Title: Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools

Authors: Dan Goldhaber, Cyrus Grout, Nick Huntington-Klein

Publication Date: April 1, 2017

Journal: Education Finance and Policy 2017; 12 (2): 197–223.
https://doi.org/10.1162/EDFP_a_00200

Peer Review Process: https://direct.mit.edu/edfp/pages/submission-guidelines

Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools

Dan Goldhaber, Cyrus Grout, and Nick Huntington-Klein

<RH>Screen Twice, Cut Once

Dan Goldhaber (Corresponding author)

Center for Education Data & Research

University of Washington Bothell

Seattle, WA 98103

dgoldhab@uw.edu

Cyrus Grout

Center for Education Data & Research

University of Washington Bothell

Seattle, WA 98103

cyrusgrout@gmail.com

Nick Huntington-Klein

Department of Economics

California State University, Fullerton

Fullerton, CA 92831

nhuntington-klein@fullerton.edu

Acknowledgements: We acknowledge support from the Institute of Education Science's Researcher-Practitioner Grant Program (grant number R305H130030) and from the National Center for the Analysis of Longitudinal Data in Education Research (CALDER). CALDER funding is through grant number R305A060018 to the American Institutes for Research from the Institutes of Education Sciences, U.S. Department of Education. This research has benefitted from the helpful input of Heather Hill, Angela Jones, Susanna Loeb, Jonah Rockoff, Mary Templeton, members of Spokane Public Schools' stakeholder group, and seminar participants at CALDER, AEFP, University of California Santa Barbara, the University of Washington, and the University of Southern California. We thank Andrew Katz, Malcolm Wolff, Patricia Martinkova, Roddy Theobald, and Nate Brown for data analyses and management support. We also thank all data providers, including Spokane Public Schools, the Washington Office of the Superintendent of Public Instruction, and the Washington School Information Processing Cooperative. All errors are solely the responsibility of the study's authors, and the views expressed are those of the authors and should not be attributed to their institutions, the study's funders, or the agencies supplying data.

Abstract

Despite their widespread use, there is little academic evidence on whether applicant selection instruments can improve teacher hiring. We examine the relationship between two screening instruments used by Spokane Public Schools to select classroom teachers, and three teacher outcomes: value added, absences, and attrition. We observe all applicants to the district (not only those who are hired), allowing us to estimate sample selection corrected models using random tally errors and variation in the level of competition across job postings as instruments. Ratings on the screening instruments significantly predict value added in math and teacher attrition, but not absences—an increase of one standard deviation in screening scores is associated with an increase of about 0.06 standard deviations of student math achievement, and a decrease in teacher attrition of 3 percentage points. The use of selection instruments may represent an efficient means of improving the quality of the teacher workforce.

<A>1. Introduction

Sophisticated human capital management practices are "essential to coordinate the work of many" (Lazear and Gibbs 2007, p. I), and an important area of focus in the field of personnel economics is how recruitment and hiring practices influence the quality of employees. The question of how to make effective applicant selection decisions is important to most any profession, but is particularly salient for America's public schools. Research over the last decade has documented the vast productivity differences between individual teachers (e.g., Aaronson, Barrow, and Sander 2007; Chamberlain 2013; Chetty, Friedman, and Rockoff 2014b). However, policies targeting teacher improvement among in-service teachers have thus far been challenging to implement successfully at scale (Yoon et al. 2007; Wayne et al. 2008). While school systems often have a good deal of choice amongst applicants, it can be quite costly to remove an ineffective teacher once hired. All of this points to the importance of making good up-front applicant selection decisions.

In this paper we analyze the relationship between two teacher selection rubrics that are used in Spokane Public Schools (SPS) and three teacher outcomes:[1] (1) value-added measures of effectiveness, (2) teacher absence behavior, and (3) the likelihood of attrition.[2] We determine whether the information collected by the rubrics is capable of identifying teachers who go on to perform well. If the rubric is successful, this is evidence that a structured subjective evaluation

[1] Spokane Public Schools is the largest school district in eastern Washington, and the second largest in the state. In 2012, the district included thirty-four elementary, six middle, and five high schools, and employed approximately 1,800 teachers who instructed 28,800 students.

[2] Value-added measures of effectiveness are a direct measure of a teacher's influence on student achievement, and absence and attrition behavior have also been empirically linked to student outcomes (Clotfelter, Ladd, and Vigdor 2011; Herrmann and Rockoff 2012). Moreover, school districts (including SPS) care about absences and higher rates of turnover independent of their relationships to student achievement because they impose costs on school districts, which must hire substitutes for absent teachers and replace teachers who exit a position.

can efficiently identify high-quality applicants. In the case of teachers, this means an opportunity to improve student outcomes and save money without disrupting in-service teachers.

Two aspects of our study are unique. First, unlike previous studies of teacher hiring, we observe all applicants, not just those who are ultimately hired. Furthermore, we observe outcomes for all hired applicants, whether they are hired by SPS or by a different public school district in Washington State. The analysis is not restricted to the subset of applicants who perform well enough to obtain a job offer from the employer. Our ability to see those who reject the offer of employment is also informative. Because teachers in 95 percent of job searches that include an offer from Spokane accept the offer, the difference between an applicant who is hired by SPS and an applicant who is not is largely a decision on the district's part, rather than due to employee preference.

Second, we are able to correct for selection bias that may arise from not being able to observe the outcomes of teachers who are not employed in public schools in Washington. Our selection-corrected estimates exploit the fact that a nontrivial proportion of the summative scores teachers receive on the selection instruments are incorrectly computed (because of procedural oversight or arithmetic mistakes), and the fact that applicants face differing amounts of competition when applying to different jobs. These factors influence the likelihood of being hired but are assumed to be otherwise unrelated to teacher quality.[3] Our use of scoring errors echoes Angrist (1998) who used errors in normalization on a military entrance exam to predict entry into the military.

We find that a one standard deviation increase in the screening scores used by SPS is significantly associated with a three percentage point decrease in teacher attrition, suggesting an

[3] Note that we use the terms "rubric" and "instrument" interchangeably when referring to the screening tools used by SPS during the hiring process.

opportunity for public schools to improve match and save money (Barnes, Crowe, and Schaefer 2007; Ronfeldt, Loeb, and Wyckoff 2013). Direct measures of teaching effectiveness can be calculated for the subsample of teachers who teach subjects with associated statewide exams. Among this group, a one standard deviation increase in screening scores is associated with an increase of 0.06 standard deviations of student math achievement. While our research context limits us to a relatively small sample size, our results imply that these hiring practices may be able to improve later student outcomes (Chamberlain 2013; Chetty, Friedman, and Rockoff 2014b). Correcting for selection into an SPS job does not meaningfully change the findings.

Public school districts often use an ad hoc hiring system relying on unquantified impressions of applicants and objective data that often have an unverified relationship with future performance. The SPS uses a process in which many of the same subjective sources of information (e.g., letters of recommendation and résumés) are coded using standardized screening instruments. The SPS's hiring process is capable of separating high-quality hires from others, and its success is notable because effective predictors of teacher success are often hard to come by. The extent to which the findings in SPS generalize to other districts depends both on the current quality of the selection processes at other districts and the depth of their applicant pools. Given that in general there are two to three times as many teachers trained each year as will be hired (Greenberg, McKee, and Walsh 2013), and that the hiring process often appears to be ad hoc (Ballou 1996), these findings strongly suggest that public schools can more efficiently select high-quality teachers through the use of well-designed applicant selection tools. This stands in contrast to Staiger and Rockoff's (2010) assertion that reliable screening at the hiring stage is unlikely to be an efficient tool for raising student achievement because there is "scant evidence that school districts or principals can effectively separate effective and ineffective teachers when they make hiring decisions" (p. 103).

<A>2. The Importance of Human Capital and Applicant Selection

Teachers can have profound effects on students, and empirical research has demonstrated their influence on students' academic achievement and longer-term life outcomes, such as college-going behavior and labor market earnings.[4] Not surprisingly, the last decade has seen a considerable amount of research and policy attention directed toward interventions that might improve the quality of the teacher workforce.[5]

Little of this research focuses on the choices school systems make in the teacher hiring process. Yet, the potential for improving workforce quality through effective hiring practices is broadly supported by research from the field of personnel economics (Heneman and Judge 2003; Shaw and Lazear 2007) and industrial psychology (see Society for Industrial and Organizational Psychology 2014 for an overview). Studies analyzing a wide variety of employers and employee groups generally find that screening tools based on biographical data (experience and training) improve the process of efficiently identifying the best applicants. A meta-analysis of assessments of education and experience by McDaniel, Schmidt, and Hunter (1988) finds that different types of screening scores have correlations in the range of 0.11 to 0.45 with measures of job performance. While suggestive, these studies are hardly definitive in that they are all based on

[4] See, for instance, Aaronson, Barrow, and Sander (2007), Hanushek and Rivkin (2010), and Goldhaber and Hansen (2013) for estimates of the effects of teachers on student test scores, and Chamberlain (2013) and Chetty, Friedman, and Rockoff (2014a) on the long-term impact of teachers.

[5] See, for instance, Goldhaber and Brewer (2000) and Glazerman, Mayer, and Decker (2006) on alternative routes to certification; Dee and Wyckoff (2013) on performance evaluation and feedback; Hill and Grossman (2013) on new evaluation processes; Springer et al. (2010) and Neal (2011) on performance incentives; Boyd et al. (2009) and Goldhaber and Hansen (2013) on teacher preparation; and Garet et al. (2001, 2011) on professional development.

estimated relationships for employees on the job—those who are not hired are not observed and the reported correlations do not account for sample selection.

Public school districts often have a significant amount of choice among potential teachers.[6] So, how do they select amongst applicants? The *Schools and Staffing Survey* suggests that districts rely on teacher licensure and graduation from a state-approved teacher education institution (USDOE 1997), even though there is little evidence that these credentials are a good proxy for teacher quality (Goldhaber and Brewer 2000; Glazerman, Mayer, and Decker 2006). Harris et al. (2010) find that principals report seeking a mix of teacher characteristics including experience, enthusiasm, pedagogical skills, content knowledge, and organizational fit, but put little weight on indicators of academic proficiency in hiring decisions (consistent with Ballou 1996 and Hinrichs 2014 but contrary to Boyd et al. 2013. Liu and Johnson (2006) report that hiring decisions for new teachers are often late, rushed, and information poor, and this is often due to factors beyond administrators' control, including budgetary timing and procedural rules imposed by collective bargaining agreements.

There is mixed evidence about whether the best applicants are the ones hired under these apparently ad-hoc approaches. A few studies look at whether schools identify the best applicants in terms of a direct measure of teacher effectiveness: value-added estimates of teacher performance. Hanushek et al. (2005) find that schools offering higher levels of compensation or better working environments are more likely to hire teachers with advanced degrees, but they do not find evidence the teachers hired by these schools tend to be more effective.

---

[6] Ingersoll and Perda (2010) report that in 2000, the ratio of all teachers in the supply pipeline to the number of teachers leaving through retirement and attrition was more than two to one. Similarly, Strauss et al. (2000) find evidence of excess supply in Pennsylvania, where 75 percent of districts hiring for various subject areas had at least three applicants per position. In elementary education, mathematics, English, and social studies, there were at least 10 applicants per position.

Staiger and Rockoff (2010) examine a natural experiment that occurred in the Los Angeles Unified School District (LAUSD). In 1997, California provided a financial incentive to limit K-3 class sizes. LAUSD more than doubled its annual hiring of new elementary school teachers but teacher pay did not increase, and the proportion of new hires without teaching credentials increased from 59 percent to 72 percent. Value-added estimates of teacher effectiveness showed no evidence of a decrease, suggesting that prior to the change the district had not been selecting teachers from the top end of the distribution, consistent with Hanushek et al. (2005). Boyd et al. (2011), on the other hand, find that schools tend to hire teachers with higher value-added when selecting from among within-district transfer applicants.

To our knowledge, only a few studies analyze the relationship between teacher applicant information and job performance. Each is limited to assessments of those applicants who are hired, which, like other studies in this vein, prevents them from adjusting for sample selection bias. Young and Delli (2002) study the use of a widely used screening instrument in two school districts—the Gallup Organization's Teacher Perceiver Interview (TPI)—and find it is modestly predictive of teacher performance as measured by principal evaluations.[7] Dobbie (2011) investigates the selection of Teach For America (TFA) teachers and their students' achievements in the first year of teaching. A one standard deviation increase in an index of measures used to select TFA applicants increases student math achievement by about 0.15 standard deviations. The findings for reading achievement were smaller and not statistically significant. Rockoff et al. (2011) examine the extent to which traditional (e.g., degree and major, passage of license exam) and nontraditional information about teacher applicants (e.g., measures of personality) are related

[7] In a subsequent meta-analysis of commercial teacher selection instruments, Metzger and Wu (2008) identify the Young and Delli (2002) analysis as the only published peer-reviewed study of the TPI. The meta-analysis finds a modest relationship between the interview score and a variety of teacher quality indicators, including principal student ratings and exam scores.

to teacher performance measures. Traditional and nontraditional information together explain about 12 percent of the expected variance in teacher effectiveness, compared with about 4 percent using only traditional information.

The above literature suggests there is likely room for improvement in teacher selection processes. This is the first study to analyze the relationship between teacher screening tools and teacher performance using data on both hired and non-hired applicants. Additionally, to the best of our knowledge, this is the first study on the use of screening tools in any hiring context that observes the performance of both hired and non-hired applicants.

<A>3. The Hiring Process and Data

<B>The Hiring Process in Spokane Public Schools

Following the intake of applications, the hiring process in SPS is comprised of three stages following the posting of a job:

(1) 21-point prescreening of potential applicants by Human Resources (HR) hiring officials;

(2) 60-point job-specific screening of applicants by school-level hiring officials; and

(3) In-person interview and hiring decision by school-level hiring officials.

To become eligible to interview for a position, job applicants progress through two stages of screening where they are evaluated entirely on the basis of submitted application materials,. The first stage, “pre-screening”, uses a 21-point rubric with three subcomponents scored from 1 to 6: Experience, Depth of Skills, and Quality of Recommendations (hereafter referred to as “Recommendations”). Scores of 1-2, 3-4, and 5-6 indicate the finding of, respectively, “some”, “satisfactory”, and “strong” evidence that the criterion is an area of strength for the applicant. The Recommendations subcomponent is multiplied by 1.5 in the summative score. Descriptions of what screeners look for when scoring these subcomponents are available in a separate online appendix (table C.1 in Appendix C) that can be accessed on *Education Finance and Policy*'s Web site at www.mitpressjournals.org/efp. The 21-point pre-screening process is conducted by SPS Human Resources personnel and is not associated with application to any particular job position.

The primary use of the 21-point score is to narrow the applicant pool to a manageable size. Following the intake of applications associated with a particular job posting, a school principal will request a list of applicants from HR based on a minimum cutoff on prescreening scores (e.g., all applicants scoring 17 or higher) and satisfying other qualifications (e.g., holding an endorsement in a particular subject area). Teachers who satisfy the principal's criteria are advanced to the second stage of the hiring process.

The second stage of the hiring process is used to select the candidates who will receive in-person interviews. This job-level screen uses a 60-point rubric with ten evaluation criteria: Certificate & Education, Training, Experience, Classroom Management, Flexibility, Instructional Skills, Interpersonal Skills, Cultural Competency, Preferred Qualifications, and Letters of Recommendation scored on the same 1 to 6 scale as the 21-point rubric (see descriptions of the subcomponents on the 60-point screening form in table C.1, which is available on the *Education Finance and Policy* Web site). Scoring of 60-point rubrics takes place at the school level and is typically conducted by the principal, but may involve multiple screeners.[8] The 60-point screening score is not only a general appraisal of teacher qualifications, as in the 21-point screening score, but also conveys the screener's impression of the quality of the match between applicant and position.

[8] See Martinkova and Goldhaber (2015) for information on the reliability of this instrument.

The highest scoring applicants are granted in-person interviews with the school. The content of these interviews are at the discretion of the principal and are not standardized. Following the interview process, the school can make a job offer to the applicant. Further detail about the hiring process is available in Appendix A on *Education Finance and Policy*'s Web site.

Structurally, the hiring process established by SPS appears to be typical of the procedures employed by other public school districts. As reported by Liu and Johnson (2006), school districts typically handle early hiring activities such as initial screening and verifying credentials centrally, with final hiring decisions being made by school-level administrators.

<B>Data

We study the pool of applicants during the 2009-12 hiring years to all certificated classroom teaching positions for which SPS used the hiring process outlined above to select an applicant. In total, 521 job postings fit this description and 2,668 applicants applied to one or more of these jobs.[9] We link data on SPS applicants to statewide teacher datasets using unique teacher certification numbers.[10] These data include licensure exam scores and subject-area endorsements provided by the Office of the Superintendent of Public Instruction (OSPI), and teacher absence data for teachers who do not work in Spokane collected by the Washington School Information Processing Cooperative (WSIPC).[11] We also link applicants to the S-275 personnel report, which

[9] One additional outlying applicant is dropped. This teacher received an erroneous 60-point screening score of 66 and was a very effective teacher. In certain subsample analyses the teacher is enough of an outlier to significantly distort estimates. In the main analysis, dropping this teacher reduces the estimated validity of the screening score slightly.

[10] Of the 2,668 unique applicants in our study sample, 2,243 have certification numbers. A number of applicants did not report a certification number in their application or had out-of-state or provisional certification. Applicants without certification numbers cannot be linked to the administrative databases that include outcomes data and so are dropped from the analysis.

[11] Absences are defined as days taken due to illness, which excludes other categories such as personal day or bereavement. The WSIPC provides record-keeping services for 271 of 295 school districts in Washington State, mostly the smaller districts in Washington. Larger school districts typically maintain in-house personnel records.

includes information on demographics, experience levels, contracts, and building assignments for all certificated employees of public school districts in Washington State. Data on school characteristics come from Public School Universe data generated by the National Center for Education Statistics (NCES). Student data come from the statewide Core Student Record System (CSRS). The CSRS dataset includes information on student demographics and the annual state math and reading test scores for students in grades 3 through 8.

These data allow us to associate teacher screening scores with student and school characteristics, as well as teacher performance. Summary statistics for the screening scores are presented in table 1.[12] We perform all analyses at the applicant-year level, averaging screening scores over multiple applications if the teacher applied to multiple jobs in the same year. Many of the 2,668 unique individuals applied to SPS in more than one year, generating a total of 4,215 unique applicant/year combinations between 2009 and 2012. Ten percent of the applicants (88 percent of whom were internal transfers) were not given a 21-point screening score, resulting in 3,946 unique applicant-year combinations screened at the 21-point stage.

Some screening score subcomponents have missing observations. Observations appear to be missing for random reasons, such as coders of 60-point scores omitting a criterion for all applicants to a particular job, or not entering individual subcomponent scores on the 21-point rubric into the computer. We adjust the data to make screening scores more comparable across jobs and schools. The process of adjustment is described in detail in Appendix B (available on the *Education Finance and Policy* Web site). Both unadjusted and adjusted scores are in table 1, and analysis is robust to the use of either unadjusted or adjusted scores.

[12] Correlations between unadjusted screening score components are in table C.2 on the *Education Finance and Policy* Web site.

Three measures are used to evaluate teacher outcomes: (1) grades 3 through 8 student performance on Washington State's annual assessments of student learning for math and reading, (2) teacher absences in 2012 and 2013, and (3) teacher retention.[13] Student performance is the most limited of these measures, as it applies only to teachers in subjects and grades with the appropriate state exams. We are able to match 301 applicant-year observations to student test score data, and 374 applicant-year observations to absence data. Teacher retention in the district is determined by matching 1,024 applicant-year observations to the S-275 personnel records for the school years ending between 2010 and 2013.[14] Observed attrition is censored at 2013.

Descriptive statistics of applicant data and teacher outcomes based on the furthest stage of the hiring process reached are presented in table 2. Of the 4,215 applicant-year combinations (which include many teachers who are never observed as employed and so do not have associated outcome data), 3,946 (or 94 percent of the total applicant/years) are pre-screened by HR using the 21-point rubric, 1,707 (40 percent) are advanced for consideration at the school level where they are scored on the 60-point rubric, 1,236 are interviewed (29 percent), and 538 (13 percent) are hired into or offered a new job in SPS. Nearly all (95 percent) of those offered a position accepted it or another position in SPS that year. Four hundred ninety-six applicant/years (12 percent) are identified as being employed in a certificated teaching position in a different district in Washington State by October of the same year they applied to SPS, either by obtaining a new job or staying in a currently held position. Including these teachers, 32 percent of

[13] Teacher outcome data are linked to the most recent screening scores. Consider a teacher employed continuously in Washington State from 2009 to 2012. If that teacher applied to SPS in both 2009 and 2010, then the 2011 and 2012 teacher performance outcomes are linked to the 2010 application, the 2010 outcomes are linked to the 2009 application, and the 2009 outcomes are not used.

[14] For all three outcomes, some matched teachers were not screened at all levels, and so clustered observation numbers in the main analyses are lower than these figures. For absences and attrition, nonclustered observation numbers are higher than these figures because multiple years of outcome data are associated with each applicant-year observation.

applicants are associated with a certificated classroom teaching position in Washington by October of the next school year, either with a new job for which they applied or one they held the year prior.

Teachers who have previous experience with SPS are more likely to progress through the hiring pipeline. Teachers who have previously worked or did student teaching in Spokane make up 11 and 36 percent of applications each year, respectively, but make up 43 and 47 percent of those who are hired. In total, about 71 percent of hired teachers had some previous SPS experience as a teacher, a student teacher, or both. This advantage is due in part because an agreement between the district and the teachers' union ensures preferential treatment for transfers, and partly due to a general preference for familiarity.

Outcomes are observed for a subset of applicants, including those hired by SPS and those employed by a different school district.[15] The summary statistics in table 2 suggest that SPS's hiring process is, on average, effective at selecting high-quality teachers. Average value-added scores generally increase as the application pool narrows.[16] Average annual absences are fairly stable across the stages of the hiring process, and hired applicants average slightly more absences. The proportion of teachers observed attriting within one year is quite stable through the hiring pipeline, but those who are hired tend to attrit less often. SPS applicants perform slightly below the state average on the state's licensure exam (WESTB), but average scores increase throughout the hiring process.[17]

---

[15] Applicants who are hired by SPS progress through every stage of the hiring process. As such, differences in average outcomes across each stage of the hiring process are driven by the outcomes of the teachers who are hired by other school districts.

[16] The value added of individual teachers is estimated based on a derivative of equation 1 in the next section. Estimates are then "shrunk" using empirical Bayes methods, rather than using the weighting described for equation 2, because empirical Bayes is more applicable outside a regression context.

[17] For those not hired by SPS nor another district, value-added scores can be observed if the un-hired applicant stays employed in a previously-held position in SPS.

<A>4. Empirical Methods

Our analysis investigates the extent to which the SPS screening instruments are predictive of student achievement, teacher absences, and teacher retention. We describe below analytic models for these primary outcomes. In each case, we run a regression of the outcome of interest on control variables and a measure of the screening scores $SCREEN_{j(tprior)}$, which is the average screening score given to teacher j during his or her most recent application year, denoted $tprior$.[18] $SCREEN_{j(tprior)}$ can include the average 21-point score (Specification 1) in year $tprior$, the average 60-point score (Specification 2) in year $tprior$, both the 21- and 60-point scores (Specification 3),[19] or the average score on one of the subcomponents of the 21- or 60-point rubric (Specification 4). In Specification 4, the effect of each subcomponent is estimated separately to avoid issues of collinearity between them. Below, we describe the models for student achievement, teacher absences, and teacher attrition.

<B>Primary Outcome Models

<C>Student Achievement

We use a two-step model to assess the relationship between teacher scores on the screening instruments and student achievement. In the first step we estimate a student achievement model, from which we draw teacher value added. Then we estimate teacher value added as a function of screening scores, taking into account the uncertainty of the estimates from the first step. As is standard in the value-added literature, this process restricts the analysis to the portion of teachers

[18] For instance, we may observe student outcome data for teacher *j* in 2013. If that teacher applied to positions in 2010 and 2011, we would use the average 21-point screening score and the average 60-point screening score from 2011 to inform the model.

[19] We also tried a specification in which we used factor analysis on the thirteen 21-point and 60-point components and set $SCREEN_{j(tprior)}$ to be these factors. However, the factors were almost exactly the same as the summative scores, and results were very similar to those in Specification 3.

who teach subjects with associated annual standard exams, that is, those teachers who teach reading or math (or both).

We use a two-step process, rather than including the screening score in the student achievement model in a one-step process, so that the coefficient on the screening score is not affected by the relationship between screening scores and student assignments:[20]

$$Y_{ijsgt} = \alpha_j + \alpha_1 Y_{i(g-1)(t-1)} + \alpha_2 X_{igt} + \varepsilon_{ijst}^{\alpha} \quad (1)$$

$$\hat{\alpha}_j = \alpha_0' + \alpha_1' SCREEN_{j(tprior)} + \alpha_g' + \alpha_t' + \alpha_y' + \varepsilon_{jst}^{\alpha'} \quad (2)$$

In equation 1, Y_{ijsgt} is the test score for each student *i* in class with teacher *j* in subject *s* (math or reading), grade *g*, and year *t*, normalized within grade, year, and subject; $Y_{i(g-1)(t-1)}$ is a vector of the student's scores in the previous grade and year in both math and reading, also normalized within grade, year, and subject; X_{igt} is a vector of student attributes in grade *g* and year *t* (gender, race, eligibility for free or reduced-price lunch, English language learner status, gifted status, special education status, learning disability status, migrant status, and homeless status). Equation 1 is estimated on the full sample of Washington Teachers for whom student test score data are available. The value-added score for teacher j is estimated as a teacher fixed effect, represented by the coefficient α_j.[21,22]

[20] The primary model results, described below, are qualitatively similar using a one-step process. However, results differ when dividing the sample by teachers who work in SPS against those who work elsewhere. This suggests that the relationship between teacher quality and students' assigned characteristics differs in and out of the district.
[21] We do not include teacher experience because experience is one of the categories used to determine an applicant's screening score.
[22] The specification of models used to estimate value-added measures of teacher performance has been extensively discussed in the academic literature; evidence suggests that teacher effects derived from models that control for prior tests (as in equation 1) tend to be highly correlated with one another (Goldhaber, Gabele, and Walch 2014), and experimental (e.g., Kane and Staiger 2008; Bacher-Hicks, Kane, and Staiger 2014) and quasi-experimental evidence (e.g., Chetty, Friedman, and Rockoff 2014b) suggests they have little to no bias. See Koedel, Mihaly, and Rockoff (2015) for a more extensive discussion.

Teacher value-added estimates $\hat{\alpha}_j$ are the dependent variable in equation 2. Since $\hat{\alpha}_j$ are estimated values, each observation in equation 2 is weighted by the inverse of the standard deviation of $\hat{\alpha}_j$. In the second step, we include indicators for grade α'_g and year α'_t, and indicators α'_y for the number of years (1, 2 or 3, but typically 1) between the hiring year $tprior$ and the year t in which performance data are observed, or the "gap". Standard errors are clustered at the teacher level. The coefficient of interest in equation 2 is α'_1, which represents the expected change in teacher effectiveness (and thus the average standardized exam score of her students) associated with a one standard deviation increase in that teacher's screening score.

We do not include school fixed effects in our primary model specification because the inclusion of school effects limits the precision of the estimates and masks differences in teacher effectiveness associated with the schools in which teachers are hired. However, one concern with this choice is the possibility that teachers who receive higher screening scores tend to sort into schools that lead to increased teacher performance (e.g., because of an effective principal). To the extent that location in better schools is driving the performance of these teachers (as measured by value added), a positive relationship between $\hat{\alpha}_j$ and $SCREEN_j$ could be biased upward. As a robustness check, we also estimate the student achievement models with school fixed effects. The results are less precise, though not substantively different from those obtained in the primary model specification, consistent with Chetty, Friedman, and Rockoff (2014a) who find that sorting is a minimal source of bias in value-added model specifications that control for prior student-level test scores.

<C>Teacher Absences

To assess the relationship between teacher scores on the screening instruments and teacher absences, we estimate the following teacher-level model using ordinary least squares (OLS) regressions with year and gap indicators, and standard errors clustered at the teacher/hiring year level:

$$A_{jt} = \beta_0 + \beta_1 T_j + \beta_2 S_{kt} + \beta_3 SCREEN_{j(tprior)} + \beta_t + \beta_y + \varepsilon_{jt}^{\beta} \tag{3}$$

where A_{jt} is the number of sick days taken by teacher j in year t; T_j is a vector of teacher characteristics (gender and race); S_{kt} is a vector of school characteristics in year t (size, student demographics, urbanicity, and level); and $SCREEN_{j(tprior)}$ is the screening score for teacher j in year *tprior*. The coefficient of interest in equation 3, β_3, represents the expected change in the number of annual teacher absences associated with a one standard deviation increase in the teacher's screening score. Standard errors are clustered at the teacher/hiring year level.

<C>Teacher Retention

To assess the relationship between teacher scores on the screening instruments and teacher retention, we estimate logit models predicting that teachers leave the district.[23] Specifically, let $p_{jkdt}(y)$ be the probability that teacher j in year t leaves the district (d) y years after being hired. Then,

$$\log\left(\frac{p_{jt}(y)}{1-p_{jt}(y)}\right) = \gamma_y + \gamma_1 T_{jt} + \gamma_2 S_{kt} + \gamma_3 SCREEN_{j(tprior)} + \gamma_t + \varepsilon_{jt}^{\gamma}. \tag{4}$$

The control variables in equation 4 are the same as those in equation 3, with the addition to the vector of teacher characteristics (T_{jt}) a series of indicators for whether the teacher holds an

[23] Attrition estimates predicting attrition from the school or from the state are reported in table C.4 (available on the *Education Finance and Policy* Web site).

endorsement in a particular subject, offering a partial control for the presence of employment opportunities outside of teaching. The coefficient of interest in equation 4, γ_3, is the expected change in the log-odds of leaving the district associated with a one standard deviation change in the teacher's screening score, all else equal. In analysis, estimates are presented as the mean marginal effect of the screening score on the probability of leaving the district. Standard errors are clustered at the teacher level.

<B>Correction for Sample Selection

It is possible that estimates from the above models are biased by sample selection. We observe teachers in all Washington public schools, and so the likely source of bias is among the teachers who leave the profession or the state entirely, which form 68 percent of the sample. In this section, we do not model selection into the statewide group observed in the main model, but the more specific selection into the SPS position to which the teacher applied. We do this because our excluded variables specifically predict success in SPS' hiring process rather than selection into a teaching position in general. Teachers who applied to SPS and who are hired into Washington State, but not into SPS, are left out of this analysis entirely, so that the selection process being modeled matches the sample in the second stage. The Heckman model identifies the validity among those hired into SPS (see Blundell and Costa Dias 2000). However, validity among the full sample is of interest. As such, if we find no bias, then these results offer support for the main model findings that include those who work elsewhere in Washington.

We address the potential for selection bias using a Heckman selection model (Heckman 1979). We generate two variables that predict whether or not an applicant is likely to be hired but are otherwise uncorrelated with that applicant's performance as a teacher. We identify the Heckman model using exclusion restrictions on these two variables:[24]

(1) A variable indicating whether an applicant was given a 60-point screening score because of the fact that he or she received a favorable tally error on the 21-point screening instrument;[25] and

(2) A measure of the amount of competition faced by an applicant: the average 21-point screening scores of the other applicants for the job.[26]

Tally errors arise from the incorrect hand-marking of 21-point screenings, which occur in 18.8 percent of applications, and lead to the applicant erroneously receiving a 60-point screening score in 3.64 percent of all job applications. These miscalculations may be due to addition mistakes, (about 38 percent of errors), not multiplying the Recommendations criterion by 1.5 (7 percent), or performing the multiplication incorrectly (56 percent).

We then estimate the following teacher effectiveness model:

$$\hat{\alpha}_j|Hired = \alpha'_0 + \alpha'_1 SCREEN_{j(tprior)} + \varepsilon^{\alpha'}_{jst} \tag{5}$$

$$Hired^* = \alpha_{H0} + \alpha_{H1} Z_t + \alpha_{H2} SCREEN_{j(tprior)} + \varepsilon^H_t \tag{6}$$

$$Hired = I(Hired^* \geq 0) \tag{7}$$

[24] We also estimate the model with a measure of competition and errors in the totaling of the 60-point scores and find similar results. However, errors in 60-point scores are not statistically significant in the first stage. In addition, this approach requires the first stage to be limited to those who received 60-point scores, which excludes an important part of the selection process.

[25] In the case of jobs with an explicit score cutoff, this is an indicator that the calculated score is below the cutoff but the written score is above the cutoff. In the case of jobs without an explicit cutoff (200 jobs), this is an indicator that the calculated score is outside the top N scores, but the written score is within the top N, where N is the number of applicants advanced to the next stage. We assume that teachers with missing component scores are without error; results are similar if teachers with missing scores are instead dropped.

[26] The exclusion restriction for this variable relies on the assumption that a rater's scoring of an applicant is unaffected by the quality of the competition, which is made more plausible by the fact that the 21-point scores we use are not job-specific. Still, to address this possibility, we perform a pseudo-Sargan test where we regress teacher performance residuals on the excluded variables. We further describe this test in section 5.

where *Hired** is the propensity for a particular applicant to be hired into SPS on the basis of his or her 21-point screening score and the above-defined excluded variables Z_t. Equations 5 through 7 are estimated as a Heckman selection model using a two-step method (Maddala 1983; Heckman 1979). We similarly estimate Heckman selection models for our other outcomes, absences and attrition, combining equations 3 and 4, in turn, with equations 6 and 7. Results are robust to the use of a linear probability model in the first stage (as in Olsen 1980), an approach that cannot rely on the nonlinearity of the first stage for identification.

Standard errors are corrected for the fact that the inverse Mills ratio is a constructed variable. In addition, we bootstrap the corrected standard errors using 1,000 bootstrap samples to correct for the fact that the first and second stages of analysis are at different levels of observation. In the first stage, each observation is a single job application, but in the second stage each observation is a teacher's value added, absences, or attrition in a particular year.

The selection-corrected estimates allow for a check on how trustworthy the main model estimates are. We show these main model estimates, which illustrate the relationship between the screening scores and teacher outcomes, in the next section.

<A>5. Results

Below we present the estimated coefficients from the primary models that describe the relationship between the hiring rubrics used by SPS and the three primary teacher outcomes. We then report findings from models that correct for sample selection.

Prior to discussing the findings for the hiring rubrics, it is worth noting several unreported regression coefficients that are generally consistent with existing empirical literature. We find that students eligible for free or reduced-price lunch score about 0.06–0.07 standard deviations lower than those who are not eligible. We additionally estimate alternate specifications with years of experience included in the model. Students assigned to first year teachers relative to those assigned to second year teachers score about 0.03–0.06 standard deviations lower on the state assessment, similar to estimates from the literature (Rockoff 2004; Rivkin, Hanushek, and Kain, 2005; Clotfelter, Ladd, and Vigdor, 2006; Goldhaber and Hansen 2013). In our other outcomes, teachers in their first or second year are predicted to be absent about one day fewer than teachers with three to five years of experience, and almost three fewer days than teachers with five to ten years of experience.

<B>Predictive Validity of Screening Scores

The predicted relationships between the 21- and 60-point screening scores and teacher outcomes are presented in table 3. The rubric scores have been normalized so that the coefficients should be interpreted as the effect of a one standard deviation change in an applicant's score on the teacher fixed effect predicting student achievement in math (Panel A column 1) and reading (Panel A column 2), total annual absences as measured in days (Panel B), and the marginal effect of the screening score on attrition from the district (Panel C).

Applicant scores on the 21-point rubric have a positive but insignificant relationship with teacher effectiveness in both math and reading (Specification 1, the top row in table 3). The relationship between the 60-point rubric and teacher effectiveness (Specification 3) is greater than the 21-point score for both subjects, and is statistically significant for math. Both the 21- and 60-point results, including the nonsignificant results, predict improvements that we find to be educationally meaningful. Students assigned to teachers who score one standard deviation higher on the 60-point rubric are predicted to have student achievement that is 0.064 standard deviations higher in math and 0.033 standard deviations higher in reading. These effects are similar, for instance, to the above-estimated difference in achievement associated with being assigned to a novice teacher versus a second-year teacher (0.03–0.06 standard deviations). Effect sizes of this magnitude also indicate the ability to meaningfully separate high- and low-quality teachers. To demonstrate this, we separate the sample into quartiles based on quartile of the 60-point screening score. Students assigned to the average teacher in the top quartile are predicted to have student achievement that is 0.158 standard deviations higher in math (significant at the 5 percent level) and 0.103 standard deviations higher in reading (significant at the 10 percent level).

Teacher absences are positively associated with the 21-point score whether it is entered into the model separately (Specification 1) or in tandem with the 60-point screening score (Specification 3), but it is not significant. The point estimate associates a one standard deviation increase in the 21-point screening score with an increase in teacher absences of about half of a day. The total 60-point screening score is small and insignificant in each specification. The lack of a significant relationship between screening scores and teacher absences is consistent with Rockoff et al. (2011). Additionally, when we include experience in the model, the coefficient magnitudes decrease substantially and remain insignificant. Previous research has found a strong positive relationship between experience and teacher absences (Clotfelter, Ladd, and Vigdor 2011; Herrmann and Rockoff 2012), plausibly because teachers with experience are more likely to be tenured (Miller, Murnane, and Willett 2008) and have children.

Perhaps the most intriguing results deal with teacher attrition, because they suggest meaningfully large effects and are based on a broad sample of teachers in the district. The 21- and 60-point scores are both predictive of lower rates of attrition from the district. A one standard deviation increase in the 60-point score predicts a decrease in attrition of about three percentage points. Given that the baseline amount of attrition in the first year is about 20 percent (see table 2), this is not a small change. Estimates predicting attrition from the school, or from the profession entirely, are presented in table C.4 (available on the *Education Finance and Policy* Web site). Coefficients are slightly larger for school attrition and smaller for attrition from the profession, but the pattern is broadly similar.

The results above are estimated on the broadest sample available for each outcome. However, a group of particular interest is new teachers, whose performance may be especially hard to predict. In table 4 we limit the sample to outcomes observed in the first year after being hired and estimate the model separately for those who have no previous teaching experience at the time of application, and those with experience. Additionally, we estimate the model separately for those who end up teaching in SPS and those who teaching elsewhere.[27] This addresses the possibility that results could simply be indicative of the returns to match quality or job satisfaction, as well as the possibility that validity is different at high scores (more likely hired into SPS) and low scores (less likely).

When splitting the sample by prior experience, we find that effect sizes for student achievement are stronger among new teachers than among experienced teachers, although due to small sample sizes these estimates are not precise. Attrition results, on the other hand, are stronger among experienced teachers, and the screening scores are not statistically significant predictors of district attrition among new teachers. However, the point estimate among new teachers in the first year is similar to the main model result from table 3, and so imprecision is

[27] We also estimate the models separately by school level (elementary, middle, and high) and find that the 60-point score predicts more strongly for student achievement and attrition in middle school (see table C.3 on the *Education Finance and Policy* Web site).

again a problem. Overall, small sample sizes prevent us from coming to strong conclusions about the ability of the screening scores to predict performance among new teachers, but point estimates suggest the screening scores may be significant predictors of applicant performance in a larger sample. Comparing teachers in and out of SPS, the coefficient estimates of interest (those on $SCREEN_{j(tprior)}$) are qualitatively similar for most outcomes when the sample is restricted to applicants employed by SPS. Results are somewhat stronger inside SPS for student achievement and attrition. This may be attributable to the aforementioned match quality issue (the selection tools being scored with SPS in mind), measurements being more reliable for those with prior district experience (Martinkova and Goldhaber 2015), or the sensitivity of the screening instrument along the performance distribution.[28]

The results discussed thus far emphasize the relationship between the summative screening scores and teacher outcomes. Another result of interest is which aspects of teacher ability are most closely related to teacher outcomes, or at least which aspects can be measured accurately enough in screening such that those measures can help predict teacher outcomes.[29] Estimates from Specification 4, which evaluates the impact of each of the subcomponents of the screening scores one at a time, are in table 5. To avoid issues of collinearity between the subcomponents (as illustrated in the correlation table in table C.2 on the *Education Finance and Policy* Web site), each coefficient in table 5 is from its own regression.

The individual subcomponents of the screening scores, like the summative screening scores, fail to predict teacher absences, with the exception of Experience, which becomes small

[28] We can further refine the results by limiting the in-SPS sample to only those who were hired for the job for which they applied, and use only 60-point screening scores collected with that particular job in mind. When we do this, math and attrition estimates lose some precision but point estimates do not change. The reading estimate becomes larger (the coefficient rises to about .05) and becomes statistically significant at the 10 percent level.
[29] Further information on the reliability of the subcomponents is in Martinkova and Goldhaber (2015)

and insignificant when teacher experience is included in the model. However, some of the individual subcomponents appear to be driving the relationships found between the summative scores and both student achievement and teacher attrition. Each of the subcomponents on the 60-point rubric receives an equal weight in the construction of the overall rating, yet there are substantial differences in their respective coefficient estimates.

For instance, the coefficient on Classroom Management is relatively large for both math and reading—for reading it is one of only three significant coefficients. Training, Flexibility, and Instructional Skills are also significant and large for math. The lack of significance for Certificate and Education is noteworthy given that certification is a measure to which many school systems give primacy (USDOE 1997). More generally, the fact that the estimated coefficients of the subcomponents of the 60-point rubric are substantially different from one another suggests that a reweighting of these subcomponents could increase the ability of the 60-point rubric rating to predict teacher effectiveness.

Training, Classroom Management, Flexibility, and Instructional Skills are also predictive of teacher attrition. In addition to these, Experience, Interpersonal Skills, and Preferred Qualifications significantly predict attrition. Like in mathematics achievement, a reweighting of the summative scores could improve attrition outcomes. In results available from the authors, we report such an "optimal weighting" which roughly doubles the predictive validity of the screening scores for math and reading value added, as well as attrition.[30]

[30] We use canonical correlation to generate a set of weights for each component such that the weighted summative score maximizes the correlation with the outcome of interest. The main model coefficients on the weighted sum are 0.144 ($p<0.01$), 0.051, -0.405, and -0.049 ($p<0.05$) for math value added, reading value added, absences, and district attrition, respectively.

<B>Correction for Selection Bias

The above findings suggest that the screening instruments are predictive of key teacher outcomes, but there is reason to be concerned these findings are biased by sample selection. As outlined in section 4, we suggest that if we find no evidence of selection bias, this is reason to support the estimates in section 5, rather than to prefer the bias-corrected estimates presented here.

We present the first stage probit estimates of the selection model in table 6. The level of observation here is at the job application level. As such, there are about ten times as many observations here than in the other analyses, which are at the applicant-year level.

The excluded variables perform as expected in the first stage model (column 1). An applicant is more likely to be hired if an erroneously high 21-point screening score leads to the applicant advancing to the 60-point screening stage of the hiring process. An applicant who faces a higher level of competition for a particular job, as measured by the average 21-point screening score among other applicants, is significantly less likely to be hired. The F-test on the excluded instruments is 103.58, suggesting sufficient exogenous variation in hiring to control for selection into the sample.

The second column of the table presents the results of a placebo test. We estimate the probability of being hired in another district in Washington after not being hired by SPS. We find no significant relationship between the excluded variables and the likelihood of being employed in another district. This supports our use of the exclusion restriction, since the excluded variables directly affect the chances of being hired by SPS, rather than representing unobserved characteristics that might be correlated with both the general probability of employment and teacher outcomes.

Second-stage estimates of the relationship between screening scores and outcomes are in table 7. Estimates are presented without selection correction, and then with the correction. The sample used here does not match table 3 since it includes only those who were hired into a position for which they applied. Precision is lost in this more restrictive sample.

Coefficients on the screening scores for all outcomes are very similar with and without the selection correction. In each case, there are only small and insignificant differences between selection-corrected and uncorrected estimates, suggesting the bias introduced by the process of sample selection does not greatly affect estimates in the main model.[31]

For each model, we test the plausibility of the exclusion restrictions by estimating a pseudo-Sargan test for over-identification. Residuals from the second stage are regressed on the set of excluded variables. The F-test of all excluded variables is a rough test on the exogeneity of the excluded variables. We cannot reject the hypothesis that the residuals are unrelated to the excluded variables except through selection, buttressing the findings from the placebo test.

<A>6. Policy Implications and Conclusions

Our findings show that the two screening rubrics used by Spokane Public Schools predict teacher effectiveness and teacher attrition—but not teacher absences—in expected ways. A one standard deviation increase in the 60-point screening score is associated with a 0.06 standard deviation increase in math achievement and a marginally significant 0.03-0.04 increase in reading achievement, comparable to the difference between a first-year teacher and a second-year teacher. The same one standard deviation increase in the screening score is associated with a

[31] The finding that differences are "small" is not well defined. In results available from the authors, we perform a power analysis and find the model has 80 percent power to detect a difference of 0.064 or larger in the achievement model. We find no such differences.

decrease in attrition of three percentage points. Since the turnover of a single teacher can cost a district about $10,000 (Barnes, Crowe, and Schaefer 2007), improved hiring practices have the ability to both improve effectiveness and save money.

The screening scores used by SPS represent the value of guided human interpretation of somewhat subjective information (such as that contained in letters of recommendation). We validate the notion that a system with this type of guidance on interpreting applications is capable of effectively identifying effective teachers and likely improves on the ad-hoc hiring processes typically seen in public schools (e.g., Ebmeier and Ng 2006; Oyer and Schaefer 2011). Our results also speak to hiring processes more broadly, and the question of how structured and computerized hiring should be. In some low-skill private sector jobs, for example, hiring is heavily based on low-cost computerized assessments. The hiring procedure that SPS uses is neither the typical ad-hoc hiring process seen in many schools, nor is it a fully computerized system that ignores difficult-to-parse information like letters of recommendation. One natural question is how much is added by SPS' use of structure as compared to a fully subjective process. Another natural question is how much is added by SPS' use of subjective criteria and coding as compared to a fully objective process, even if such a process may not be tenable in the teacher labor market.

We can get a clearer sense of the value of the subjective information these assessments provide by estimating what predictive power the screening scores offer beyond what could be achieved by making hiring decisions on the basis of objective, observable factors alone. In table C.6, we show the predictive validity of factors derived from objective criteria (as shown in table C.5; tables are available on the *Education Finance and Policy* Web site), with and without screening scores. The student-level standard deviation in teacher effectiveness predicted by the purely objective factors is 0.077 for math and 0.063 for reading, as compared with the total standard deviations of 0.296 and 0.219, as derived from equation 1.[32] Available objective information in the application explains about 7 percent of the variance in teacher effectiveness in math and 8 percent in reading. With the addition of screening scores, the estimated standard deviation in the measured teacher effect increases to 0.089 for math and 0.071 for reading, so that the explained portion of the variance of teacher effectiveness rises to nine percent in math (a 30 percent increase) and 10 percent in reading (a 25 percent increase). These results can be compared to Rockoff et al. (2011), who find that nonstandard criteria increase the standard deviation of estimated teacher effectiveness by 50 percent.

These screening scores are related to significant gains in student achievement and attrition, and offer important data beyond what is typically available in a teacher's hiring packet. However, our findings for SPS may not generalize across all school districts. The SPS district is seen as a desirable place to work in eastern Washington and does not face the same competition for teacher labor as other districts. They also hire a high percentage (more than 70 percent) of its workforce from among people who already have experience there. Although we find the screening scores do predict outcomes for those who end up working outside SPS, it is possible that the predictive abilities of the screening rubrics are aided by screener familiarity with those who write the letters of recommendation. That said, these results are consistent with the wider literature on screening at the hiring stage in other industries as well as in teaching.

Teachers have a strong and lasting effect on their students (Chetty, Friedman, and Rockoff 2014b). The idea of improving the quality of the teacher workforce through more effective hiring is appealing given the high dollar and political costs of dismissing ineffective

[32] These estimates are larger than estimated for statewide value added. Statewide math and reading value added have variances of 0.19 and 0.18, respectively (Goldhaber and Theobald 2013).

teachers who are in-service (Treu 2014), and empirical evidence that other teacher performance interventions (such as professional development or performance incentives) tend to have only marginal impacts on productivity (Yoon et al. 2007).

Unlike previous work on screening processes, including work outside of the realm of teacher labor markets, we are able to observe applicants who are not employed in the district for which they applied. The observation of un-hired applicants strengthens the analysis by allowing us to predict performance outside the district and thereby correct for sample selection bias. Given the novelty of this work to the employee screening literature, our results provide general support for the broader use of standardized screening instruments to process subjective applicant information in public education and other sectors of employment.

We show a strong relationship between the performance on selection instruments and some measures of in-service teacher quality. This relationship likely overstates what is possible in terms of improving the teacher workforce as a whole because school systems compete with one another in the teacher labor market. Nevertheless, since many school districts rely on far more informal processes for selecting teachers, and likely lose some potentially talented teachers to other occupations at the hiring stage, there appears to be substantial room for improving the quality of the teacher workforce through greater use and refinement of formal selection instruments. It is worth noting that SPS's screening process utilizes applicant information that is already collected by most school districts during the hiring process (i.e., letters of recommendation, education and credentials, and professional experience). Moreover, SPS developed its screening tools in-house and the only costs that we are aware of involve those associated with the amount of time SPS personnel spend screening applicants. Thus, from a financial standpoint, the use of selection instruments may represent an efficient means of improving the quality of the teacher workforce.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25(1): 95-135.

Angrist, Joshua D. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* 66(2): 249-288.

Bacher-Hicks, Andrew, Thomas Kane, and Douglas O. Staiger. 2014. Validating teacher effect estimates using changes in teacher assignments in Los Angeles. NBER Working Paper No. 20657.

Ballou, Dale. 1996. Do public schools hire the best applicants? *The Quarterly Journal of Economics* 111(1): 97-133.

Barnes, Gary, Edward Crowe, and Benjamin Schaefer. 2007. The cost of teacher turnover in five states: A pilot study. Washington, DC: National Commission on Teaching and America's Future.

Blundell, Richard, and Monica Costa Dias. 2000. Evaluation methods for non-experimental data. *Fiscal Studies* 21(4): 427-468.

Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2009. Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis* 31(4): 416-440.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, Matthew Ronfeldt, and James H. Wyckoff. 2011. The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management* 30(1): 88-110.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2013. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics* 31(1): 83-117.

Chamberlain, Gary E. 2013. Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences of the United States of America* 110(43): 17176-17182.

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

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014b. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review* 104(9): 2633-2679.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41(4): 778-820.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2011. Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy* 6(3): 399-438.

Dee, Thomas, and James Wyckoff. 2013. Incentives, selection, and teacher performance: evidence from IMPACT. NBER Working Paper No. 19529.

Dobbie, Will. 2011. Teacher characteristics and student achievement: Evidence from Teach For America. Unpublished paper, Harvard University.

Ebmeier, Howard, and Jennifer Ng. 2006. Development and field test of an employment selection instrument for teachers in urban school districts. *Journal of Personnel Evaluation in Education* 18(3): 201-218.

Garet, Michael S., Andrew C. Porter, Laura Desimone, Beatrice F. Birman, and Kwang Suk Yoon. 2001. What makes professional development effective? Results from a national sample of teachers. *American Education Research Journal* 38(4): 915-945.

Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Marian Eaton, Kirk Walters, Mengli Song, Seth Brown, Steven Hurlburt, Pei Zhu, Susan Sepanik, and Fred Doolittle. 2011. *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington DC: U.S. Department of Education, Institute of Education Sciences.

Glazerman, Steven, Daniel P. Mayer, and Paul T. Decker. 2006. Alternative routes to teaching: The impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management* 25(1): 75-96.

Goldhaber, Dan, and Dominic Brewer. 2000. Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* 22(2): 129-145.

Goldhaber, Dan, and Michael Hansen. 2013. Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica* 80(319): 589-612.

Goldhaber, Dan, and Roddy Theobald. 2013. Managing the teacher workforce in austere times: The determinants and implications of teacher layoffs. *Education Finance and Policy* 8(4): 494-527.

Goldhaber, Dan, Brian Gabele, and Joe Walch. 2014. Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy* 1(1): 28-39.

Greenberg, Julie, Arthur McKee, and Kate Walsh. 2013. Teacher prep review: A review of the nation's teacher preparation programs. Washington, DC: National Council on Teacher Quality.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. The market for teacher quality. NBER Working Paper No. 11154.

Hanushek, Eric A., and Steven G. Rivkin. 2010. Generalizations about using value-added measures of teacher quality. *American Economic Review* 100(2): 267-271.

Harris, Douglas N., Stacey A. Rutledge, William K. Ingle, and Cynthia C. Thompson. 2010. Mix and match : What principals really look for when hiring teachers. *Education Finance and Policy* 5(2): 228-246.

Heckman, James J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1): 153-161.

Heneman III, Herbert G., and Timothy A. Judge. 2003. *Staffing organizations*. 4th edition. Middleton, WI: McGraw-Hill/Irwin House.

Herrmann, Mariesa A., and Jonah E. Rockoff. 2012. Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics* 30(4): 749-782.

Hill, Heather C., and Pam Grossman. 2013. Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review* 83(2): 371-385.

Hinrichs, Peter. 2014. What kind of teachers are schools looking for? Evidence from a randomized field experiment. Federal Reserve Bank of Cleveland Working Paper No. 14-36.

Ingersoll, Richard M., and David Perda. 2010. Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal* 47(3): 563-594.

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

Kodel, Cory, Kata Mihaly, Jonah E. Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47: 180-195.

Lazear, Edward, and Michael Gibbs. 2007. *Personnel economics for managers*. 2nd Edition. Available http://facultysobweb.bcit.ca/kevinw/4800/documents/Person-econ-ch_1to6.pdf. Accessed 13 June 2016.

Liu, Edward, and Susan M. Johnson. 2006. New teachers' experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly* 42(3): 324-60.

Maddala, Gangadharrao S. 1983. *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

Martinkova, Patricia, and Dan Goldhaber. 2015. Improving teacher selection: The effect of inter-rater reliability in the screening process. Center for Education Data & Research Working Paper No. 2015-7.

Metzger, Scott A., and Meng-Jia Wu. 2008. Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research* 78(4): 921-940.

McDaniel, Michael A., Frank L. Schmidt, and John E. Hunter. 1988. A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology* 41(2): 283-309.

Miller, Reagen T., Richard J. Murnane, and John B. Willett. 2008. Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis* 30(2): 181-200.

Neal, Derek. 2011. The design of performance pay in education. In *Handbook of the economics of education*, Vol. 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, pp. 495-550. Amsterdam: North-Holland.

Olsen, Randall J. 1980. A least squares correction for selectivity bias. *Econometrica* 48(7): 1815-1820.

Oyer, Paul, and Scott Schaefer. 2011. Personnel economics: Hiring and incentives. In *Handbook of Labor Economics*, Vol. 4, edited by David Card and Orley Ashenfelter, pp. 1769-1823. Amsterdam: Elsevier B.V.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73(2): 417-458.

Rockoff, Jonah E. 2004. The Impact of Individual Teachers on Students' Achievement: Evidence from Panel Data. *American Economic Review* 94 (2): 247–252.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2011. Can you recognize an effective teacher when you recruit one? *Education Finance and Policy* 6(1): 43-74.

Ronfeldt, Matthew, Susanna Loeb, and Jim H. Wyckoff. 2013. How teacher turnover harms student achievement. *American Educational Research Journal* 50(1): 4-36.

Shaw, Kathryn, and Edward Lazear. 2007. Personnel economics: The economist's view of human resources. 21(4): 91-114.

Society for Industrial and Organizational Psychology. 2014. *Employment testing*. Available www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx. Accessed October 20, 2015.

Springer, Matthew G., Dale Ballou, Laura S. Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2010. *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Available www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf. Accessed 7 June 2016.

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

Strauss, Robert P., Lori R. Bowes, Mindy S. Marks, and Mark R. Plesko. 2000. Improving teacher preparation and selection: Lessons from the Pennsylvania experience. *Economics of Education Review* 19(4): 387-415.

Treu, Rolf M. 2014. *Vergara vs. State of California Tentative Decision*. Available http://studentsmatter.org/wp-content/uploads/2014/06/Tenative-Decision.pdf. Accessed 7 June 2016.

U.S. Department of Education (USDOE). 1997. *Credentials and tests in teacher hiring: What do districts require?* Washington DC: U.S. Department of Education.

Wayne, Andrew J., Kwang Suk Yoon, Pei Zhu, Stephanie Cronen, and Michael S. Garet. 2008. Experimenting with teacher professional development: Motives and methods. *Educational Researcher* 37(8): 469-479.

Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. Shapley. 2007. *Reviewing the evidence on how teacher professional development affects student* *achievement* (Issues & Answers Report REL 2007-No. 033). Washington DC: U.S. Department of Education, Institute of Education Sciences.

Young, I. Phillip, and Dane A. Delli. 2002. The validity of the teacher perceiver interview for predicting performance of classroom teachers. *Educational Administration Quarterly* 38(5): 586-612.

Table 1. Applicant Screening Scores: Descriptive Statistics

		Unadjusted:					**Adjusted:**				
		Obs.	Mean	SD	Min.	Max.	Obs.	Mean	SD	Min.	Max.
21-Point Pre-Screening Rubric											
Rater Total Rating		3,946	16.1	(2.4)	4.5	21	3,946	16.1	(2.4)	4.5	21
Calculated Total Rating		2,633	15.9	(2.4)	3.5	21	3,946	16.2	(3.8)	3.5	21
21-Point Components	Experience	2,631	4.4	(0.8)	1	6	3,946	4.3	(1.4)	1	6
	Depth of Skills	2,633	4.8	(0.9)	1	6	3,946	4.5	(1.3)	1	6
	Recommendations	2,632	4.5	(0.9)	1	6	3,946	5.0	(1.0)	1	6
60-Point Screening Rubric											
Rater Total Rating		1,695	37.9	(7.5)	10	61	1,707	41.3	(7.3)	10	61
Calculated Total Rating		1,707	37.9	(7.5)	10	59	1,707	41.3	(7.3)	10	59
60-Point Components	Certificate and Education	1,671	5.1	(1.0)	0	6	1,707	5.0	(1.0)	0	6
	Training	1,702	3.9	(1.2)	0	6	1,707	3.9	(1.2)	0	6
	Experience	1,706	4.0	(1.1)	0	6	1,707	4.0	(1.1)	0	6
	Management	1,700	4.1	(1.0)	0	6	1,707	4.0	(1.0)	0	6
	Flexibility	1,703	4.2	(1.0)	0	6	1,707	4.2	(1.0)	0	6
	Instructional Skills	1,706	4.1	(1.0)	0	6	1,707	4.1	(1.0)	0	6
	Interpersonal Skills	1,703	4.4	(1.0)	0	6	1,707	4.4	(1.0)	0	6
	Cultural Competency	1,702	4.0	(1.0)	0	6	1,707	4.0	(1.0)	0	6
	Preferred Qualifications	1,470	3.9	(1.3)	0	6	1,707	3.6	(1.4)	0	6
	Letters of Rec.	716	4.2	(1.1)	0	6	1,707	4.1	(0.8)	0	6

Table 2. Outcome Variable Summary Statistics

Panel A: Applicant Information	All	21-Pt Rating	60-Pt Rating	Interview	Hired/ Offered	Hired Elsewhere
Total Obs. (Teacher/Yr.)	4,215	3,946	1,707	1,236	538	496
Total Proportions	1.00	0.94	0.40	0.29	0.13	0.12
Certificated Employment Experience in Year Applied						
No Experience	0.83	0.84	0.69	0.63	0.49	0.53
SPS District	0.11	0.09	0.22	0.28	0.43	0.03
Other District	0.07	0.07	0.09	0.09	0.08	0.44
Calculated Experience	3.18	3.22	3.85	3.70	3.23	4.41
	(4.66)	(4.63)	(5.01)	(4.73)	(4.22)	(5.28)
Student Teaching in SPS	0.36	0.37	0.40	0.42	0.47	0.29
21-Point Pre-Screening	NA	16.10	16.99	17.16	17.27	16.46
Rubric Summative Rating		(2.38)	(2.22)	(2.20)	(2.20)	(2.24)
60-Point Screening	NA	NA	41.31	43.60	45.61	40.09
Rubric Summative Rating			(7.29)	(6.13)	(5.75)	(6.76)
WESTB Average	-0.07	-0.07	-0.03	-0.02	0.02	-0.04
(Standardized statewide)	(0.75)	(0.75)	(0.74)	(0.75)	(0.70)	(0.75)

Panel B: Outcomes[a]	All	21-Pt Rating	60-Pt Rating	Interview	Hired/ Offered	Hired Elsewhere
Value-Added						
Math	-0.05	-0.07	-0.03	-0.03	-0.01	-0.08
(N=196 Teacher/Yr.)	(0.21)	(0.20)	(0.21)	(0.21)	(0.21)	(0.19)
Reading	-0.08	-0.09	-0.08	-0.07	-0.07	-0.09
(N=202 Teacher/Yr.)	(0.17)	(0.17)	(0.17)	(0.18)	(0.18)	(0.17)
Absences (N=374 Teacher/Yr.)						
Total Annual Absences	6.92	6.62	7.38	7.51	7.26	5.28
	(5.35)	(5.09)	(5.25)	(5.33)	(5.34)	(5.10)
Attrit within 1 Year (N=1024 Teacher/Yr.)						
School	0.46	0.46	0.45	0.44	0.40	0.48
District	0.30	0.31	0.29	0.27	0.20	0.40
K-12 WA Public Schools	0.22	0.23	0.22	0.21	0.17	0.18
Attrit within 3 Years (N=774 Teacher/Yr.)						
School	0.46	0.46	0.45	0.43	0.40	0.52
District	0.31	0.33	0.30	0.29	0.21	0.43
K-12 WA Public Schools	0.22	0.23	0.22	0.22	0.17	0.19

Notes: No experience, experience in SPS, and experience in other districts determined by identifying applicants as being employed in a certificated teaching position. Value-added scores are estimated as a derivative of equation 1. WESTB scores are centered at mean zero at the state level with standard deviations of approximately 0.20 and 0.16 for math and reading, respectively (depending on year).

[a] Observation numbers in the Outcomes panel represent the number of applications (at the teacher/year level) with associated outcome data.

Table 3. Predicting Teacher Effectiveness with Summative Screening Scores

Panel A:	Math	Reading
(Spec. 1) 21-Point Score	N = 220 (184)[a]	N = 229 (189)
21-Point Score	0.032	0.024
	(0.022)	(0.015)
(Spec. 2) 60-Point Score	N = 152 (127)	N = 151 (126)
60-Point Score	0.064**	0.033
	(0.028)	(0.025)
(Spec. 3) 21- *and* 60-Point Scores	N = 130 (106)	N = 128 (104)
21-Point Score	0.016	0.029
	(0.025)	(0.021)
60-Point Score	0.030	0.003
	(0.035)	(0.032)

	Panel B: Annual Absences	**Panel C:** District Attrition
(Spec. 1) 21-Point Score	N = 453 (335)	N = 1,210 (617)
21-Point Score	0.416	-0.019*
	(0.300)	(0.011)
(Spec. 2) 60-Point Score	N = 287 (213)	N = 1,265 (633)
60-Point Score	-0.083	-0.030***
	(0.508)	(0.011)
(Spec. 3) 21- *and* 60-Point Scores	N = 272 (205)	N = 1,092 (560)
21-Point Score	0.415	-0.026**
	(0.699)	(0.012)
60-Point Score	-0.126	-0.027**
	(0.534)	(0.013)

Notes: For attrition, marginal effects at the mean of the data are presented. R^2 values are approximately 0.1 for teacher effectiveness and attrition, and 0.1-0.2 for teacher absences.
[a] The number of clusters in each analysis is presented in parentheses next to the total number of observations.
* p < 0.10; ** p < 0.05; *** p < 0.01.

Table 4. Split Sample Analyses

		Teacher Math Effect	Teacher Reading Effect	Total Annual Absences	District Attrition
		(1)	(2)	(3)	(4)
Panel A: Split by Previous Experience					
Main Model (First Year Only)	60-Point Rubric Summative Score	0.069**	0.054*	-0.212	-0.053***
		(0.033)	(0.032)	(0.502)	(0.017)
	Observations	94 (75) [a]	103 (83)	120 (120)	625 (625)
	R-squared	0.278	0.080	0.254	0.056
Split Sample (First Year Only)	**Zero Previous Teaching**				
	60-Point Rubric Summative Score	0.062	0.186*	-0.307	-0.021
		(0.124)	(0.099)	(2.727)	(0.029)
	Observations	24 (22)	26 (24)	24 (24)	190 (190)
	R-squared	0.329	0.264	0.732	0.231
	One or More Years Previous Teaching				
	60-Point Rubric Summative Score	0.031	0.003	-0.373	-0.067***
		(0.036)	(0.029)	(0.494)	(0.023)
	Observations	70 (53)	77 (59)	96 (96)	435 (435)
	R-squared	0.165	0.089	0.350	0.051
Panel B: Split by Employed in Spokane or Elsewhere after Hiring Process					
Main Model	60-Point Rubric Summative Score	0.064**	0.039	-0.083	-0.030***
		(0.028)	(0.031)	(0.508)	(0.011)
	Observations	152 (127)	151 (126)	287 (213)	1,265 (633)
	R-squared	0.162	0.089	0.169	0.073
Split Sample	**Inside Spokane**				
	60-Point Rubric Summative Score	0.068**	0.032	0.236	-0.027**
		(0.033)	(0.031)	(0.555)	(0.012)
	Observations	117 (94)	116 (94)	198 (146)	966 (469)
	R-squared	0.115	0.099	0.191	0.158
	Outside of Spokane				
	60-Point Rubric Summative Score	0.058	-0.024	-0.606	-0.018
		(0.047)	(0.035)	(1.119)	(0.030)
	Observations	35 (34)	35 (32)	89 (67)	299 (171)
	R-squared	0.542	0.214	0.272	0.266

Notes: All regressions displayed in this table are run with identical controls and predictor variables as the primary outcome models above. For attrition, we present marginal effects at the mean of the data. For the new teacher split, observation numbers do not match because only first year outcomes are used. For the SPS split, subsample observation numbers do not add up to the full sample because some teachers teach both inside and outside SPS, and in attrition subsample regressions because some observations were dropped due to perfect prediction in the subsample. * p < 0.10; ** p < 0.05; *** p < 0.01.

[a] The number of clusters in each analysis is presented in parentheses next to the total number of observations.

Table 5. Predicting Teacher Effectiveness with Screening Score Subcomponents

Panel A: 21-Point Subcomponents[b]	Math N = 220 (184)[a]	Reading N = 229 (189)	Yearly Absences N = 453 (335)	District Attrition N = 1,210 (617)
Experience	0.013	0.009	0.410*	-0.075
	(0.020)	(0.015)	(0.246)	(0.102)
Skills	0.020	0.009	0.102	-0.154*
	(0.019)	(0.014)	(0.297)	(0.086)
Recommendations	0.041*	0.029*	0.186	-0.132
	(0.023)	(0.016)	(0.264)	(0.094)

Panel B: 60-Point Subcomponents	Math N = 152 (127)	Reading N = 151 (126)	Yearly Absences N = 287 (213)	District Attrition N = 1,265 (633)
Certificate & Education	0.025	-0.000	0.298	0.003
	(0.040)	(0.029)	(0.549)	(0.012)
Training	0.062**	0.040	0.150	-0.020*
	(0.030)	(0.025)	(0.547)	(0.012)
Experience	0.037	0.010	1.116**	-0.028**
	(0.034)	(0.027)	(0.453)	(0.011)
Classroom Management	0.131**	0.043*	-0.291	-0.025**
	(0.032)	(0.026)	(0.500)	(0.010)
Flexibility	0.090**	0.032	-0.175	-0.026**
	(0.032)	(0.029)	(0.602)	(0.011)
Instructional Skills	0.059*	0.033	-0.410	-0.031***
	(0.033)	(0.026)	(0.574)	(0.011)
Interpersonal Skills	0.037	0.010	-0.553	-0.037***
	(0.037)	(0.028)	(0.474)	(0.011)
Cultural Competency	0.016	-0.004	-0.004	-0.012
	(0.026)	(0.023)	(0.477)	(0.011)
Preferred Qualifications	0.028	0.041	0.369	-0.025**
	(0.031)	(0.026)	(0.644)	(0.012)
Letters of Recommendation	-0.062	-0.070**	-0.297	-0.009
	(0.045)	(0.023)	(0.425)	(0.013)

Notes: For attrition, we present marginal effects at the mean of the data. Each coefficient is from its own regression. R^2 values are approximately 0.1 for teacher effectiveness and attrition, and 0.1-0.2 for teacher absences.
[a] The number of clusters in each analysis is presented in parentheses next to the total number of observations.
[b] Each subcomponent coefficient is estimated in a separate regression.
* p < 0.10; ** p < 0.05; *** p < 0.01.

Table 6. Generalized First Stage Predicting Being Hired for Heckman Selection

	Hired	**Placebo (Hired Elsewhere)**
21-Pt Screen	0.008***	0.028***
	(0.001)	(0.008)
Excluded Variables:		
Error in Teacher's Favor	0.014***	0.002
	(0.002)	(0.017)
21-Pt Screen	-0.012***	-0.005
Competition	(0.002)	(0.014)
Observations	41,866 (3,937) [a]	18,236 (1,329)
F(Excluded Variables)	103.58***	0.938

Notes: Models are estimated using probit with an unreported constant term. Marginal effects from a probit regression are presented. No additional controls are included.
[a] The number of clusters in each analysis is presented in parentheses next to the total number of observations.
*** $p < 0.01$.

Table 7. The Effect of Screening Scores on Outcomes, With and Without Selection Correction

	Math		Reading	
VARIABLES	(1)	(2)	(3)	(4)
60-Pt Screen	0.064	0.062	0.027	0.015
	(0.082)	(0.058)	(0.089)	(0.092)
Mills Ratio (λ)		-0.085		-0.247
		(0.220)		(0.228)
Observations	73 (59) [a]		69 (56)	
R-Squared	0.163		0.113	
OverID p-value		0.755		0.604
	Absences		1-Year District Attrition	
60-Pt Screen	0.218	0.162	-0.009	-0.012
	(1.142)	(1.181)	(0.091)	(0.097)
Mills Ratio (λ)		-0.602		-0.050
		(2.883)		(0.207)
Observations	140 (106)		189 (185)	
R-Squared	0.322		0.532	
OverID p-value		0.659		0.832

Notes: Estimates are produced using Specification 3 as presented in table 3, except that the sample is limited to those hired into SPS in the sampling window and the selection correction as generated in table 5 is included in models (2) and (4). For attrition, marginal effects at the mean of the data are presented.
[a] The number of clusters in each analysis is presented in parentheses next to the total number of observations.