, but not higher attendance, and these relationships appear to be driven by teacher characteristics that are more difficult to observe (e.g., classroom management ability as demonstrated during screening) rather than those characteristics that are easier to observe and more commonly studied in the literature (e.g., certification and education).

An important consideration when estimating relationships between applicant characteristics and teacher effectiveness is that they may be biased by relationships between unobserved applicant attributes and the probability of being hired. For example, consider the possibility that both observed applicant undergraduate GPA and unobserved applicant charisma are independently predictive of teachers' contributions to student achievement, and that applicants with low GPAs are hired only if they have particularly high charisma. Even if GPA and charisma are unrelated in the population of applicants, this will tend to bias estimates of the relationship between GPA and teacher effectiveness downward as teachers with high GPAs and average charisma are compared to teachers with lower GPAs but higher charisma.

Goldhaber et al. (2016) are able to address this kind of selection concern by exploiting plausibly exogenous variation in the probability that applicants are hired. Specifically, some applicants to Spokane are accidentally given erroneous screening scores (making them eligible to be hired when they otherwise would not be) or face lower levels of competition from applicants to the same position. This allows relationships between screening performance and subsequent effectiveness to be estimated for applicants whose hiring was essentially random and thus not subject to selection on unobservable applicant attributes. Their results suggest that the magnitude of such bias is small.

Using data on teacher applicants in Washington, D.C., Jacob et al. (2016) find that applicants are more likely to be hired if they have prior experience, but no more (and perhaps less) likely to be hired if they have higher undergraduate GPAs or college entrance exam scores, attended more competitive schools, or have a master's degree. Despite being no more likely to be hired, applicants with these stronger academic backgrounds have superior subsequent performance if hired as measured by a composite multiple-measures evaluation outcome, as do applicants who perform better during screening on a written assessment of teaching knowledge, during an interview, or during a teaching audition, or who possess a graduate degree. Again, the authors attempt to correct for selection bias using plausibly exogenous variation in hiring probabilities – in this case, discontinuities in the probability of advancing through the stages of screening for similarly-performing applicants – and again the apparent bias in their uncorrected estimates is generally small.

More recently, Sajjadiani et al. (2018) find that information contained in prospective teachers' resumes, unlikely to be widely used during hiring but discernable through machine learning techniques, is predictive of subsequent work outcomes. Moreover, their simulations suggest that incorporating this information when making hiring decisions would meaningfully improve the composition of hired applicants, suggesting potential gains to acquiring and using additional information about applicants during the hiring process

Although these studies yield important information about districts' abilities to screen potential teachers for quality, there is a clear need for more research – in new contexts and using different screening methods – to identify the teacher attributes that are meaningfully predictive during screening and hiring, the circumstances in which those attributes are predictive, and the extent to which hiring outcomes can be improved by imposing more rigorous – and costly – new teacher screening systems. We contribute to this literature using data from a new screening system recently adopted by the Los Angeles Unified School District (LAUSD). In LAUSD prospective teachers apply to the district office, where they are screened before school administrators make hiring decisions. Beginning in 2013, LAUSD began reforming and standardizing this screening process, and as of the 2014-15 school year applicants were screened using the Multiple Measures Teacher Selection Process, a highly standardized system with eight components (e.g., a writing sample and the delivery of a sample lesson), each scored according to rubrics aligned to district goals (e.g., employee evaluation criteria). In addition to potentially changing the quality distribution of teachers who are eventually hired, these reforms result in the capture of many applicant characteristics that are not typically quantified in administrative data. This allows for novel analyses both of applicants' relative employment prospects and of the predictive validity of various measures of applicant quality. We thus attempt to answer three interrelated questions about new teacher hiring: 1) *Which applicant characteristics, as measured during screening, are predictive of effectiveness after hire?*; 2) *Could the information collected during screening be used in the hiring process more effectively?*; 3) *Has the quality of new teacher hires improved in LAUSD as the new screening system has been adopted?*

LAUSD's Multiple Measures Teacher Selection Process

In LAUSD prospective teachers apply directly to the district office, and the district estimates that it receives approximately 10,000 applicants for approximately 1,250 certificated positions each year. These applicants are subsequently screened by human resources (HR) specialists, and to make this screening process manageable the received applications are eliminated from the screening process in stages. Until recently this occurred in two stages, with applicants first being eliminated if they failed to submit completed applications or did not meet minimum certification requirements (i.e., an undergraduate degree and teaching certification). Applicants who were not eliminated at this first stage then would be invited in for an interview with the district based on district hiring needs, and remaining applicants would be eliminated solely on the basis of interview performance.

The new process, referred to as the Multiple Measures Teacher Selection Process (MMTSP), changes the second stage of this process, essentially replacing the old central office interview. Whereas before 2014-15, all minimally eligible teachers were eligible to be interviewed (based on need for teachers with specific certifications), and the site-based interview provided the sole information beyond minimal qualifications used in the hiring determination, under the MMTSP applicant evaluations became more standardized, rigorous, and explicitly aligned to district goals such as the specific criteria by which LAUSD teachers are evaluated. The new MMTSP process is illustrated in Figure 1. In the first stage applicants are, as before, selected out on the basis of their applications' completeness (automatically checked by the district's digital application platform) and a manual review by HR staff to ensure that applicants meet minimum (unscored) certification criteria (e.g. possession of an undergraduate degree). The district estimates that the first stage of screening excludes as many as half of all applications, most commonly because applications are incomplete or because applicants do not hold required credentials or have applied to positions for which there are few, if any, vacancies. Applicants can note on their initial applications their interest in working in each of the district's six regional "local districts," although this information does not inform eligibility for employment at any stage in the process.

In the second stage, applicants who pass the minimal eligibility requirement are then scored on eight separate assessments. The subarea scores sum to a possible total of 100 points, and to be eligible for employment applicants must obtain minimum passing scores on several of these assessments as well as a total score of at least 80. These assessments are scored according to rubrics, many of which are explicitly aligned to the districts' Teaching and Learning Framework (TLF), the criteria by which classroom teachers are evaluated during classroom observations.

Table 1 summarizes the eight individual criteria on which applicants are evaluated under the MMTSP. Based on their initial applications, each applicant receives up to 10 points on the basis of their *undergraduate GPA*. Another 10 points are awarded for *subject matter preparation*, which are determined by applicants' scores on their subject-specific licensure exams or, when the exam requirement has been waived, again on the basis of their GPAs. Unlike most of the other assessment scores, which can be as low as zero or one, the lowest possible subject matter score is eight. A small number of points are also granted for meeting any of several miscellaneous criteria that the district considers desirable, and these points are awarded on a binary basis with applicants receiving either all of the points or none of them. In particular, three *background points* are given to candidates who have certain prior LAUSD (non-teaching) experience (e.g., serving as a teaching assistant), have specific prior leadership (e.g., military) experience, possess a master's (or higher) degree, or are recruited through Teach for America.[3] Finally, two *preparation points* can be given to applicants who attended a school highly-ranked by *U.S. News & World Report*, who can show evidence of prior teaching effectiveness (e.g., student achievement data), or who majored in their credential subject field (or, if multi-subject, majored in a core academic subject or liberal studies).[4]

Applications who meet minimum eligibility requirements are also screened in a "pre-interview" stage with two components. The first component is the solicitation of standardized electronic *professional references*. Candidates are rated by their references on such attributes as "professionalism" and "ethical conduct" and, if appropriate, aspects of their teaching (e.g., "classroom management") on a scale ranging from "ineffective" to "highly effective." Applicants are given a score of up to 20 points on the basis of these ratings, with any "ineffective" ratings resulting in a score of zero

---

[3] The district hires up to 35 Teach For America teachers each year, all who teach in special education placements. District HR personnel explained that they give background points to TfA-recruited teachers because of evidence that TfA teachers produce strong results in mathematics.

[4] We observe only whether applicants received background or preparation points, not the specific reason why they did so.

and the candidate's elimination. The second component of this pre-interview process is the offsite completion of an online writing task in which applicants respond to a series of vignettes, describing how they would navigate situations they might face as teachers. The resulting *writing sample* is given a score of up to 15 points by HR staff on the basis of a rubric aligned with the district's TLF as well as on overall grammar and organization, and applicants with a score below 11 on the writing sample are automatically eliminated.

After the pre-interview stage applicants are invited in to complete the final stages of the screening process, though the district estimates that approximately one-third decline to do so for various reasons. This leaves between 3,000 and 5,000 applicants per year who are in fact brought in to the district office for *interviews* and to give *sample lesson* demonstrations, again both scored using rubrics explicitly aligned to the TLF. The interview is structured (i.e., using pre-planned questions), is designed to assess both knowledge of teaching and attitudes toward work, and is worth up to 25 points for applicants with scores below 20 resulting in disqualification. The sample lesson demonstration is administered to two HR specialists playing the scripted roles of students and is worth up to 15 points, and applicants are disqualified by scores below 11.

The district accepts and screens applications on a year-round basis. Applicants who receive a total score of at least 80 and the minimum required scores on each scored screening assessment (e.g., at least 11 on the interview) are placed on the district's eligibility list, which is then given to personnel specialists and school administrators who have related vacancies to fill.[5] These site-based administrators then interview the candidates in whom they are interested, based on information provided to them by the district. Importantly, actual screening scores are never provided to personnel specialists or school

[5] School administrators have considerable flexibility in how, if at all, eligible candidates will be further screened at the school site (e.g., through on-site interviews). Personnel specialists can also indicate to school administrators if specific candidates list preferences for their geographic region.

administrators as long as the applicant "passes" the screening with a score of at least 80.[6] This is also the case if the score would have been over 80 but the applicant fails to meet an individual cut point on one element of the screening process (e.g., scoring at least an 11 on the sample lesson).

There are two exceptions by which applicants can be placed on the eligibility list despite failing to obtain a minimum overall score or screening component score.[7] First, though all candidates must be assessed by HR, a school principal can request that a particular candidate receive an exception to the score requirements. Principals are then free to hire that candidate if they so choose, but they are notified in writing by the district that the applicant does not meet the standard screening criteria. Second, applicants who fail to meet one of the individual assessment cut points, or fail to achieve an overall score of 80, are given a blind review by a panel of HR specialists. This review incorporates the full range of submitted application materials (minus identifying information). Following this review, applicants that the panel deems sufficiently high-quality are added back to the eligibility list.[8] As shown in Figure 1, approximately 200 applicants each year are granted exemptions through one of these two avenues. Such applicants remain on the eligibility list for up to one year, and if they are not hired may reapply and be screened again.[9]

[6] Communications with district personnel indicate that school-site administrators are not provided with pre-employment evaluation scores due to confidentiality issues related to California Education Code.
[7] These exceptions pertain to the eight scored screening assessments rather than, for example, legal certification requirements.
[8] Communications with district personnel indicate that this panel is particularly concerned with ensuring that applicants are not uniformly dismissed for low undergraduate GPAs, and are willing to use GPAs from Masters' degree programs up to award points up to the minimum required cut-point of 8 out of 10 for applicants with sufficient graduate GPAs. This panel also reviews cases in which references or HR staff raise serious and specific concerns about the fitness of applicants who would otherwise be deemed eligible to hire.
[9] Because we do not observe applications to the district that fail to reach the eligibility list we cannot know with certainty how frequently applicants reapply. Of the 5,396 unique individuals observed on the eligibility list over the three years in this study, only 80 appear more than once and none appear more than twice. Among these 80 individuals, correlations between screening scores received in their first and second appearances on the eligibility list are mostly weak to moderate, ranging from r = -.06 (professional references) to r = .63 (GPA), with overall scores correlating at r = .14.

Simple regressions predicting screening scores based on teachers' certification area and quarter of first eligibility (shown in Appendix Table 1) show that applicants with math and special education certifications have lower overall scores, and lower sub-scores across most of the individual component scores, relative to applicants with elementary certifications. Relative to applicants who apply over the summer (July-September), winter (January – March) and spring (April – June) applicants have higher screening scores overall and in most subareas, whereas fall applicants (October-December), who effectively add to the "late hire" pool after the school year starts, have significantly lower overall and subarea scores. In addition, relative to applicants with elementary certifications, all specialized teachers are significantly more likely to receive a minimum score exemption to be placed on the eligibility list, and summer applicants are least likely to get a minimum score exemption. Late fall applicants are two times more likely to have an exception to be placed on the eligibility list.

Table 2 highlights the screening score and sub-scores of applicants who are eventually employed by LAUSD and those who are not. The average score of both sets of applicants exceeds the baseline cut point of 80, but employed applicants score approximately two points higher and there is a much smaller standard deviation, suggesting less variability in the assessed quality of employed relative to unemployed applicants (also seen by the minimum score of 17 for the latter group). Average sub-scores are generally slightly higher in all areas for employed relative to unemployed applicants, and the variation is greater for the unemployed group across all sub-scores. Exceptions include the subject matter score, which is virtually the same between the two groups, and score exceptions; fewer teaches placed in the eligibility pool due to exemptions are eventually hired. Teachers certified to teach special education are more likely to be employed and those certified to teach social studies are less. This likely reflects demand in LAUSD.

Data & Methods

Ultimately, an average of approximately 1,700 applicants are placed on the eligibility list each year (where they can remain for up to one year), and approximately three-quarters of applicants placed on the eligibility list are subsequently hired by the district. Importantly for our study, over the course of three years (2014-15 through 2016-17) we observe only those individuals who are placed on the eligibility list and their hiring outcomes in LAUSD, or 5,476 applications in total, approximately 10 percent of which are present despite the applicant failing to meet a minimum score requirement (i.e., having received an exception as described above), though we are unable to observe the precise reason for which they received an exception.[10,11] We observe teacher effectiveness outcomes through the 2015-16 school year.[12] For the purposes of most analyses below screening scores are standardized to have a mean of zero and standard deviation of one across all applications on the eligibility list.

Correlations between each individual assessment score and the overall score (shown in Appendix Table 2) are generally moderate in strength, with individual correlations ranging from $r = .17$ to $r = .58$. However, the individual assessment scores are much more weakly correlated with one another, with no correlation between any two screening assessment scores greater than $r = .38$. This suggests that the different screening assessments capture distinct information and are not heavily redundant with one another.[13] It also has the implication that the results presented below do not

---

[10] Importantly, we cannot ascertain if eligible applicants drop out of the process of their own volition, either by refraining to move on in the process if invited or declining offers of employment.

[11] As can be seen in Table 2, we observe slightly fewer individual scores because in a small number of cases no score was recorded or a score was discarded as erroneous for lying outside the range of possible scores. (Based on correspondence with HR staff, we replace missing background or preparation points with zeros.) We also set overall scores to missing if they were constructed on the basis of an erroneous sub-score. However, the results presented below are essentially unchanged if erroneous scores are used or if applications with any erroneous score are dropped altogether.

[12] Our observations also include a small number of applicants – less than one percent of the total – who were screened during an earlier pilot phase of the MMTPS and hired in the 2013-14 school year, though in practice this does not affect results.

[13] These correlations are only for applicants on the eligibility list and who have thus completed the entire screening process. We do not observe other applicants, and therefore cannot determine assessment score correlations for the entire population of applicants. Among applicants who are present on the eligibility list because they received a

change substantially if the assessment scores are used to predict teachers' outcomes individually rather than jointly.[14]

MMTSP scores of individuals who are hired by the district can also be merged to other administrative data linking teachers to schools and students, allowing us to observe these teachers' evaluation outcomes, mobility, attendance, and (in the case of math and English/language arts teachers in tested grades) value-added measures (VAMs) of effectiveness. These data also include student demographic information, including student race, English language learner status and eligibility for free- or reduced-price lunch or special educations services.

Evaluation data are provided to us by the LAUSD's human resources division. They come from 2014-15 and 2015-16 implementations of LAUSD's Educator Development and Support: Teachers (EDST) program, by which teachers are evaluated on the basis of a combination of classroom observations, pre- and post-observation conferences, and other work products (e.g., department meeting agendas, student work samples, or parent call logs). During these years EDST evaluations were required for all non-permanent (i.e., non-tenured) teachers as well as a subset of permanent teachers.[15] As part of their EDST evaluations teachers receive ratings on 15 priority elements of the TLF, which address teachers' planning and preparation of instruction, classroom environments, delivery of instruction, and professional growth.[16] Teachers also receive a final summative rating, though the details of these ratings differ across the two years we utilize here. In 2014-15 teachers received focus element ratings

score exception, correlations between scores are in some cases larger and in some cases smaller than for all applicants on the eligibility list, though still never greater than $r = .39$. Additionally, score correlations are more frequently negative among applicants receiving exceptions, particularly when considering sample lesson scores, consistent with these exceptions being granted more frequently in cases where applicant performance was neither uniformly strong nor uniformly weak across assessments.

[14] Note that even applicants' GPA and subject matter scores, which can in principle both be awarded based on applicants' GPAs, are not highly correlated. This is likely due to the fact that subject matter scores can be no lower than eight, while GPA scores can range from one to 10.

[15] Permanent teachers were to be evaluated every other year or, with sufficient and satisfactory experience, as infrequently as every five years.

[16] The TLF consists of 61 elements total, and teachers and evaluators may choose to utilize additional components though in practice few do so formally.

on a four-point scale ranging from “ineffective” to “highly effective” and received a final summative rating of either “below standard performance” or “meets standard performance.” In 2015-16 the number of rating categories for the focus elements was reduced to three with the elimination of the “highly effective” rating and the number of final rating categories was increased to three by the addition of a rating of “exceeds standard performance.” Though initially intended to be made on the basis of two formal classroom observations, during the 2014-15 school year the number of required formal observations was reduced to one for teachers who performed sufficiently well during their initial observations.

We focus on two outcomes from these evaluations. First, we consider the probability that a teacher receives a final summative rating of “below standard performance.” This rating category represents a virtually identical share of all final evaluation ratings in 2014-15 (4.2 percent) as when the additional “exceeds standard performance” category was added in 2015-16 (4.3 percent), suggesting that the change in the number of categories did not substantially increase differentiation by evaluators among lower-performing teachers. Second, we average teachers’ ratings across the TLF elements for which they received ratings by converting their focus element ratings to numeric values of one to three, giving the two highest ratings a value of three in 2014-15 to allow comparability across years. Again, the availability of the higher rating category appears to increase differentiation primarily among higher-rated teachers; the share of focus area ratings receiving a three or four in 2014-15 (77 percent) is very similar to the share receiving a three in 2015-16 (75 percent).[17]

Teacher mobility data come from annual district-wide employee records associating teachers with school sites at the beginning of the year. Because employee identifiers link teachers over time in these data it is possible to observe whether a teacher has changed sites or left district employment

[17] Results are nearly identical if instead of being compressed in 2014-15 ratings are averaged on a one-to-three or one-to-four-point scale in their respective years and then standardized to have a mean of zero and standard deviation of one within each year. Results are available upon request.

between one year and the next. We identify teachers as switching schools in a given year if in the subsequent year they appear in a different LAUSD school and as leaving the district if they are not employed by the district at all the following year. We do not count teachers as moving or exiting if their school sites in the current year were closed in a given year of if they retire.

HR records provided by the district also include attendance records for all staff. For each teacher we observe the number of hours they missed assigned work hours for legally-protected reasons (e.g., jury duty, military leaves, or leaves protected by the Family Medical Leave Act) as well as those that were missed for unprotected reasons (e.g., regular illness days, bereavement leaves, or personal necessity days), with six hours representing a full teacher workday for administrative purposes. The district also constructs an overall attendance rate for each teacher, defined as the percentage of contracted work hours for which a teacher was present. Each of these variables is defined separately for certificated and non-certificated attendance in the event that an employee holds both a certificated and non-certificated (i.e., non-teaching) position in the district. Because our focus is teachers we use only the certificated attendance data for any given employee, though in practice this makes little difference as few teachers hold a simultaneous non-certificated position.

We construct value-added measures of teachers' contributions to student achievement in math and ELA using student-level test score and demographic data. We link teachers to students using annual report card data, and students are linked to all teachers for whom they are indicated as receiving instruction in math or ELA, respectively, with each student-teacher link weighted on the basis of the share of the student's instructional time for which they are assigned to that teacher. We then estimate teacher VAMs in each year as the teacher fixed effect in a regression of each students' math or ELA achievement (standardized within test and year) on their achievement in the prior year (in both subjects, similarly standardized) and a vector of student characteristics (e.g., indicators for students' gender, race, free- or reduced-price lunch eligibility, and gifted, special education, and English learner status), and a set of teacher fixed effects.[18] Math and ELA VAMs are estimated separately, as are VAMs at the elementary and secondary level. These teacher VAMs are then standardized to have a mean of zero and standard deviation of one across all teachers in each subject-level-year, and teachers with both elementary and secondary students being given an overall subject area VAM that is the mean of their elementary and secondary VAMs.

Student achievement data, provided to us by LAUSD's Office of Data and Accountability, come from California's statewide standardized tests. These data present a challenge because the state recently transitioned from the previous Standardized Testing and Reporting (STAR) system, used through 2012-13, to a new set of tests aligned to the Common Core State Standards, collectively referred to as the California Assessment of Student Performance and Progress (CAASPP) system beginning 2014-15. This transition from the STAR system to the CAASPP system included a transition year – 2013-14 – during which student test scores were not released. Thus even within the CAASPP period it is not possible to estimate VAMs identically in all years because prior year test scores are not available during the 2014-15 school year. We attempt to address this by controlling for students' test scores two years prior, using data from the CST in 2012-13. This has the advantage of increasing our sample size, which is helpful given that the CAASPP system reduces the numbers of grade levels in which students are tested and thus the number of teachers for whom VAMs can be estimated in any single year.

However, whether controlling for student achievement lagged twice is adequate to remove bias in the VAM estimates is not obvious, and there is evidence that VAM estimation is sensitive to differences in tests and test administration (e.g., Lockwood et al., 2007; Papay, 2011) and to transitions across testing regimes, particularly in ELA (Backes et al., 2018). Our results do differ somewhat, though

[18] This is similar to what Hock and Isenberg (2017) describe as the "full roster" method to account for shared responsibility for students across teachers. Detailed information about how VAMs are estimated is available upon request.

not in a consistent direction, if data from only 2015-16 are used, though even when VAMs from 2014-15 are included the sample sizes on these outcomes are relatively small, as shown in Table 2, which summarizes these new teacher outcome variables.

These combined teacher-level data can then be used to answer our first two research questions. To evaluate whether the information collected during screening is predictive of subsequent employment performance (RQ 1), and similar to Goldhaber et al. (2017), we estimate the following model:

$$outcome_{ist} = \theta_0 + \theta_1 S_i + \theta_2 X_{ist} + \theta_3 D_{st} + \sum_{c=1}^{C} \alpha^c C_i^c + \gamma_t + \varepsilon_{ist} \qquad (1)$$

Here *outcome* is a measure of teacher *i*'s value-added contribution to student achievement in school *s* in year *t*, or, alternatively, teacher *i*'s attendance outcomes or average EDST ratings as defined above. To predict teacher final binary evaluation outcomes we use logistic regression to predict the odds that a new teacher is given an unsatisfactory rating. To understand the relationship between screening score and teacher mobility (exit the district or switch schools within the district, relative to staying in the same school), we use a multinomial logistic regression. The predictors of interest are teacher screening scores and are contained (individually or jointly) in *S*.[19] Recall that the MMTSP awards background and preparation scores for a variety of miscellaneous applicant attributes, such as leadership or being recruited via Teach for America, and for each of those scores applicants either receive all of the points or none of them. Because these background and preparation points can each take only two values, we include them as dummy variables indicating whether the points were awarded or not. X is a set of teacher characteristics, including an indicator of whether the teacher was hired despite failing to meet minimum eligibility requirements and indicators of the number of years since hire and whether the teacher holds a graduate degree.[20]

[19] Some components of the screening system, such as fully-digitized inputs of some screening performance indicators, were phased in over time, but results do not change substantially if the earliest screening periods are excluded. Results available upon request.

[20] A concern with controlling for possession of a graduate degree is that prospective teachers can earn background points during screening for possessing one. However, background points are a small fraction of all points and can

Model 1 also includes *D*, a set of school characteristics that may be associated with teacher outcomes including grade level and district region indicators, and the share of students in the school who are neither white nor Asian, who are English learners, or who are eligible for free- or reduced-price lunch (FRL) or special education services.[21] To allow for the possibility that teachers have systematically different outcomes if they teach different subjects or in different years we include sets of dummy variables indicating teacher certification subject area (e.g., elementary vs. mathematics) (C_i^c) and school years (γ_t). Standard errors are clustered at the teacher level because individual teachers may be observed in multiple years after they are hired.[22]

An important consideration in the estimation of equation (1) is the possibility that teachers' selection into employment in LAUSD and/or into specific kinds of schools may bias the estimated relationships between screening scores and eventual outcomes. Although both Goldhaber et al. (2017) and Jacob et al. (2016) find little evidence of selection bias in their studies of conceptually similar processes in Spokane, Washington and Washington, D.C., we are nonetheless concerned with two potential sources of bias. First, it is possible that applicants who perform better during the screening process have other unobservable or unassessed characteristics that are associated with both screening scores and subsequent effectiveness (variously measured). This will tend to produce coefficients that are in some sense biased if teacher effectiveness is determined by these unobserved factors (e.g., physical health) rather than the attributes directly assessed during screening *per se* (e.g., GPA). Given that the

be awarded for several other reasons; only 37 percent of individuals who receive background points possess a graduate degree in their first year of employment. Controlling for possession of a graduate degree therefore matters little in practice and allows for the possibility that individuals may acquire additional degrees after completing screening.

[21] In results not shown but available upon request, we also estimate models replacing observable school characteristics with a school fixed effect. For most outcome measures, this substantially reduces our effective sample size due to many school sites having only one newly-hired teacher or no variation in the outcome of interest (e.g., unsatisfactory evaluation outcomes). Most estimates are not strongly sensitive to the choice of specification, though we discuss differences below when relevant.

[22] Standard errors change only very slightly, and not in a consistent direction, if they are instead clustered on schools. Results available upon request.

purpose of the MMTSP is primarily the prediction, rather than the causal explanation, of effectiveness we do not view this kind of bias as a significant problem.

Second, it is possible that estimated relationships between screening scores and teacher outcomes are biased by the selection of employed teachers into particular kinds of teaching placements. For instance, if teachers with higher scores are more likely to be recruited into and accept positions in schools with "easier" working conditions (e.g., higher test scores, fewer students who qualify for free- or reduced-price lunches, or with principals who are more lenient in evaluations), this would produce a spurious relationship that reflects selection into school environments rather than true effectiveness. We are not fully able to rule out such selection bias, but we do attempt to test for it in several ways, including examining self-reported applicant school preferences, checking to see whether estimates are sensitive to controls for observable school characteristics, and by making within-school comparisons through the use of a school fixed effect. As we will describe in greater detail below, none of these strategies are perfect, but they leave us with the general impression that, as in earlier work, selection bias of this sort is minimal. In addition, we are aided by the fact that, in LAUSD, school administrators do not observe applicants' overall or element screening scores; site administrators only know whether or not a teacher has achieved a score over the eligibility cut point of 80 out of 100. This means that site administrators cannot select for the highest scoring teachers, or for the highest scoring teachers on a specific element of the MMTSP. Moreover, individual teacher applicants do not know their screening scores, and therefore similarly cannot attempt to match themselves into specific school environments based on screening score or sub-scores.

This aspect of the LAUSD process, then, enables us to assess whether the underlying score provides a true signal of later teacher quality, measured by various outcomes. Of course, the MMTSP scores may serve as proxies for characteristics administrators can observe in one way or another in site-based interviews, but this is in itself telling about the signal provided by the MMTSP. As such, our results are important for theoretical and practical reasons because they help to illuminate principals' revealed hiring preferences and because in hiring it is often prediction (e.g., of employee outcomes), rather than causation, that is of interest.

To evaluate whether information collected during screening could be used more effectively (RQ 2) we first ask the simple question: Are teachers with higher screening scores more likely to be employed in LAUSD? If the answer is "yes," even without hiring administrators' (principals') knowledge of applicants' actual scores, then the implication is that the screening scores, and potentially sub-scores, could be proactively used in employment decisions. We estimate a series of linear probability models to assess the probability that a teacher is hired by the district as a function of their screening performance:

$$H_{it} = \alpha_0 + \alpha_1 S_{it} + \sum_{c=1}^{C} \alpha^c C_i^c + \sum_{q=1}^{Q} \alpha^q Q_i^q + \gamma_t + e_i \qquad (2)$$

Here H is an indicator of whether teacher i was hired from the eligibility list in year t.[23] S again is either teachers' overall screening scores or a vector of sub-scores collected during screening to determine whether those characteristics are individually or jointly predictive of the probability of being hired, as well as an indicator of whether the applicant received a minimum score exception. C_i^c is a set of dummy variables indicating each teacher's subject area certification (e.g., elementary or mathematics), since the teacher labor market is substantially segmented by certification area. Because LAUSD collects applications on a rolling basis throughout the year applicants may face very different labor market conditions depending upon when they apply. This will be true both across years (e.g., because of changes to the budget or rates of retirement) and at different times throughout the year (e.g., because hiring is concentrated over a few months or because the eligibility list varies in the number of teachers it contains). Equation (2) therefore includes both a set of indicators for the quarter

[23] We also estimate hiring probabilities using logistic regression and obtain very similar results. Results are also similar if we estimate a second linear probability model using only those observations predicted to have probabilities between zero and one in the first regression, a procedure suggested by Horrace and Oaxaca (2006) for mitigating bias in linear specifications.

in which the applicant entered the eligibility list (Q_i^q) and a calendar year fixed effect (γ_t). Standard errors are again clustered at the individual level since some applicants appear on the eligibility list multiple times.

Next, we ask if assigning different weights to the various sub-scores might enhance the ability of the MMTSP to better predict the outcomes of interest: teachers' VAMs, absences, mobility and evaluation outcomes. We use simple canonical correlation to consider the potential implications of weighting screening assessments differently.

Our third and final research question asks whether the adoption of the new screening system has improved hiring outcomes in LAUSD. Knowing who gets hired and their likely effectiveness under the new teacher hiring system is suggestive of the effectiveness of the overall hiring system. For example, ideally screening captures teacher characteristics that predict effectiveness and teachers with those desirable characteristics are more likely to be hired. However, a complete understanding of the system's effectiveness requires a counterfactual hiring outcome to which the observed hiring outcomes can be compared. Unfortunately, we do not observe the employment and effectiveness outcomes of applicants unless they are eventually hired by LAUSD, so we cannot know whether the applicants who are hired are more or less effective than those who are not.

To partially circumvent this problem we utilize publicly available school-level data for other schools outside of the LAUSD traditional public school system to generate comparison groups to LAUSD for use in a difference-in-difference (DiD) regression. While we do not observe individual teacher outcomes in non-LAUSD schools, publicly available data – specifically, staff data files from the California Department of Education (CDE) - do contain records for individual teachers linking them to schools and years of experience in their current district. Thus while we cannot observe these teachers' individual outcomes (or track individual teachers over time), using these and other public school-level data made available by the CDE we can observe some aggregate outcomes at their schools (e.g., math and

English/language arts test scores) as well as school-level student demographics from the 2004-5 through the 2016-17 school years. It is therefore possible to estimate the relationship between the presence of newly-hired teachers and those aggregate test score outcomes at schools across California, as well as to estimate how that relationship changes uniquely in LAUSD schools after the adoption of the new hiring system.

For example, one might plausibly assume that a larger share of newly hired teachers is associated with lower student achievement in a school at year-end, perhaps because newly-hired teachers tend to be less experienced and thus less effective, or because there are disruptive effects of turnover as such. However, the quality of those newly-hired teachers likely depends on the process by which they were hired. Thus, if the new hiring system is an improvement over the status quo ante, newly hired teachers should be more effective and the relationship between newly hired teachers and achievement should be attenuated. We test this hypothesis by estimating a difference-in-difference model:

$$score_{sdt} = \beta_1 newteach_{sdt} + \beta_2 (newteach * lausdtps * post)_{sdt} + \beta_3 (newteach * post)_{sdt} + \beta_4 (newteach * lausdtps)_{sdt} + \beta_5 (lausdtps * post)_{sdt} + +\beta_6 post_{sdt} + \beta_7 D_{sdt} + \delta_s + \gamma_t + \mu_{sdt} \tag{3}$$

where *score* is the average math or ELA test score for school *s* in in district *d* in year *t*, standardized (at the school level) across all schools in the state within year. *newteach* is the share of teachers in a school who are new to their district. *lausdtps* is an indicator for LAUSD traditional (i.e., non-charter, so subject to the new MMTSP) public schools (TPSs). We use three comparison groups of schools in different iterations of the DiD: 1) TPSs in the nine next-largest districts in the state (after LAUSD, which is the largest);[24] 2) other TPSs in Los Angeles county but not in LAUSD itself; and 3) charter schools in LAUSD (which are not bound by district hiring policies but are all plausibly subject to the

[24] After LAUSD, the next largest districts (by enrollment) in California in 2015-16 according to the CDE are the unified districts of San Diego, Long Beach, Fresno, Elk Grove, San Francisco, Santa Ana, Capistrano, Corona-Norco, and San Bernadino. LAUSD has four times more students than San Diego.

same, e.g., labor market, forces impacting LAUSD TPSs). Summary statistics for the included LAUSD TPSs, as well as each set of comparison schools, are presented in Table 3.

post indicates the period after the screening reform (i.e., beginning in 2014-15). The coefficient of primary interest is $\hat{\beta}_2$, as it estimates the extent to which the relationship between newly-hired teachers and school-level achievement changes in LAUSD TPSs (and only LAUSD TPSs) after the adoption of the new hiring system. *D* is the same set of student demographic controls included in the models above, δ_s is a set of school fixed effects to control for unobserved, time-invariant heterogeneity between schools, and γ_t is a set of year fixed effects to control for statewide changes over time. μ_{sdt} is an error term. Standard errors are clustered at the higher of the school or district level, as appropriate.

The primary identifying assumption of the DiD approach is that treated and non-treated units – in this case, schools – would have parallel trends in their outcomes in the absence of treatment. In the case of LAUSD during this time period there are at least two major challenges to this assumption. First, it is not possible to rule out the influence of other time-varying factors in LAUSD, such as the adoption of a new collective bargaining agreement with district teachers. We attempt to mitigate this concern by focusing on the presence of newly-hired teachers in particular, who would likely be most sensitive to changes in district hiring policy. Nevertheless, the potential for confounding by other LAUSD policies remains.

Second, while the DiD approach has the advantage of utilizing data from untreated schools and thus helps to rule out contemporaneous regional or statewide changes in school operations and outcomes, estimates will nevertheless be biased if treated and comparison schools have different preexisting outcome trajectories. This is particularly concerning given that, as is shown in Table 3, there are substantial differences between LAUSD and all three comparison groups in terms of the average ELA and math scores and the percentage of new teachers in the district (especially when comparing to charter schools in LAUSD). Below we attempt to test this assumption directly, for example by allowing schools to have their own linear time trends in some specifications, and also attempt to identify schools that may serve as appropriate comparisons. Unfortunately, no single district is obviously comparable to LAUSD, which is not only the second largest school district in the country by enrollment but also more than four times larger than the next largest district in California. We therefore utilize the three comparison groups of schools described above to see if estimates are sensitive to the choice of counterfactual.

Results

RQ 1: Which Applicant Characteristics are Predictive of Effectiveness after Hire?

VAMs. We first examine which applicant characteristics predict various measures of teacher effectiveness. Results considering teachers' VAM outcomes are provided in columns 1 through 6 of Table 4. We find that higher overall screening performance is associated with larger teacher contributions to student achievement in regularly-tested subjects; a one standard deviation increase in overall screening score is associated with teacher-level VAMs that are 16 (10) percent of a standard deviation higher in ELA (math). Thus, our result suggest that, as in other contexts (e.g., Boyd et al., 2008), while newly-hired teachers are less effective on average than other, likely more experienced teachers in the district, the MMTSP screening is able to discern variation in effectiveness prior to hire. This variation is meaningful; back-of-the-envelope estimates suggest that an applicant with the minimum passing screening score of 80 would require roughly a year to be as effective at raising student achievement in math as an applicant with the average score (roughly 85) and, given diminishing returns to experience, perhaps more than four years to be as effective in ELA.[25,26]

[25] In LAUSD, novice (newly-hired) teachers, on average, are less effective as measured by VAMs; they are below the district average by 34 (42) percent of a standard deviation in ELA (math) in their first year after hire. In addition, there are steep returns to effectiveness in the early years; we observe within-teacher returns to experience between teachers' first and second years in the classroom of approximately four (seven) percent of a standard deviation in ELA (math).

[26] Unlike most of the teacher outcomes discussed below, these VAM estimates are sensitive to the choice of how to control for teachers' school contexts, becoming smaller and insignificant when the school controls are replaced with

Coefficients on individual screening scores are qualitatively similar in both subjects, though mostly smaller in magnitude for math. It may be that the MMTSP, which includes some focus on verbal communication and writing, may simply be less effective at detecting differences in quality (as measured by VAMs) for math teachers than those teaching ELA. In both subject areas, relationships between sub-scores and VAMs appear to be driven at least in part by applicant scores on their sample lesson and whether or not they received background or preparation points, though these relationships are only marginally, if at all, significant in some cases. Since preparation points are frequently rewarded for evidence of prior teaching effectiveness, both they and the sample lesson assessment may serve as relatively direct evidence of teaching skill. No other screening measure is significantly positively associated with teacher VAM, and for several scores the coefficients are in fact slightly negative. This is not because screening components are redundant; recall that correlations between the individual screening assessment scores are generally weak. In results not shown but available upon request we find that coefficients on screening component scores change very little whether they are entered into the model individually or simultaneously.

Two other potential indicators of teacher effectiveness do not appear to add additional value. Teachers with graduate degrees have VAMs that are indistinguishable from those who have only a BA, consistent with most of the previous evidence discussed above.[27] Additionally, applicants teaching in the district despite failing to meet a minimum screening score requirement might be expected to be more effective once hired because they will tend to have been actively identified by either school administrators or HR staff for an exemption. However, we find no evidence that this is the case with

a school fixed effect. However, because VAMs can often be estimated only for one teacher per school, including a school fixed effect substantially reduces the usable variation among new teachers, dropping the number of schools (teachers) from which we can estimate screening score relationships by at least 78 percent (59 percent) in both math and ELA. Potential issues of teacher sorting, including on student achievement and growth, are considered further below.

[27] The presence of few newly-hired teachers with doctoral degrees makes estimating results separately for masters and doctoral degrees infeasible.

respect to teachers' contributions to student achievement; coefficients on the dummy variable indicating exceptions are insignificant after controlling for the fact that that these applicants tend to have lower screening scores on average. We do not observe why these individuals received an exception and so cannot offer much in the way of additional interpretation. However, to the extent that such exemptions are particularly discretionary these results are consistent with prior work indicating that subjective or relatively unstructured assessments during job screening are often unreliable predictors of employee performance (e.g., Delli & Vera, 2003).[28]

Evaluations. Results from regressions predicting teachers' summative evaluation ratings are presented in Table 5. We find that higher screening scores are associated with significantly lower odds of a teacher receiving an unsatisfactory final evaluation rating. A one standard-deviation increase in overall screening score is associated with 57 percent lower odds of an unsatisfactory rating among all teachers, and as much as 73 percent lower odds among elementary teachers. Varying sample sizes and baseline odds make coefficients and their significance difficult to compare across types of teachers, but sample lesson performance appears to be particularly predictive of evaluation outcomes for elementary and secondary teachers. Undergraduate GPA scores are especially predictive of evaluation performance for elementary and special education teachers, while subject matter scores are somewhat more predictive for secondary teachers, perhaps reflective of the relative importance of subject matter expertise relative to general academic ability in more academically-specialized classrooms.

In addition to being less likely to receive unsatisfactory final evaluation ratings, applicants with higher screening scores receive higher ratings on average across the focus TLF elements on which they

---

[28] As discussed above, because student-level standardized test results are not available for the 2013-14 school year, VAMs in 2014-15 are estimated controlling for students' achievement from two years prior (i.e., in 2012-13). When samples are restricted to teachers in the 2015-16 school year coefficients are qualitatively similar both for overall scores and screening sub-scores, though generally larger in magnitude for ELA, smaller in magnitude for math, and estimated less precisely due to sample size limitations. The acquisition of new data from the 2016-17 school year should allow for new, more precise estimates that rely on weaker assumptions about VAM bias. Details about how VAMs are constructed are available from the authors.

are evaluated, with a standard deviation increase in screening score associated with an average score on these EDST ratings that is 0.05 points higher on the one-to-three-point scale we use here. This amounts to a difference of 14 percent of a standard deviation. The individual screening assessment sub-scores that are predictive of overall evaluation ratings are, for the most part, the same as those that predict the average EDST rating, though subject matter scores, which are marginally significantly predictive of overall evaluation outcomes, do not predict average EDST scores. Again, applicants who received a score exception or who possess a graduate degree are generally not significantly more likely to receive a better evaluation rating.[29]

Attendance. Columns 7 through 15 of Table 4 provide results about the relationships between screening score and teachers' attendance. We find that screening performance is also predictive of teachers' attendance, with a standard deviation increase in overall screening score associated with an increase in the share of potential (i.e., contracted) hours for which a teacher is actually present at work of 0.3 percentage points, or an additional 3.3 hours of work in 182-day work year (slightly over one-half of a contracted 6-hour work day). Here again the receipt of preparation points is substantially predictive, but so too are subject matter, GPA, and, especially, professional reference scores. An applicant with a one standard deviation increase in professional reference score will be absent 8.8 fewer hours in a given year, or one and one-half days less than the average new teacher. These results may indicate that these assessments, and especially references, capture aspects of teachers' conscientiousness or work ethic, though we cannot rule out other possibilities (e.g., that they are proxies for applicants' physical health). Other screening measures are not predictive of attendance rates.

Recall that district administrative records distinguish hours for which an employee is absent for legally-protected reasons from those that are unprotected. Tellingly, no aspect of screening

[29] In results available upon request overall screening performance is also shown to be predictive of lower odds of unsatisfactory performance on each of the 15 teaching standards by which teachers are mostly commonly evaluated, usually significantly so.

performance is predictive of teachers' protected absences, but screening performance is significantly related to unprotected absences; a standard-deviation increase in overall screening performance is associated with a decrease in unprotected absences of more than three hours, or just over half of the six-hour contracted teacher workday in LAUSD, in essence almost entirely driving the overall absence rate results. The individual screening scores that are predictive of attendance rate are also predictive of unprotected absences, especially the professional reference sub-score; a one standard deviation increase in that score is associated with a teacher taking 7.76 fewer hours of unprotected absences in a year. Interestingly, being granted the "preparation points" is associated with four fewer hours of unprotected absences a year, and having a higher subject matter score is associated with 1.74 fewer hours of unprotected absences a year. Given that unprotected absences, but not protected absences, are to a large extent discretionary from a teacher's point of view, this pattern of results is at least consistent with the idea that the screening process is discerning real features of applicant quality. This is perhaps also consistent with previous research indicating that screening applicants for their attitudes toward work (e.g., using professional references) can help to select workers who are less likely to shirk (e.g., Huang & Cappelli, 2010). As was the case with teacher VAMs, employees receiving minimum screening score exceptions have attendance that is at best no better than that of other new hires with similar scores and possession of a graduate degree is in general not significantly predictive of teacher attendance; signs on these coefficients indicate if anything lower attendance.

Retention. Table 6 provides results from a multinomial logistic regression predicting teachers' propensities to switch schools within the district or exit the district altogether, relative to staying in the same school. Screening performance is not significantly predictive of teachers' mobility; only writing scores and the receipt of preparation points are predictive of remaining in the same school or district, respectively, between one year and the next. These results are somewhat surprising in light of earlier work by Goldhaber et al. (2017), which found that in Spokane teacher screening ratings predicted teachers' propensities to remain in their schools. In LAUSD we find that teachers with one standard deviation higher overall ratings appear to have lower odds of switching schools and leaving the district (by seven and ten percent, respectively, or by just over half a percentage point each on average), but these results are not statistically significantly different from zero.

Teacher sorting. As discussed above, a primary concern when interpreting these results is that teachers will tend to sort into schools and classrooms in ways that are related to our measures of effectiveness and also predictable based on their screening performance. For example, if applicants who perform well during screening are especially likely to be hired into schools where teachers are evaluated leniently, this will tend to create a relationship between screening performance and teacher outcomes that is not driven by the validity of the screening assessments. That is, it may be that screening scores are more valid measures of new teachers' placements than they are of teacher quality *per se*.

We run several robustness tests to assess the magnitude of new teacher sorting. First, we examine whether or not novice teachers' screening scores are associated with their employment in schools in the top or bottom quartiles of various measures of school context: non-Asian minority students; free- or reduced-price lunch students; English Language Learners; Special Education students; and schools' prior ELA and math achievement and growth. Results are shown in Appendix Table 3. We find only very limited evidence of sorting. Overall screening score is associated with a decreased likelihood of working in high minority schools and an increased likelihood of working in schools in the top quartile of prior year ELA achievement. However, screening score is also associated with an increased probability of working in schools with the most English Language Learners, and there are no statistically significant differences in the estimated relationships between screening score and propensity to work in schools with the highest or lowest proportions of students in poverty, special education students, or previous year ELA achievement growth or math achievement level or growth.[30,31]

Next, we examine teacher sorting through applicants expressed interest in working in each of the district's six regional school districts. We find that teachers with higher screening scores are slightly less likely to indicate that they are interested in working in the district's east or south regions but otherwise do not express geographic preferences that are significantly different from those of other applicants. Given that schools in these two local districts tend to be considered the "hardest to staff" (i.e., hosting more of the low-income, minority and EL populations in the district), these results may suggest some initial bias on the part of new teacher applicants away from difficult schooling contexts. It is important to note, however, that these preferences are expressed optionally and are non-binding. They are at most suggestive of modest sorting of teachers on the basis of screening performance.

Second, we estimate relationships between overall screening performance and teacher outcomes replacing the school-level controls with a school fixed effect (available upon request). In all regressions including fixed effects, sample sizes are substantially diminished. Nonetheless, with the exception of VAMs, discussed above, estimates are not strongly sensitive to this choice though the coefficient on unprotected hours absent shrinks from -3.11 (p = .02) to -2.53 (p = .06) and the coefficient predicting school switching changes direction without gaining significance. Other estimates are essentially unchanged, suggesting that the estimates above are largely robust even within schools,

[30] These models control only for teacher certification area and school year. We do not control for quarter of initial hiring eligibility because these variables are correlated with, and may therefore obscure, sorting on screening scores. However, in practice this makes little difference to the results, available upon request.

[31] As when estimating VAMs, classifying schools based on students' prior achievement and prior achievement growth is complicated by changing testing regimes and missing data in 2013-14. As with the VAM measures, figures in Appendix Table 2 are based on student achievement (or achievement growth) data from two years prior when necessary. If only new teachers in 2015-16 are used (requiring no such double lag) the coefficients estimating sorting on prior achievement shrink in magnitude and, in the case of prior ELA achievement, lose significance. Unfortunately the years under consideration do not allow prior student growth to be estimated without lagging prior achievement twice and comparing scores across testing regimes, though the acquisition of student testing data for 2016-17 should make this possible in the future.

where teachers might be expected to have very similar experiences (e.g., with respect to how they are evaluated or the expectations for attendance).

Thus, while we are unable to definitively rule out the possibility that the relationships observed between screening performance and teacher outcomes are biased by unobserved differences in new teachers' placements, several checks suggest that observable teacher sorting is at most modest. This bolsters the interpretation that screening scores reflect, at least in part, authentic differences in prospective teacher quality and, by implication, that LAUSD's screening system is genuinely sensitive to several aspects of teacher quality.

Summary. In sum, screening performance is predictive of several aspects of teacher effectiveness, including contributions to student test score gains, evaluated performance, and attendance, though not teacher mobility. The magnitudes of these relationships vary, but they are not obviously explicable in terms of teacher sorting and are likely to be practically meaningful. Back of the envelope estimates suggest that had the district replaced every new hire with a score below 85 in 2015-16 with an applicant scoring 85 (and providing no minimum score exceptions), outcomes for newly-hired teachers in LAUSD would improve substantially; teachers in their first year in the district in 2015-16 would have had VAMs that were six and three percent of a standard deviation larger in ELA and math, respectively, would have had collectively 240 fewer unprotected days absent, and would have been 1.6 percentage points less likely to receive an unsatisfactory final evaluation rating (from a baseline probability of approximately 4.7 percent). Of course, raising screening standards in this way may not be practical, especially for hard-to-staff teaching positions, but districts may nevertheless benefit from doing so when possible.[32]

[32] Conversely, given that applicants receiving minimum score exceptions appear to be no more effective than what is indicated by their overall scores, for harder positions to fill it may be preferable to slightly reduce minimum passing requirements rather than providing exemptions for applicants with scores far below passing thresholds.

RQ 2: Could the Information Collected during Screening be Used More Effectively?

Variation in teacher screening performance.

We now turn to if and how screening information might be used more effectively in LAUSD. The evidence above suggests that the screening instruments employed by LAUSD can, to varying degrees, discern prospective teacher quality. However, the extent to which screening instruments impact hiring outcomes depends not only on the validity of the instruments but also on whether the information they discern alters the probability that applicants are eventually hired. To some extent LAUSD's system alters these probabilities mechanically by excluding most lower-scoring applicants from ever reaching the eligibility list. An additional relevant question is whether higher-scoring applicants are also more likely to be hired conditional on being placed on the eligibility list. Table 7 presents estimated relative probabilities that applicants on the eligibility list are eventually hired as teachers in the district. Columns 1-4 present results for all applicants and columns 5-16 provide results for applicants with different certification areas: Elementary, Science, Math, ELA, Social Studies and Special Education.

Application timing and certification. We first report on non-screening characteristics associated with eventual employment in LAUSD. Column 1 provides estimates of relationships between certification area and application timing and employment unconditional on any aspect of screening performance. We find that applicants who enter the eligibility list between October and December are more likely to be hired than those entering at any other time despite their relatively low screening scores (shown in Appendix Table 1), perhaps reflecting a relatively limited teacher supply; only 14 percent of applicants on the eligibility list enter during that period. Recall as well that these applicants have relatively high probabilities of having received minimum score exceptions, perhaps another indicator of the tightness of the labor market during this time. To the extent that score exceptions are granted to circumvent a limited supply of prospective teachers, this may again point to potential advantages of relaxing cut scores in at least some circumstances.

Compared to elementary teachers, special education and P.E. teachers are approximately 10 percentage points more likely, science teachers are approximately five percentage points more likely, and social studies teachers 20 percentage points less likely, to be hired. This is true regardless of whether screening performance is controlled for, and is likely indicative of staffing needs in the district. These results are consistent with earlier work in Washington state (Goldhaber & Theobald, 2013).

Screening performance. Shifting to our relationships of interest, despite not being provided to school administrators, applicant screening scores are meaningfully predictive of subsequent employment in the district as a teacher; a standard deviation increase in overall screening score (1 SD = 5.32 points) is associated with an increase in the probability of being hired of approximately six percentage points. This is true when we estimate the relationship for applicants from all certification types (column 3), and remains positive and almost always statistically significant across the various certification types. The relationship is especially strong for applicants with math and special education certifications. Many of the individual screening scores are also predictive of eventual hire; after controlling for certification area and quarter of first eligibility, applicants are more likely to be hired if they have higher interview scores, sample lesson scores, writing scores, and/or professional reference scores, or if they receive preparation points. Undergraduate GPA and subject matter scores are not predictive, nor is receiving background points.[33] These regression results largely confirm the results from our summary statistics presented in Table 2.

Individuals who are present on the eligibility list despite failing to meet a minimum score requirement are substantially less likely to be subsequently employed as teachers in LAUSD. Conditional on certification area and, importantly, period of first eligibility, these individuals are 21 percentage points less likely to be hired. This difference is more than halved once screening performance is

[33] Results are very similar if GPA is used rather than GPA screening score. Recall that subject matter scores exhibit little variation by construction, and that receipt of background and preparation points can only be received in fixed increments and are thus included in these models using dummy variables.

controlled for, but remains statistically significant.[34] This is perhaps surprising given that these individuals will tend to be present only if they have been actively chosen for an exception by either a school administrator or HR screening specialists, and principals only know whether or not an applicant is eligible for employment in her school, not the applicant's score or if the applicant is eligible because of an exception (unless the principal herself asked for the exception). This suggests that, although principals may make a special request to be able to consider a specific applicant who does not pass the district's screening assessment, principals are still sensitive to the qualities assessed during screening or to warnings from the district office that candidates do not meet standard screening criteria.

Columns five through 16 present results for each of the largest certification areas separately. Overall screening scores are consistently positively associated with hiring probability, and often significantly so, with a one standard deviation increase in hiring score predicting increased probabilities of subsequent employment of anywhere from four percentage points (for elementary and ELA teachers) to eight percentage points (for special education teachers). The relationship is weakest for science teachers, for whom a one standard deviation increase in overall score is associated with only a three percentage point increase in propensity for hire, and the relationship is not significant at traditional levels. The relationship is less precisely estimated for individual screening scores, and in some cases results are quite consistent across certification types but in others they vary. *Ceteris paribus*, measures of academic background and preparation are generally not predictive of subsequent employment; for example, GPA and subject matter scores are not predictive of subsequent employment overall or for any subject area. However, interview performance, sample lesson performance and references are differentially predictive of employment across certification types.

[34] That the magnitude of the coefficient shrinks when conditioned on scores is unsurprising given that scores are predictive of employment and that score exceptions are by definition required only for relatively low scores. Though not shown, we observe a similar pattern within applicant subject areas.

That there is a linear relationship between score and hiring probability overall and across different certification areas is conceptually interesting. Because administrators do not observe these scores they cannot directly influence administrators' hiring decisions. They appear, however, to serve as proxies for characteristics that principals care about, perhaps including communication abilities, personality traits, professionalism, or other teaching skills.

To the extent that differences in the relationship between screening scores and hiring probability across areas of teacher certification reflect genuine variation in demand, they suggest that research into principle hiring preferences should distinguish more carefully between different kinds of teachers.[35] And here again it may be that different screening criteria are appropriate for different kinds of teachers, as the pool of available teachers appears to be substantially tighter in some subject areas than others. More than 80 percent of eligible special education teachers are eventually hired, for example, compared to 75 percent of eligible applicants overall. A tight labor supply might help to explain why, as discussed above, special education teachers on the eligibility list have substantially higher odds than elementary teachers of having received a minimum score exception. Because applicants generally do not compete across certification areas and can have substantially different hiring probabilities once screened, differential performance across screening assessments may suggest that there are gains to be had from differentiating screening criteria across subject areas (e.g., by lowering cut scores for some subject areas) if the alternative is relying more heavily on score exceptions.

Reweighting screening assessments. Because some screening sub-scores are more predictive than others or contribute more to applicants' overall scores, a natural question is whether screening performance could be used differently to better predict these outcomes. A comprehensive analysis of which teacher outcomes to emphasize and how to optimize screening for those outcomes would involve

[35] However, because we observe only final employment outcomes and not job offers we cannot be certain that these hiring patterns reflect administrator preferences.

complicated considerations about individual districts' capacities and priorities, but a simple example can illustrate many of the relevant issues. Specifically, and similar to Goldhaber et al. (2017), we consider four of the outcomes above – ELA VAMs, unprotected hours absent, departure from the district, and the receipt of an unsatisfactory final evaluation rating – and for each outcome use canonical correlation to identify weights for each screening sub-score that maximize the correlation between overall screening scores and that outcome, producing four new (i.e., reweighted to maximize each of the four outcomes) overall scores for each applicant. We then rerun the models above for each of those four outcomes, replacing applicants' original scores with each of the reweighted scores to see how the predictive validity of overall screening scores vary when they have been reweighted to better predict different outcomes.

The coefficients on overall screening score from each regression are presented in columns two through five of Table 8. The first column provides the original coefficients presented above for comparison.[36] Reading from left to right across the table shows how the predictive validity of applicants' overall screening scores for each outcome vary as scores are reweighted to better predict different outcomes. For example, weighting screening scores to better predict ELA VAMs increases the coefficient on the (standardized) overall score by about 12 percent, from 0.16 to 0.18. Looking down the columns, however, shows that this gain comes with a trade-off; reweighting scores in this way (i.e., to predict ELA VAM) reduces the coefficient when predicting unprotected hours absent from the original -3.11 to a much smaller, and now statistically insignificant, -0.43. This is a recurring pattern across all four teacher outcomes considered here; in each case it is possible to increase the magnitude of the coefficient predicting one outcome, in some cases substantially, by reweighting scores to predict that

[36] Canonical correlations are conducted unconditional on school controls or other teacher controls (e.g., graduate degree). This has the implication that the weights produced may not fully maximize the predictive power of overall scores within our sample after adjusting for other variables in the models, but it perhaps better reflects the reality that these other control variables may be difficult for district screening staff to optimize around in practice.

outcome, but in nearly every case doing so reduces the magnitude of the coefficients predicting other outcomes. While the evidence presented here suggests that districts can meaningfully predict teacher effectiveness through careful screening, they may also need to think carefully about what teacher attributes they value most highly and weigh the costs of prioritizing some attributes over others.

Summary. Given that higher performance during screening is predictive of both subsequent employment and subsequent effectiveness, the evidence presented here suggests that hiring in LAUSD under the MMTSP is sensitive to teacher quality. It is possible that there would be gains to differentiating screening requirements by time of year or certification area to respond to differences in labor supply, but we cannot directly test that possibility at present. It is also possible that screening assessments could be utilized differently to better predict outcomes of interest to the district, but this would likely entail trade-offs in the form of reduced ability to predict other, potentially equally important outcomes.

RQ 3: Has the Quality of New Teacher Hires Improved in LAUSD as the New Screening System has been adopted?

Even if LAUSD's new teacher hiring system is sensitive to applicant quality, it is not clear whether it is more effective than LAUSD's previous system. Unfortunately, an evaluation of the hiring reform is hampered by the fact that we do not observe teacher-level hiring outcomes in other districts and thus cannot directly estimate whether these teacher-level outcomes have changed uniquely in LAUSD during this time. In Table 9 we present results from attempts to estimate these outcomes indirectly in a difference-in-difference framework, examining how the relationship between newly-hired teachers and school-level achievement changes uniquely in LAUSD traditional public schools (TPSs) during this time using the three sets of comparison schools described above. For each set of comparison schools both ELA and math achievement are considered, and models in odd (even) columns estimated without (with) school-specific linear time trends.

Coefficients found in the first row in Table 9 indicate that prior to 2014-15 and in all three sets of comparison schools, the presence of teachers newly-hired by the district was associated with lower school-level achievement in both ELA and math. Depending on the specification, comparison group, and test subject, in these schools during this time an additional percentage point of teachers in the school in their first year was associated with a change in achievement of anywhere from 0.000 to -0.008 school-level standard deviations. This is consistent with the notion that newly-hired teachers are either less effective than other teachers (e.g., because they are novices) or are proxies for other circumstances that are detrimental to student achievement (e.g., because their presence is indicative of teacher turnover). As indicated by the coefficients on the interaction terms in rows three and four, this relationship was perhaps more negative in LAUSD TPSs than in the comparison schools, though whether this is indicative of relatively poor hiring processes or something else (e.g., different causes of turnover or hiring across districts) is not clear.

As shown in row two, there is some indication that these relationships changed in the comparison schools beginning in 2014-15, and perhaps changed differently across the set of comparison groups, becoming more positive in some schools, more negative in others, and remaining unchanged in some. What is of primary interest, however, is whether the relationship between newly-hired teachers and achievement changed differently in LAUSD TPS during this time. The coefficients of interest are thus those in the three-way interactions between the share of new-to-the-district teachers in a school, the indicator of the post-reform period, and the indicator of whether the school is in LAUSD or, in the final model, an LAUSD TPS. Across both subject areas and all comparison groups coefficients are positive, and often substantially so, indicating that the relationship between the presence of newly-hired teachers and achievement in LAUSD TPSs has become more positive (or less negative) since the adoption of the new hiring system relative to comparison schools. This is consistent with uniquely improved hiring outcomes in LAUSD during this time as a consequence of the new teacher screening system.

However, as discussed above, the identifying assumption of this approach is that schools in both the treated and untreated groups would have had similar trends in their outcomes in the absence of the hiring reform. As a check on this assumption we run alternative specifications of each model, presented in the even-numbered columns, in which each school is allowed to have its own linear trend in achievement over time. These coefficients become substantially smaller in magnitude, often shrinking to statistical insignificance. Although coefficients are always positive and thus consistent with the hypothesis of improved hiring outcomes in LAUSD we cannot rule out the possibility of unobserved time-varying school factors confounding the observed relationships. And even if changes unique to LAUSD are correctly identified in these models, we cannot rule out the effects of other LAUSD reforms or changes taking place during this time that may be related to schools' achievement trajectories. The evidence is thus perhaps suggestive of improved hiring outcomes, but by no means conclusive.[37]

Discussion & Policy Implications

Despite widespread agreement that teacher quality is important for students and school systems, very little is known about how school districts should hire teachers. This is due in part to the fact that defining teacher quality is difficult and contentious, but also to the fact that observable teacher characteristics are often weak predictors of teacher effectiveness and extant literature provides even less guidance about identifying effective teachers *ex ante* during the hiring process. We contribute to

[37] In results not presented but available upon request we also consider an interrupted time series model using teacher-level data from LAUSD extending as far back as 2007-8 and the VAM, attendance, evaluation, and mobility outcomes considered above for newly-hired teachers. These results are also consistent with somewhat improved hiring outcomes in LAUSD since 2014-15, especially for ELA VAMs and attendance. However, these estimates are imprecise and difficult to interpret not only because of contemporaneous statewide changes in standardized testing regimes but also districtwide changes in teacher evaluation protocols and the fact that we do not observe trends in these outcomes in other districts.

this literature using screening data from the Los Angeles Unified School District and newly-hired teacher outcomes for as many as three years of employment after screening.

LAUSD's new teacher screening assessments appear to accurately discern several aspects of teacher quality. Applicants' overall performance during screening is positively, significantly, and meaningfully associated with their subsequent contributions to student achievement, attendance, and evaluation outcomes, and these relationships do not appear to be driven to a large extent by the differential sorting of teachers into classroom placements, though such factors cannot be definitively ruled out. The district may therefore benefit from its policy of excluding most low-performing applicants from employment eligibility. Even among teachers eligible to be hired, performance during screening is predictive of subsequent employment in the district, suggesting that the screening process may be measuring applicant characteristics that are important to school administrators and that school administrators are sensitive to teacher quality. We also find time series and inter-district evidence consistent with the hypothesis that hiring outcomes have improved uniquely in LAUSD's traditional public schools relative to several sets of comparison schools, though we cannot rule out alternative explanations.

There is important variation in which components of screening are predictive of different teacher outcomes. For example, a sample lesson assessment is meaningfully predictive of teacher effectiveness whether measured in terms of contributions to student achievement or more subjective classroom-observation based evaluation ratings. Professional references are predictive of teacher attendance and evaluation ratings, as are measures of academic and subject matter preparation. Preparation points, though difficult to interpret given their complicated composition, are predictive of teacher VAM and attendance and deserving of further study. Additionally, screening performance is not predictive of teachers' retention in their school or the district. This variation in predictive validity across teacher outcomes points to likely challenges for districts attempting to screen teachers more rigorously;

consistent with prior work finding that teacher quality is not easily measured along a single dimension (e.g., Kraft, forthcoming) we find that selecting teachers more deliberately to achieve one outcome (e.g., VAM) appears to frequently entail trade-offs with respect to other outcomes (e.g., attendance) even among the limited set of outcomes we consider here.

Much remains to be learned about how best to hire teachers, including how to differentiate screening and hiring processes on the basis of subject area, grade level, and labor supply. Further research, in LAUSD and elsewhere, should help to illuminate ways in which these hiring processes can be further improved. For example, there are reasons to think that it may be useful to lower screening requirements during times of year or for specific subject areas when the availability of applicants is low or their performance unusually weak. Additionally, in some cases hiring appears more closely related to aspects of screening performance that do not predict teacher effectiveness (e.g., interview performance) than to those linked to teacher outcomes (e.g., undergraduate GPA). This may indicate ways in which district-level screening can improve hiring outcomes by prioritizing applicant attributes that tend to be underrated at the school level.

Additionally, virtually nothing is known about the longer-term implications of potential teacher screening and hiring reforms, including whether and under what circumstances they produce net improvements to hiring outcomes and whether they have dynamic effects on the quality of prospective teachers entering the labor market. At the same time, we contribute to a small but growing body of literature suggesting that it is possible to collect information about prospective teachers prior to hire that can be used to inform and improve hiring by schools and districts. Given that many administrators appear to have substantial discretion when making hiring decisions, perhaps moreso than after teachers have been hired, and that teacher quality has important consequences for students and schools new teacher screening may prove to be an important lever for improving educational quality in many settings.

References

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, *27*(5), 483–503. https://doi.org/10.1016/j.econedurev.2007.03.002

Associated Press. (2008, June 28). Superintendent: Bad Teachers Hard to Fire. *MSNBC*. Retrieved from http://www.nbcnews.com/id/25430476/ns/us_news-education/t/superintendent-bad-teachers-hard-fire/#.Wk5ZZVWnGpo

Backes, B., Cowan, J., Goldhaber, D., Koedel, C., Miller, L. C., & Xu, Z. (2018). The Common Core Conundrum: To What Extent Should We Worry That Changes to Assessments Will Affect Test-Based Measures of Teacher Performance? *Economics of Education Review*, *62*, 48–65. https://doi.org/10.1016/j.econedurev.2017.10.004

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). 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), 793–818. https://doi.org/10.1002/pam.20377

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103–115. https://doi.org/10.1016/j.jue.2009.05.001

Cannata, M., Rubin, M., Goldring, E., Grissom, J. A., Neumerski, C. M., Drake, T. A., & Schuermann, P. (2017). Using Teacher Effectiveness Data for Information-Rich Hiring. *Educational Administration Quarterly*, *53*(2), 180–222. https://doi.org/10.1177/0013161X16681629

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, *44*(6), 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449–465. https://doi.org/10.1016/j.econedurev.2010.12.010

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, *XLI*(4), 778–820. https://doi.org/10.3368/jhr.XLI.4.778

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682. https://doi.org/10.1016/j.econedurev.2007.10.002

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). Are Teacher Absences Worth Worrying About in the United States? *Education Finance and Policy*, *4*(2), 115–149. https://doi.org/10.1162/edfp.2009.4.2.115

Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing Elements in the Discussion of Teacher Shortages. *Educational Researcher*, *45*(8), 460–462. https://doi.org/10.3102/0013189X16679145

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does Teacher Certification Matter? Evaluating the Evidence. *Educational Evaluation and Policy Analysis*, *23*(1), 57–77. https://doi.org/10.3102/01623737023001057

Delli, D. A., & Vera, E. M. (2003). Psychological and Contextual Influences on the Teacher Selection Interview: A Model for Future Research. *Journal of Personnel Evaluation in Education*, *17*(2), 137–155. https://doi.org/10.1023/B:PEEV.0000032425.42878.f3

Dobbie, W. (2011). *Teacher Characteristics and Student Achievement: Evidence from Teach For America*. Working Paper, Cambridge, MA.

Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, *4*(6), 540–547. https://doi.org/10.1080/17439760903157232

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review*, *13*(1), 1–17. https://doi.org/10.1016/0272-7757(94)90019-1

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, *14*(1), 1–21. https://doi.org/10.1016/0272-7757(94)00031-Z

Engel, M., Jacob, B. A., & Curran, F. C. (2014). New Evidence on Teacher Labor Supply. *American Educational Research Journal*, *51*(1), 36–72. https://doi.org/10.3102/0002831213503031

Feng, L., & Sass, T. R. (2013). What makes special-education teachers special? Teacher training and achievement of students with disabilities. *Economics of Education Review*, *36*, 122–134. https://doi.org/10.1016/j.econedurev.2013.06.006

Goldhaber, D. (2007). Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness? *Journal of Human Resources*, *XLII*(4), 765–794. https://doi.org/10.3368/jhr.XLII.4.765

Goldhaber, D. (2016). In Schools, Teacher Quality Matters Most. *Education Next*, *16*(2), 56–62.

Goldhaber, D., & Brewer, D. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145. https://doi.org/10.3102/01623737022002129

Goldhaber, D., & Brewer, D. J. (1998). When Should We Reward Degrees for Teachers? *The Phi Delta Kappan*, *80*(2), 134–138.

Goldhaber, D., & Brewer, D. J. (2001). Evaluating the Evidence on Teacher Certification: A Rejoinder. *Educational Evaluation and Policy Analysis*, *23*(1), 79–86. https://doi.org/10.3102/01623737023001079

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2017). Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools. *Education Finance and Policy*, 1–58. https://doi.org/10.1162/EDFP_a_00200

Goldhaber, D., & Theobald, R. (2013). Managing the Teacher Workforce in Austere Times: The Determinants and Implications of Teacher Layoffs. *Education Finance and Policy*, *8*(4), 494–527. https://doi.org/10.1162/EDFP_a_00111

Griffith, D., & McDougald, V. (2016). *Undue Process: Why Bad Teachers in Twenty-Five Diverse Districts Rarely Get Fired*. Thomas B. Fordham Institute.

Hanushek, E. A. (1992). The Trade-off between Child Quantity and Quality. *Journal of Political Economy*, *100*(1), 84–117.

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. *The Economic Journal*, *113*(485), F64–F98. https://doi.org/10.1111/1468-0297.00099

Hanushek, E. A., & Rivkin, S. G. (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics*, *4*(1), 131–157. https://doi.org/10.1146/annurev-economics-080511-111001

Hanushek, E. A., Rivkin, S. G., Rothstein, R., & Podgursky, M. (2004). How to Improve the Supply of High-Quality Teachers. *Brookings Papers on Education Policy*, (7), 7–44.

Hanushek, E. A., Rivkin, S. G., & Schiman, J. C. (2016). Dynamic Effects of Teacher Turnover on the Quality of Instruction. *Economics of Education Review*, *55*, 132–148. https://doi.org/10.1016/j.econedurev.2016.08.004

Harris, D. N., & Rutledge, S. A. (2010). Models and Predictors of Teacher Effectiveness: A Comparison of Research about Teaching and Other Occupations. *Teachers College Record*, *112*(3), 914–960.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7–8), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Hock, H., & Isenberg, E. (2017). Methods for Accounting for Co-Teaching in Value-Added Models. *Statistics and Public Policy*, *4*(1), 1–11. https://doi.org/10.1080/2330443X.2016.1265473

Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, *90*(3), 321–327. https://doi.org/10.1016/j.econlet.2005.08.024

Huang, F., & Cappelli, P. (2010). Applicant Screening and Performance-Related Outcomes. *The American Economic Review*, *100*(2), 214–218.

Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2016). *Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools* (Working Paper No. 22054). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22054

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631. https://doi.org/10.1016/j.econedurev.2007.05.005

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180–195. https://doi.org/10.1016/j.econedurev.2015.01.006

Kraft, M. A. (forthcoming). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources*.

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review*, *28*(1), 49–57. https://doi.org/10.1016/j.econedurev.2007.10.007

Ladd, H. F., & Sorensen, L. C. (2016). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. Education Finance and Policy, 12(2), 241–279. https://doi.org/10.1162/EDFP_a_00194

Levin, J., & Quinn, M. (2003). *Missed Opportunities: How We Keep High-Quality Teachers out of Urban Classrooms*. New York, NY: The New Teacher Project. Retrieved from https://eric.ed.gov/?id=ED481608

Liu, E., & Johnson, S. M. (2006). New Teachers' Experiences of Hiring: Late, Rushed, and Information-Poor. *Educational Administration Quarterly*, *42*(3), 324–360. https://doi.org/10.1177/0013161X05282610

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V.-N., & Martinez, J. F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement*, *44*(1), 47–67. https://doi.org/10.1111/j.1745-3984.2007.00026.x

Metzger, S. A., & Wu, M.-J. (2008). Commercial Teacher Selection Instruments: The Validity of Selecting Teachers Through Beliefs, Attitudes, and Values. *Review of Educational Research*, *78*(4), 921–940. https://doi.org/10.3102/0034654308323035

Milanowski, A. T., & Odden, A. R. (2007, April 6). *A New Approach to the Cost of Teacher Turnover*. School Finance Redesign Project Working Paper 13. Retrieved from http://www.crpe.org/publications/new-approach-cost-teacher-turnover

Miller, R. T., Murnane, R. J., & Willett, J. B. (2008). Do Teacher Absences Impact Student Achievement? Longitudinal Evidence From One Urban School District. *Educational Evaluation and Policy Analysis*, *30*(2), 181–200. https://doi.org/10.3102/0162373708318019

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145. https://doi.org/10.1016/0272-7757(94)90003-5

Painter, S. R. (2000). Principals' Perceptions of Barriers to Teacher Dismissal. *Journal of Personnel Evaluation in Education*, *14*(3), 253–264. https://doi.org/10.1023/A:1008144215978

Papay, J. P. (2011). Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates Across Outcome Measures. *American Educational Research Journal*, *48*(1), 163–193. https://doi.org/10.3102/0002831210362589

Papay, J. P., & Kraft, M. A. (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. https://doi.org/10.1016/j.jpubeco.2015.02.008

Rice, J. K. (2003). *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*. Economic Policy Institute, 1660 L Street, N.

Robertson-Kraft, C., & Duckworth, A. (2014). True Grit: Trait-Level Perseverance and Passion for Long-Term Goals Predicts Effectiveness and Retention among Novice Teachers. *Teachers College Record*, *116*(3).

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. American Economic Review, 94(2), 247–252. https://doi.org/10.1257/0002828041302244

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2010). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, *6*(1), 43–74. https://doi.org/10.1162/EDFP_a_00022

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, *50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Ryan, A. M., & Tippins, N. T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management*, *43*(4), 305–318. https://doi.org/10.1002/hrm.20026

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, *50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Sajjadiani, S., Sojourner, A., Kammeyer-Mueller, J., & Mykerezi, E. (2018). *Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover* (Working Paper).

Strunk, K. O. (2014). The role of collective bargaining agreements in the implementation of teacher quality reforms: Perils and possibilities. In F. M. Hess & M. Q. McShane (Eds.), *Teacher Quality 2.0: Will Today's Reforms Hold Back Tomorrow's Schools?* (pp. 155–180). Cambridge, MA: Harvard Education Press.

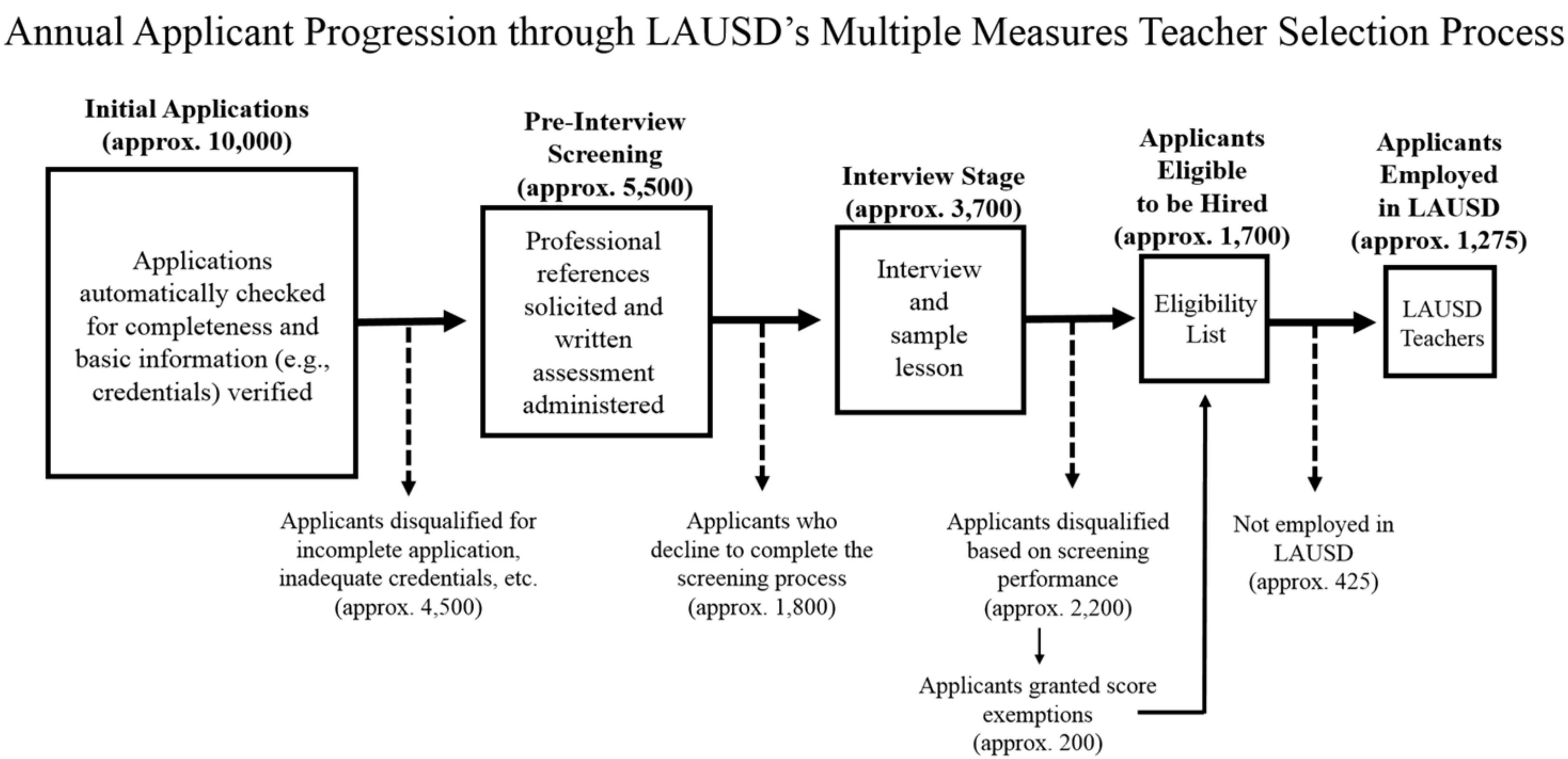
Strunk, K. O., Marsh, J., & Bruno, P. (2017). *An Effective Teacher for Every Student: Developing a Research Agenda to Further Policy Change*. Policy Analysis for California Education.

Wayne, A. J., & Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research*, *73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Wiswall, M. (2013). The dynamics of teacher quality. Journal of Public Economics, 100, 61–78. https://doi.org/10.1016/j.jpubeco.2013.01.006

Figures

Figure 1



Note. Figures are approximate and illustrative.

Tables

Table 1 – Eligibility Criteria for Prospective Teachers in LAUSD

Criterion	Description	Minimum Points Possible	Maximum Points Possible	Minimum Passing Score
Interview	*Structured, conducted by one HR specialist.*	0	25	20
Professional References	*Collected from student teaching or other past professional experience.*	0	20	16
Sample Lesson	*Delivered to and evaluated by two HR specialists.*	0	15	11
Writing Sample	*Timed (45 minutes) responses to hypothetical student-related scenarios.*	1	15	11
GPA	*Scored based on verified undergraduate GPA.*	1	10	N/A
Subject Matter	*Based on subject-matter licensure test scores or, if waived, GPA score.*	8	10	N/A
Background	*For any of: certain prior LAUSD (non-teaching) experience, prior leadership (e.g., military experience), possession of a graduate degree, or Teach for America experience.*	0	2	N/A
Preparation	*For any of: attendance at school highly-ranked by U.S. News & World Report, evidence of prior teaching effectiveness (e.g., student achievement data), or major in credential subject field or, if multi-subject, core academic subject/liberal arts.*	0	3	N/A
Overall		10	100	80

Note. Points are awarded in accordance with criterion-specific rubrics aligned to district goals (e.g., employee evaluation criteria).
Applicants may be placed on the eligibility list despite scoring below the minimum passing score at the request of a school administrator or upon a review of application materials by human resources staff.

Table 2 – Summary Statistics for Newly-Screening Teachers

	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Screening Scores	Applications Resulting in Employment					Applications not Resulting in Employment				
Overall Score	4034	85.01	4.27	46	99	1322	83.07	7.51	17	100
Interview Score	4116	21.53	1.42	0	25	1357	21.12	2.74	0	25
Sample Lesson Score	4117	12.29	1.57	0	15	1356	11.80	2.45	0	15
Writing Score	4111	12.74	1.11	3	15	1349	12.55	1.39	1	15
Reference Score	4115	18.15	1.84	0	20	1344	17.52	3.65	0	20
GPA Score	4091	8.64	1.40	1	10	1344	8.57	1.57	1	10
Subject Matter Score	4090	8.91	0.61	8	10	1349	8.92	0.63	8	10
Received Backgr. Points	4117	0.57	0.49	0	1	1359	0.54	0.50	0	1
Received Prep. Points	4117	0.54	0.50	0	1	1359	0.52	0.50	0	1
Received Score Exception	4117	0.08	0.27	0	1	1359	0.18	0.38	0	1
Certification Area										
Elementary	4117	0.34	0.47	0	1	1359	0.38	0.48	0	1
Science	4117	0.05	0.21	0	1	1359	0.04	0.19	0	1
Math	4117	0.04	0.20	0	1	1359	0.06	0.23	0	1
SPED	4117	0.34	0.48	0	1	1359	0.24	0.43	0	1
ELA	4117	0.09	0.29	0	1	1359	0.08	0.28	0	1
Foreign Language	4117	0.02	0.13	0	1	1359	0.02	0.15	0	1
Social Studies	4117	0.04	0.20	0	1	1359	0.11	0.31	0	1
Arts	4117	0.03	0.17	0	1	1359	0.03	0.17	0	1
P.E.	4117	0.03	0.16	0	1	1359	0.02	0.13	0	1
Multiple Subjects	4117	0.02	0.14	0	1	1359	0.02	0.15	0	1
Quarter Eligible										
Jan-March	4117	0.15	0.36	0	1	1359	0.14	0.35	0	1
April-June	4117	0.28	0.45	0	1	1359	0.35	0.48	0	1
July-September	4117	0.42	0.49	0	1	1359	0.40	0.49	0	1
October-December	4117	0.15	0.36	0	1	1359	0.11	0.31	0	1
Post-Hire Outcomes (2,069 teachers, 2013-14 through 2015-16)										
MA or Doctorate	3270	0.39	0.49	0	1					
ELA VAM	899	-0.25	1.16	-7.4	5.6					
Math VAM	725	-0.32	1.11	-4.5	4.2					
Attendance Rate	3265	97.10	3.89	25	100					
Protected Hours Absent	3265	9.07	43.62	0	557					
Unprotected Hours Absent	3265	31.47	36.73	0	612					
Below Standard Eval.	2819	0.02	0.14	0	1					
Average EDST Rating	2851	2.77	0.29	1	3					
Switch School	3188	0.11	0.31	0	1					
Leave LAUSD	3208	0.06	0.24	0	1					
School Characteristics										
% Non-Asian Minority	3270	0.88	0.19	0	1					
% FRL	3270	0.82	0.20	0	1					
% SPED	3270	0.13	0.09	0	1					
% EL	3270	0.26	0.16	0	.85					
Elementary	3270	0.49	0.50	0	1					
Middle School	3270	0.19	0.39	0	1					
High School	3270	0.27	0.44	0	1					
Other Grade Arrangement	3270	0.06	0.24	0	1					
Local District										
Central	3270	0.20	0.40	0	1					
East	3270	0.13	0.34	0	1					
Northeast	3270	0.16	0.37	0	1					
Northwest	3270	0.14	0.35	0	1					
South	3270	0.16	0.36	0	1					
West	3270	0.20	0.40	0	1					

Table 3 – Summary Statistics for Difference-in-Difference Analysis (2004-5 through 2016-17)

	LAUSD TPSs (705 Unique Schools)					TPSs in Nine Next Largest Districts (777 Unique Schools)					Other LA County TPSs (1,148 Unique Schools)					Charter Schools in LAUSD (345 Unique Schools)				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Average ELA Score	6739	-0.41	0.79	-3.1	2.7	8220	0.01	0.95	-3.9	3.4	11215	0.12	0.99	-3.6	3.3	2418	0.08	0.90	-2.9	2.6
Average Math Score	6665	-0.27	0.91	-2.5	2.9	8218	0.05	0.96	-3.2	3.3	11141	0.16	1.00	-3.4	3.4	2418	-0.00	1.06	-2.5	3.0
%age New to District	6743	4.07	6.22	0	88.7	8227	4.94	7.11	0	100	11241	6.02	8.82	0	100	2419	28.49	27.95	0	100
% EL	6743	0.35	0.19	0	1	8227	0.30	0.22	0	1	11241	0.26	0.18	0	1	2419	0.22	0.19	0	1
% FRL	6743	0.79	0.19	0	1	8227	0.65	0.28	0	1	11241	0.58	0.30	0	1	2419	0.70	0.30	0	1
% Non-Asian Minority	6743	0.88	0.17	0	1	8227	0.65	0.26	0	1	11241	0.70	0.28	0	1	2419	0.78	0.29	0	1
% SPED	6743	0.12	0.06	0	0.89	8227	0.12	0.06	0	1	11241	0.11	0.07	0	1	2419	0.11	0.06	0	0.69

Note. Test scores are standardized at the school level across all schools in the state in a given year.

Table 4 – OLS Regressions Predicting New Teacher VAMs and Attendance

	VAM						Attendance								
	ELA			Math			Percentage Hours Worked			Protected Hours Absent			Unprotected Hours Absent		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Overall Score		0.16*			0.10+			0.30*			0.57			-3.11*	
		(0.06)			(0.06)			(0.13)			(1.23)			(1.31)	
Interview Score			-0.07			-0.01			-0.12			-0.51			1.49
			(0.05)			(0.05)			(0.11)			(1.24)			(1.09)
Sample Lesson			0.17**			0.08			-0.18			0.63			1.38
Score			(0.07)			(0.06)			(0.12)			(1.03)			(1.03)
Writing Score			0.01			0.00			0.00			0.26			0.25
			(0.04)			(0.04)			(0.06)			(0.95)			(0.58)
Professional			-0.06			-0.02			0.81*			-0.01			-7.76+
Reference Score			(0.12)			(0.10)			(0.40)			(1.17)			(4.29)
Undergraduate			0.07			0.03			0.16+			0.58			-1.94*
GPA Score			(0.05)			(0.05)			(0.08)			(0.85)			(0.88)
Subject Matter			0.04			-0.01			0.18*			1.10			-1.74*
Score			(0.05)			(0.05)			(0.09)			(0.87)			(0.87)
Received			0.10			0.16+			-0.01			-2.55			-0.07
Background Points			(0.09)			(0.09)			(0.15)			(1.73)			(1.47)
Received			0.18*			0.14+			0.39*			0.14			-4.06**
Preparation Points			(0.08)			(0.08)			(0.15)			(1.68)			(1.48)
Minimum Score	-0.25	-0.06	0.02	-0.15	0.01	0.05	-0.28	-0.08	-0.17	0.35	-1.01	-0.04	4.39	2.47	3.06
Exception	(0.17)	(0.22)	(0.23)	(0.16)	(0.19)	(0.21)	(0.44)	(0.44)	(0.45)	(3.68)	(3.60)	(3.90)	(4.77)	(4.74)	(4.67)
Graduate Degree	0.02	0.01	0.00	0.02	-0.00	-0.03	-0.21	-0.25	-0.17	1.29	1.72	2.36	1.76	2.17	1.30
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.17)	(0.17)	(0.17)	(1.64)	(1.67)	(1.73)	(1.55)	(1.62)	(1.61)
Years Since Hire (Reference Group = 1)															
2	0.22*	0.17	0.16	0.24*	0.24*	0.20+	-0.24	-0.32	-0.11	5.70*	5.57*	6.14**	2.55	3.17	1.37
	(0.10)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.22)	(0.23)	(0.23)	(2.22)	(2.29)	(2.37)	(2.06)	(2.10)	(2.19)
3	0.59*	0.50*	0.49+	0.44	0.42	0.35	0.06	-0.16	0.12	17.93	11.64	12.63	-1.31	1.00	-1.38
	(0.25)	(0.25)	(0.26)	(0.33)	(0.32)	(0.33)	(0.44)	(0.48)	(0.46)	(12.05)	(11.22)	(11.28)	(4.01)	(4.41)	(4.29)
Constant	1.22+	1.12	0.99	1.98**	2.01**	1.73*	97.50***	97.12***	97.24***	-6.33	-7.13	-4.74	21.09	25.23+	24.23+
	(0.70)	(0.71)	(0.69)	(0.68)	(0.69)	(0.68)	(1.51)	(1.58)	(1.53)	(7.96)	(8.11)	(8.19)	(14.50)	(15.05)	(14.72)
Observations	899	870	872	725	703	706	3265	3168	3191	3265	3168	3191	3265	3168	3191
Teachers	646	626	628	511	495	498	2067	2010	2026	2067	2010	2026	2067	2010	2026
R-sq	0.10	0.11	0.12	0.18	0.20	0.20	0.03	0.04	0.05	0.01	0.01	0.01	0.04	0.04	0.06

Note. Standard errors clustered on teachers in parentheses. Screening scores are standardized to have a standard deviation of one.
All models include year and teacher certification area indicators, school grade level and district region indicators, and the share of students in the school who are non-Asian racial minorities, FRL-eligible, SPED, and English learners.
+ p<.1, * p<.05, ** p<.01, *** p<.001

Table 5 –Regressions Predicting Teacher Evaluation Outcomes

	“Below Standard” Final Evaluation Rating														
	All			Elementary			Secondary			Special Education			Average EDST Rating		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Overall Score		0.43***			0.27**			0.64			0.42**			0.05***	
		(0.09)			(0.14)			(0.27)			(0.13)			(0.01)	
Interview Score			0.86			1.12			1.18			0.87			0.04**
			(0.13)			(0.78)			(0.84)			(0.14)			(0.01)
Sample Lesson Score			0.60**			0.36*			0.40**			0.81			0.04***
			(0.11)			(0.17)			(0.12)			(0.17)			(0.01)
Writing Score			0.94			1.01			1.11			0.80			-0.00
			(0.13)			(0.27)			(0.28)			(0.20)			(0.01)
Professional			0.61**			0.26*			0.66			0.57			0.01
Reference Score			(0.10)			(0.17)			(0.28)			(0.25)			(0.01)
Undergraduate GPA			0.73**			0.55**			1.04			0.71*			0.02**
Score			(0.09)			(0.10)			(0.39)			(0.11)			(0.01)
Subject Matter Score			0.75^{+}			0.74			0.68^{+}			0.69			-0.01
			(0.12)			(0.28)			(0.15)			(0.19)			(0.01)
Received			1.48			0.77			1.58			1.48			0.00
Background Points			(0.48)			(0.44)			(1.12)			(0.66)			(0.01)
Received Preparation			0.71			0.67			0.86			0.54			0.02
Points			(0.22)			(0.39)			(0.54)			(0.26)			(0.01)
Minimum Score	2.95*	1.35	0.76	1.36	0.54	0.31	2.64	2.57	0.24	5.25**	2.28	2.59	-0.09*	-0.04	0.01
Exception	(1.26)	(0.69)	(0.51)	(0.84)	(0.43)	(0.29)	(2.37)	(2.56)	(0.30)	(2.72)	(1.36)	(1.91)	(0.04)	(0.04)	(0.04)
Graduate Degree	1.39	1.48	1.17	0.81	0.74	0.52	3.49*	3.89*	3.15^{+}	1.12	1.22	1.03	0.01	0.01	0.01
	(0.39)	(0.42)	(0.36)	(0.41)	(0.40)	(0.30)	(2.00)	(2.30)	(2.17)	(0.49)	(0.55)	(0.50)	(0.01)	(0.01)	(0.01)
Years Since Hire (Reference Group = 1)															
2	0.63	0.82	0.71	0.50	0.73	0.51	0.41^{+}	0.52	0.35^{+}	1.37	1.53	1.41	0.09***	0.08***	0.08***
	(0.21)	(0.27)	(0.25)	(0.27)	(0.40)	(0.30)	(0.22)	(0.29)	(0.20)	(0.62)	(0.70)	(0.70)	(0.02)	(0.02)	(0.02)
3													0.10	0.12^{+}	0.10
													(0.07)	(0.07)	(0.06)
Constant													2.94***	2.92***	2.90***
													(0.06)	(0.06)	(0.07)
Observations	2748	2668	2688	520	507	508	909	878	888	664	645	653	2851	2766	2786
Teachers	1831	1777	1792	520	507	508	596	578	584	664	645	653	1888	1833	1848
R-sq													0.07	0.09	0.10

Note. Models 1-12 are logistic regressions; coefficients are odds ratios.
Standard errors clustered on teachers in parentheses. Screening scores are standardized to have a standard deviation of one.
All models include year and teacher certification area indicators, school grade level and district region indicators, and the share of students in the school who are non-Asian racial minorities, FRL-eligible, SPED, and English learners.
$^{+}$ p<.1, * p<.05, ** p<.01, *** p<.001

Table 6 – Multinomial Logistic Regressions Predicting Teacher Mobility Outcomes

	Score Exceptions		+ Overall Score		+ Individual Screening Scores	
	Switch School	Leave District	Switch School	Leave District	Switch School	Leave District
Overall Score			0.93	0.90		
			(0.09)	(0.11)		
Interview Score					1.12	1.09
					(0.12)	(0.20)
Sample Lesson Score					0.89	0.94
					(0.07)	(0.10)
Writing Score					0.86*	1.02
					(0.05)	(0.08)
Professional Reference Score					1.13	0.88
					(0.19)	(0.09)
Undergraduate GPA Score					1.09	1.05
					(0.08)	(0.10)
Subject Matter Score					0.95	0.99
					(0.06)	(0.09)
Received Background Points					0.97	1.05
					(0.13)	(0.18)
Received Preparation Points					0.96	0.59**
					(0.12)	(0.10)
Minimum Score Exception	1.33	1.62+	1.47	1.68+	1.31	1.62
	(0.41)	(0.47)	(0.49)	(0.52)	(0.48)	(0.56)
Graduate Degree	0.97	1.12	0.96	1.12	0.98	1.07
	(0.12)	(0.18)	(0.12)	(0.18)	(0.13)	(0.18)
Years Since Hire (Reference Group = 1)						
2	0.93	0.94	1.02	1.04	0.99	0.99
	(0.15)	(0.17)	(0.18)	(0.19)	(0.18)	(0.19)
3	0.46	0.23	0.66	0.31	0.61	0.29
	(0.34)	(0.24)	(0.49)	(0.31)	(0.46)	(0.29)
Observations	3187	3187	3092	3092	3115	3115
Teachers	2055	2055	1999	1999	2015	2015

Note. Coefficients are relative risk ratios relative to the probability of staying in the same school. Standard errors clustered on teachers in parentheses. Screening scores are standardized to have a standard deviation of one. All models include year and teacher certification area indicators, school grade level and district region indicators, and the share of students in the school who are non-Asian racial minorities, FRL-eligible, SPED, and English learners.

+ p<.1, * p<.05, ** p<.01, *** p<.001

Table 7 – Linear Probability Models Predicting Employment as Teacher in LAUSD

	All				Elementary		Science		Math		ELA		Soc. Studies		SPED	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Overall Score			0.06***		0.04**		0.03		0.06*		0.04*		0.05^{+}		0.08***	
			(0.01)		(0.01)		(0.04)		(0.02)		(0.02)		(0.03)		(0.01)	
Interview Score				0.02***		0.03*		-0.06		0.02		-0.01		0.11*		0.02^{+}
				(0.01)		(0.01)		(0.05)		(0.02)		(0.02)		(0.05)		(0.01)
Sample Lesson Score				0.02**		0.01		0.13**		-0.03		0.02		0.04		0.03*
				(0.01)		(0.02)		(0.04)		(0.03)		(0.03)		(0.04)		(0.01)
Writing Score				0.02*		0.01		-0.04		0.05*		0.01		-0.03		0.02*
				(0.01)		(0.01)		(0.04)		(0.02)		(0.02)		(0.03)		(0.01)
Professional Reference Score				0.03***		0.02		0.02		0.03		0.05**		0.06*		0.06***
				(0.01)		(0.01)		(0.03)		(0.02)		(0.02)		(0.03)		(0.01)
Undergraduate GPA Score				0.01		-0.01		0.02		0.02		0.01		0.05		0.01
				(0.01)		(0.01)		(0.03)		(0.03)		(0.02)		(0.03)		(0.01)
Subject Matter Score				-0.00		-0.01		-0.00		-0.04		0.01		-0.03		0.01
				(0.01)		(0.01)		(0.03)		(0.03)		(0.02)		(0.03)		(0.01)
Received Background Points				0.01		0.03		0.01		0.10*		-0.01		0.02		0.02
				(0.01)		(0.02)		(0.06)		(0.05)		(0.04)		(0.06)		(0.02)
Received Preparation Points				0.02^{+}		0.03		-0.01		-0.01		0.01		0.04		0.01
				(0.01)		(0.02)		(0.05)		(0.06)		(0.04)		(0.06)		(0.02)
Minimum Score Exception		-0.21***	-0.09***	-0.07**	-0.06	-0.06	-0.01	0.16	-0.34***	-0.40***	-0.21*	-0.13	-0.18	-0.09	-0.05	-0.03
		(0.02)	(0.02)	(0.03)	(0.05)	(0.05)	(0.12)	(0.13)	(0.10)	(0.11)	(0.09)	(0.09)	(0.11)	(0.14)	(0.04)	(0.04)
Quarter of First Eligibility (Reference Group = July-September)																
Jan-March	0.01	0.01	0.00	0.00	-0.03	-0.03	-0.18	-0.17	0.03	0.01	0.11*	0.12*	0.16^{+}	0.15	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.12)	(0.13)	(0.09)	(0.10)	(0.05)	(0.05)	(0.09)	(0.09)	(0.03)	(0.03)
April-June	-0.06***	-0.06***	-0.06***	-0.06***	-0.10***	-0.10***	-0.08	-0.12^{+}	-0.21***	-0.21**	-0.09^{+}	-0.07	0.13^{+}	0.13^{+}	-0.04^{+}	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)	(0.02)	(0.02)
October-December	0.03^{+}	0.04*	0.05**	0.05**	0.10**	0.09**	-0.12	-0.14	-0.35**	-0.39**	0.12^{+}	0.12^{+}	0.11	0.12	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.14)	(0.14)	(0.12)	(0.12)	(0.06)	(0.07)	(0.09)	(0.09)	(0.03)	(0.03)
Certification Area (Reference Group = Elementary)																
Science	0.05	0.05^{+}	0.05^{+}	0.06^{+}												
	(0.03)	(0.03)	(0.03)	(0.03)												
Math	-0.05^{+}	-0.04	-0.04	-0.04												
	(0.03)	(0.03)	(0.03)	(0.03)												
SPED	0.09***	0.10***	0.11***	0.11***												
	(0.01)	(0.01)	(0.01)	(0.01)												
ELA	0.03	0.03	0.03	0.03												
	(0.02)	(0.02)	(0.02)	(0.02)												
Foreign Language	-0.04	-0.03	-0.03	-0.03												
	(0.05)	(0.04)	(0.05)	(0.05)												
Social Studies	-0.20***	-0.19***	-0.20***	-0.20***												
	(0.03)	(0.03)	(0.03)	(0.03)												
Arts	0.03	0.04	0.04	0.04												
	(0.04)	(0.04)	(0.04)	(0.04)												
P.E.	0.12**	0.12***	0.13***	0.12***												
	(0.04)	(0.03)	(0.03)	(0.03)												
Multiple Subjects	0.00	0.01	0.00	0.01												
	(0.04)	(0.04)	(0.04)	(0.04)												
Constant	0.76***	0.75***	0.73***	0.72***	0.73***	0.70***	0.87***	0.84***	0.82***	0.78***	0.76***	0.77***	0.51***	0.47***	0.81***	0.81***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.10)	(0.11)	(0.09)	(0.11)	(0.07)	(0.08)	(0.10)	(0.12)	(0.04)	(0.05)
Year Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5476	5476	5356	5367	1888	1888	234	234	248	248	469	472	312	312	1703	1710
Individuals	5396	5396	5280	5289	1865	1864	231	231	246	246	462	465	304	303	1693	1700
R-sq	0.04	0.06	0.07	0.06	0.04	0.04	0.05	0.10	0.26	0.27	0.10	0.10	0.08	0.10	0.07	0.06

Note. Standard errors clustered on individuals in parentheses. Screening scores are standardized to have a standard deviation of one.
$^{+}$ p<.1, * p<.05, ** p<.01, *** p<.001

Table 8 – Coefficients Using Reweighted Overall Scores

Predicted Teacher Outcome	Unadjusted	*Screening Scores Weighted to Predict:* ELA VAM	Unprotected Hours Absent	Leave District[a]	Unsatisfactory Final Evaluation[b]
ELA VAM	0.16*	0.18**	0.10	0.15*	0.14*
	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
Unprotected Hours Absent	-3.11*	-0.43	-7.70**	-2.18	-2.59*
	(1.31)	(0.86)	(2.49)	(1.42)	(1.07)
Leave District[a]	0.90	0.91	0.83+	0.75*	0.90
	(0.11)	(0.10)	(0.10)	(0.09)	(0.11)
Unsatisfactory Final Evaluation[b]	0.43***	0.51***	0.61*	0.51***	0.42***
	(0.09)	(0.09)	(0.14)	(0.10)	(0.09)

Note. Standard errors clustered on teachers in parentheses. Each coefficient is from a separate regression predicting the outcome in the left column using overall scores reweighted to better predict the outcome listed on the top row. Screening scores are standardized to have a standard deviation of one. All models are as described in Tables 6 & 7.

[a] Multinomial logistic regressions. Coefficients are relative risk ratios compared to staying in the same school.

[b] Logistic regressions. Coefficients are odds ratios.

+ $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$

Table 9 – Fixed Effect Regressions Predicting School-Level Achievement

	Ten Largest Districts (TPS Only)				Los Angeles County (TPS Only)				LAUSD TPS vs. LAUSD Charters[a]			
	ELA		Math		ELA		Math		ELA		Math	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
%age New Teachers	-0.005**	-0.000	-0.008**	-0.000	-0.001	0.000	-0.003	0.000	-0.002*	-0.001	-0.004***	-0.002*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
%age New x Post-Reform	-0.004	-0.006^{+}	0.005	-0.003	-0.000	0.000	0.006	0.004	0.004***	0.002^{+}	0.008***	0.005**
	(0.003)	(0.003)	(0.008)	(0.006)	(0.003)	(0.002)	(0.005)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
%age New x LAUSD	-0.001	0.001	-0.005*	0.002	-0.004**	-0.000	-0.008**	0.001				
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)				
%age New x TPS									-0.004**	0.000	-0.010***	0.003
									(0.001)	(0.001)	(0.002)	(0.002)
%age New x Post-Reform x LAUSD	0.014**	0.008*	0.020*	0.009	0.012***	0.004	0.022***	0.006				
	(0.004)	(0.003)	(0.007)	(0.006)	(0.003)	(0.002)	(0.005)	(0.004)				
%age New x Post-Reform x TPS									0.006^{+}	0.001	0.017**	0.002
									(0.003)	(0.003)	(0.005)	(0.004)
Post-Reform (2014-15+)	0.013	-0.379***	-0.038	-0.840***	0.078*	3.904***	-0.099^{+}	2.046***	-0.189***	-1.876***	-0.351***	-1.366***
	(0.052)	(0.065)	(0.098)	(0.131)	(0.039)	(0.335)	(0.056)	(0.459)	(0.050)	(0.056)	(0.082)	(0.076)
LAUSD x Post-Reform	-0.249**	-0.251**	-0.396**	-0.334*	-0.256***	-0.235***	-0.320***	-0.330***				
	(0.053)	(0.069)	(0.086)	(0.139)	(0.037)	(0.041)	(0.059)	(0.073)				
Post-Reform x TPS									-0.085	-0.121*	-0.217*	-0.305***
									(0.052)	(0.062)	(0.087)	(0.091)
School & Year Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Time Trends	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	14959	14959	14883	14883	17382	17382	17234	17234	9157	9157	9083	9083
Schools	1482	1482	1481	1481	1850	1850	1847	1847	1050	1050	1050	1050
R-sq	0.91	0.95	0.81	0.91	0.93	0.96	0.86	0.93	0.88	0.94	0.80	0.92

Note. Standard errors clustered on districts in parentheses. Test scores are standardized at the school level across all schools in the state each year. All models include school-level (log) enrollment and shares of students who are English learners, FRL-eligible, non-Asian minorities, or eligible for SPED services.

[a] Standard errors clustered on schools in parentheses.

$^{+}$ p<.1, * p<.05, ** p<.01, *** p<.001

Appendix Table 1 – Regressions Predicting Standardized Screening Scores

	Overall Score (1)	Interview Score (2)	Sample Lesson Score (3)	Writing Score (4)	Reference Score (5)	GPA Score (6)	Subject Matter Score (7)	Background Score (8)	Preparation Score (9)	Received Minimum Score Exception[a] (10)
Certification Area (Reference Group = Elementary)										
Science	0.01	-0.04	-0.17*	-0.00	-0.06	-0.04	0.15*	-0.07	0.30***	1.60+
	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.39)
Math	-0.24***	-0.14*	-0.23***	-0.27***	-0.06	-0.09	-0.23**	-0.06	0.27***	2.29***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.51)
SPED	-0.24***	-0.19***	-0.24***	-0.10**	-0.04	-0.18***	-0.30***	0.22***	-0.11***	1.86***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.22)
ELA	-0.02	0.07	-0.22***	0.13*	-0.13*	0.04	-0.22***	-0.00	0.32***	1.72**
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.31)
Foreign	-0.16	-0.27***	-0.04	-0.27**	-0.34+	0.01	-0.01	0.03	0.37***	1.88*
Language	(0.10)	(0.07)	(0.09)	(0.09)	(0.18)	(0.10)	(0.11)	(0.10)	(0.09)	(0.56)
Social	-0.09	-0.05	-0.30***	-0.03	-0.07	-0.01	-0.14*	-0.07	0.40***	1.58*
Studies	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.34)
Arts	0.08	-0.06	-0.13+	-0.19*	-0.08	0.31***	-0.14	-0.23**	0.68***	1.58+
	(0.08)	(0.05)	(0.07)	(0.08)	(0.10)	(0.05)	(0.09)	(0.08)	(0.07)	(0.41)
P.E.	-0.09	-0.04	0.14	-0.31***	0.06	-0.32**	-0.55***	-0.10	0.35***	1.37
	(0.10)	(0.10)	(0.08)	(0.09)	(0.08)	(0.11)	(0.10)	(0.09)	(0.09)	(0.39)
Multiple	-0.10	0.03	-0.21*	-0.02	-0.26+	0.06	-0.01	0.11	0.24**	2.11**
Subjects	(0.10)	(0.07)	(0.09)	(0.11)	(0.14)	(0.08)	(0.10)	(0.09)	(0.09)	(0.59)
Quarter of First Eligibility (Reference Group = July-September)										
Jan-March	0.13**	0.07+	0.04	0.01	-0.03	-0.04	-0.09*	0.27***	0.23***	1.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.14)
April-June	0.13***	0.13***	0.14***	-0.02	0.03	0.01	0.07*	0.12***	0.07*	0.78*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.09)
October-	-0.24***	-0.21***	-0.20***	-0.17***	-0.05	-0.14**	-0.08+	-0.08+	0.13**	2.05***
December	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.29)
Constant	0.53***	0.36***	0.50***	0.23***	0.11*	0.28***	0.20***	0.29***	-0.03	
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.04)	(0.06)	(0.05)	(0.06)	
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5356	5473	5473	5460	5459	5435	5439	5462	5464	5476
Individuals	5280	5393	5393	5381	5379	5356	5359	5383	5385	5396
R-sq	0.04	0.03	0.03	0.02	0.01	0.02	0.03	0.04	0.06	

Note. Standard errors clustered on individuals in parentheses. Screening scores are standardized to have a standard deviation of one.

[a] Logistic regression. Coefficients are odds ratios.

+ p<.1, * p<.05, ** p<.01, *** p<.001

Appendix Table 2 – Correlations between Screening Scores

	Overall Score	Interview Score	Sample Lesson Score	Writing Score	Reference Score	GPA Score	Subject Matter Score	Background Score	Preparation Score
Overall Score	1.00								
Interview Score	0.58^{***}	1.00							
Sample Lesson Score	0.56^{***}	0.38^{***}	1.00						
Writing Score	0.45^{***}	0.29^{***}	0.23^{***}	1.00					
Reference Score	0.54^{***}	0.07^{***}	0.08^{***}	0.07^{***}	1.00				
GPA Score	0.38^{***}	0.09^{***}	0.09^{***}	0.07^{***}	0.04^{**}	1.00			
Subject Matter Score	0.28^{***}	0.09^{***}	0.08^{***}	0.07^{***}	0.03^{*}	0.35^{***}	1.00		
Background Score	0.17^{***}	0.06^{***}	0.00	0.01	-0.01	-0.08^{***}	-0.05^{***}	1.00	
Preparation Score	0.28^{***}	0.04^{**}	-0.02	-0.04^{**}	0.01	0.01	-0.02^{+}	0.03^{*}	1.00

$^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

Appendix Table 3 – Logistic Regressions Predicting Employment in Schools in the top and bottom Quartiles of Student Demographic Characteristics

	Odds of initial employment after hire in schools in the top or bottom quartile of students based on…															
	Non-Asian Minority		FRL		EL		SPED		Prior ELA Achievement		Prior Math Achievement		Prior ELA Growth		Prior Math Growth	
	Top Quartile (1)	Bottom Quartile (2)	Top Quartile (3)	Bottom Quartile (4)	Top Quartile (5)	Bottom Quartile (6)	Top Quartile (7)	Bottom Quartile (8)	Top Quartile (9)	Bottom Quartile (10)	Top Quartile (11)	Bottom Quartile (12)	Top Quartile (13)	Bottom Quartile (14)	Top Quartile (15)	Bottom Quartile (16)
Overall Score	0.80^{***}	1.22^{**}	0.95	1.05	1.27^{**}	1.07	0.90	1.04	1.17^{*}	0.90^{+}	1.09	0.95	1.08	1.02	1.02	1.04
	(0.05)	(0.09)	(0.06)	(0.07)	(0.09)	(0.07)	(0.06)	(0.08)	(0.09)	(0.06)	(0.08)	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)
Minimum Score Exception	0.93	0.92	1.15	0.75	1.16	1.17	1.10	0.75	1.13	0.86	0.90	1.05	0.84	1.15	0.72	1.09
	(0.19)	(0.22)	(0.23)	(0.18)	(0.28)	(0.25)	(0.22)	(0.20)	(0.27)	(0.19)	(0.22)	(0.23)	(0.21)	(0.28)	(0.19)	(0.25)
Observations	2386	2386	2386	2386	2156	2386	2386	2386	2355	2355	2354	2354	2350	2343	2351	2344

Note. Standard errors in parentheses. Coefficients are odds ratios. Screening scores are standardized to have a standard deviation of one.
All models include dummy variables indicating teacher certification area and school year.
$^{+}$ p<.1, * p<.05, ** p<.01, *** p<.001